Modulating the epigenetic state promotes the reprogramming of transformed cells to pluripotency in a line-specific manner

Xiuling Fu[1], Qiang Zhuang[1], Isaac A. Babarinde[1], Liyang Shi[1], Gang Ma[1], Haoqing Hu[2], Yuhao Li[1], Jiao Chen[1], Zhen Xiao[1], Boping Deng[1], Li Sun[1], Ralf Jauch[2], Andrew P. Hutchins[1]*

[1]Shenzhen Key Laboratory of Gene Regulation and Systems Biology, Department of Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, 518055, China.

[2]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

*Correspondence: andrewh@sustech.edu.cn

Key findings

- Somatic cell reprogramming of transformed cells is context-specific
- Inhibition of MEK converts some cell lines to reprogramming-capable
- Transformed cell lines are characterized by a hyperactive chromatin state
- MEK inhibition indirectly affects chromatin to enable reprogramming

Abstract

Somatic cell reprogramming and oncogenic transformation share surprisingly similar features, yet transformed cells are highly resistant to reprogramming. There must be barriers that block transformed cells from reprogramming, but the nature of those barriers is unclear. In this study, we generated a systematic panel of transformed mouse embryonic fibroblasts (MEFs) using a variety of oncogenic transgenes, and discovered transformed cell lines that remain compatible with reprogramming when transfected with *Oct4*/*Sox2*/*Klf4*/*Myc*. By comparing the reprogramming-capable and incapable transformed lines we identified multiple stages of failure in the reprogramming process. Some transformed lines failed very early, whilst other lines seemed to progress through a normal-looking reprogramming process. Finally, we show that MEK inhibition overcomes one critical reprogramming barrier by indirectly suppressing a hyperactive epigenetic state in some of the transformed cells. This study reveals that the barriers underlying resistance to reprogramming vary between the different transformation methods.

Keywords: Somatic cell reprogramming; oncogene; transformation.

Introduction

Transformed cells and embryonic stem cells (ESCs) have a remarkable list of similarities [1]. Both cell types have a relaxed chromatin state [2], adopt a glycolysis-biased metabolism despite the availability of oxygen [3, 4], can undergo an epithelial-mesenchymal transition (EMT) [5], and form teratomas when injected into immunocompromised mice [6]. Transformed cells acquire features reminiscent of embryonic development, such as increased cellular plasticity and the up-regulated expression of pluripotent genes, including *OCT4*, *NANOG*, and *SOX2* [7]. Indeed, expression of pluripotent-specific genes in patient tumor samples correlate with poor clinical outcomes [8, 9]. For example, overexpression of *Oct4* led to epithelial lesions [10], and *NANOG* expression is associated with colorectal [11] and prostate cancer [12].

Somatic cells can be reprogrammed to induced pluripotent stem cells (iPSCs) by the transfection of a cocktail of transgenes, particularly *Oct4* (*Pou5f1*), *Sox2*, *Klf4*, and *Myc* [13]. Curiously, despite being an artificial process, the reprogramming of somatic cells to iPSCs passes through distinct phases, reminiscent of a developmental program [14-16]. Tumorigenic transformation also passes through a series of distinct phases in a process that has similarities to somatic cell reprogramming [1, 17]. Some studies suggest direct links between reprogramming and cancer. For example, transient *in vivo* activation of reprogramming factors leads to tumor formation [18, 19], and cancer-associated mutations in the transcription factor *SOX17* can confer reprogramming capability to the normally incapable SOX17 [20].

There have been reports on reprogramming primary human cancer cells to an embryonic state, including cancerous cells from liver, gastrointestinal tract, and

sarcoma cell lines [21-24]. Additionally, blood cancer cells seem more amenable to be reprogrammed: there are reports on reprogrammed cells generated from human KBM7 cells, T cell acute lymphoblastic leukemia, acute myeloid leukemia [25-27], lymphoblastoid cells [28, 29] and chronic myeloid leukemia cells [24, 30-32]. However, these blood cells seem to be the exception, since in the general transformed cells are resistant to reprogramming [33]. Importantly, it is unclear how close those iPSC-like cells are to genuine iPSCs, and the iPSC-like cells may not be completely reprogrammed. For example, whilst the overall gene expression of the reprogramed cells is closer to iPSCs than the parental lines, it remains distinct from genuine ESCs/iPSCs [23, 28, 34]. It is also possible that reprogramming from primary tissues is selecting for bystander untransformed cells [35]. Finally, even when reprogramming appears possible, the efficiency is much lower than the already low efficiency of wild-type reprogramming. This is a curious contradiction. Considering the similarities between cancer cells and iPSCs and the pathways used to generate them, it seems that reprogramming should be easier in transformed cells as cancer cells often have reduced barriers to cell type conversion.

To explore the barriers blocking the reprogramming of transformed cells, we generated a panel of eight artificial transformed mouse cell lines. Seven of these lines can acquire OCT4-GFP+ reporter expression and pluripotent characteristics using a conventional OSKM-reprogramming protocol. Albeit at very low efficiency. Of the remaining three reprogramming-incapable lines, we show that the defects in reprogramming are line-specific. Reprogramming is a phased process and some lines fail in the early phases, others in the later phases. Compared to wild-type cells, the reprogramming-incapable lines show a heightened ‘hyperactive’ chromatin state and

demonstrate global increases in chromatin accessibility and histone acetylation. Some of the transformed lines could be converted to be reprogramming-capable by inhibiting MEK signaling which moderates the active chromatin state and leads to decreased histone acetylation.

Results

Mouse transformed cell lines are recalcitrant to reprogramming

To explore reprogramming in transformed cell lines we first attempted to reprogram mouse cell lines that were originally derived from a tumor or that were spontaneously transformed. These cell lines have been grown in culture for an extended period and their passage number is unknown. We chose three mouse cell lines: 3T3-L1, a spontaneously immortalized fibroblastic cell line; 4T1, a metastatic breast tumor cell line; and N2a (Neuro-2a), a neuroblastoma cell line (**Figure 1a**). We attempted to reprogram these cell lines using OSKM in serum+Vc conditions [36]. For the 3T3-L1 and 4T1 (but not the N2a), we observed colonies that morphologically resembled iPSCs on day 15 (**Figure 1a**). Low levels of NANOG protein could be detected by immunofluorescence in some 3T3-L1 and 4T1 colonies (**Figure 1b**). We attempted to manually pick the iPSC-like colonies and establish iPSC lines. However, contaminating 3T3-L1 and 4T1 cells would outcompete any iPSC-like colonies, and based on morphology, the cultures would rapidly revert to homogenous 3T3-L1 or 4T1 cultures. Hence, we sorted reprogrammed cells positive for ICAM and negative for CD44 (ICAM+/CD44-), which marks the late stage of reprogramming in wild-type MEFs [37]. ICAM+/CD44- cells were sorted from the reprogrammed 3T3-L1 and 4T1 cells, and we detected small numbers of ICAM+/CD44- cells at D15 of reprogramming

110 in 3T3-L1 and 4T1, but not N2a (**Figure 1c, d**). qRT-PCR of the ICAM+/CD44- cells
111 indicated that *Esrrb* and *Nanog* were up-regulated in the 4T1, and *Essrb*, *Nanog,* and
112 endogenous-*Pou5f1* in 3T3-L1 (**Figure 1e**). However, when the sorted cells were
113 replated they rapidly reverted to the original cell type morphology and there was no
114 evidence of iPSC-like cells (**Figure 1f**). Overall, these results highlight the difficulty in
115 reprogramming transformed cell lines to iPSCs.
116

117 **Features of untransformed MEF senescence in relation to reprogramming**

118 We first explored wildtype MEF reprogramming to understand the blocks on
119 reprogramming in the normal case so that we can rule them out in the transformed cell
120 lines. Wildtype MEFs can only reprogram at early passages [38], and the efficiency
121 declines rapidly before complete failure after passage 4 (**Figure S1a**), when using a
122 conventional reprogramming OSKM (*Oct4*, *Sox2*, *Klf4*, and *Myc*) system [36]. RNA-
123 seq of MEFs from passages 1-6 revealed changes in cell cycle genes, including the
124 downregulation of cyclins and other positive cell cycle regulators and the activation of
125 cell cycle inhibitors, particularly *Cdkn1a* (**Figure S1b**). We divided the samples into
126 MEFs that could reprogram successfully (P1, 2) or failed (P5, 6) and measured
127 significantly differentially expressed (DE) genes. In total, 337 genes were significantly
128 up-regulated and 419 significantly down-regulated (**Figure S1c, d**). Gene ontology
129 (GO) and gene set enrichment analysis (GSEA) of the differentially expressed genes
130 indicated that the down-regulated genes were related to the cell cycle and extracellular
131 matrix. Up-regulated genes were related to cell migration/adhesion, MAPK activity,
132 apoptosis, and inflammation (**Figure S1e, f**), possibly caused by reactive oxygen
133 species (**Figure S1f**). The inflammatory response is exemplified by the up-regulation

134 of cytokines and chemokines in the P5,6 MEFs (**Figure S1g**). We defined the up-
135 regulated gene set as the 'MEF senescent signature' and the down-regulated gene
136 set as a 'reprogramming permissive signature' (**Supplementary Table S1**).
137

138 **Generation of a panel of oncogenic transformed MEF cell lines**

139 As the above-transformed cell lines have been maintained in culture for an extended
140 time they have likely accumulated many genetic alterations to adapt to the cell culture
141 environment. These changes may permanently impair their ability to reprogram.
142 Consequently, to achieve a controlled system we generated immortalized MEFs using
143 9 different combinations of factors to represent a spectrum of transformation. The
144 factors were chosen to cover different methods of immortalization: oncogenic
145 transcription factors (*Myc*, Hras^{G12V}, Mef2d, p53DD), viral transforming factors (SV40T,
146 E1A), apoptotic factors (Bcl2), and an engineered epigenetic factor (Hdac7SA; serines
147 at positions 178, 344, and 479 substituted with alanine, to block nuclear export) [39-
148 41]. The sequences of the transformation factors are in **Supplementary Table S2**.
149 Not all of these factors could successfully transform MEFs, and, except for SV40T, at
150 least two factors were required for successful transformation (**Figure S2a, b**). The
151 immortalized cell lines were maintained for at least 1 month to remove any non-
152 transformed background MEFs. After 1 month resulting lines still expressed the
153 appropriate transgenes, as confirmed by RNA-seq, RT-qPCR, and Western blot
154 (**Figure S2c-e**). The exception was the p53DD.Myc lines, as the p53DD transgene
155 was silenced at the protein level, whilst *Myc* could be detected by both RT-qPCR and
156 Western blot (**Figure S2c-e**). This suggests that both transgenes are required to
157 generate immortal MEFs, but once immortalized, p53DD becomes dispensable.

158 We next looked at the features of the transformed cells. All of the immortalized
159 MEFs proliferated faster than the wild-type MEFs (**Figure S2f**). We tested five lines
160 for the capability to form tumors in nude mice, Hras.E1A, SV40T, and Hras.Myc could
161 form tumors, whilst Bcl2.Myc and Hdac7SA.Mef2d could not (**Figure S2g, h**). Hoechst
162 staining of cross sections through the tumors indicated various differentiation layers,
163 although mainly mesoderm (**Figure S2i**). Aneuploidy is a common feature of cancer,
164 although the role of aneuploidy in transformation remains unclear [42]. A normal
165 karyotype is required for post-implantation embryonic development [43], but may not
166 be incompatible with pluripotency [44-46]. To rule out the impact of karyotype
167 abnormalities on reprogramming capability we confirmed a normal karyotype for four
168 of the lines that the subsequent study will mainly focus on (**Figure S2j**).
169

170 **Somatic cell reprogramming reveals a spectrum of reprogramming capability**

171 Using this panel of immortalized MEF cell lines we reprogrammed them using a
172 lentiviral OSKM system with vitamin C, which is known to promote reprogramming [36].
173 We used transformed lines derived from OG2 MEFs, which contain an Oct4-GFP
174 reporter [36]. Surprisingly, several lines could give rise to GFP+ colonies. We labeled
175 the lines as 'succeeding' (Hdac7SA.Mef2d, p53DD.Myc), 'struggling' (Bcl2.Myc,
176 Hdac7SA.E1A, p53DD.E1A, Hdac7SA.Myc), and 'failing' (Bcl2.E1A, Hras.E1A,
177 Hras.Myc, and SV40T) (**Figure 2a, b**). It should be noted that reprogramming the
178 transformed lines was less efficient than wild-type MEFs, except for the p53DD.Myc
179 line. Complete reprogramming requires a passage before the pluripotency gene
180 expression program can be stably established. This process partially relies on the fast-
181 dividing iPSCs outcompeting the slow-growing/senescent wildtype MEFs, however,

182 the transformed MEFs were also highly proliferative (**Figure S2f**) and would compete
183 with the reprogrammed iPSCs. Hence, we FACS sorted the GFP+ cells and passaged
184 them to establish iPSC lines in the absence of transformed MEFs. Pluripotency marker
185 genes were robustly expressed only in the sorted cells (**Figure 2c**), and we validated
186 the expression of these genes by RT-qPCR and Western blot (**Figure S3a, b**).
187 Comparison of the RNA-seq to a panel of pluripotency genes and somatic genes
188 indicated that the gene expression levels were close to ESCs (**Figure 2c**), and cross-
189 correlation and principal component analysis (PCA) indicated the cells were similar to
190 ESCs (**Figure 2d, e**). DPre was used to compare the gene expression in the GFP+
191 sorted lines to a panel of cell types [47, 48], and identified most lines as ESC-like
192 (**Figure S3c, d**). DPre also suggested that some ESC lines (e.g. Hdac7SA.Mef2d
193 GFP+ line #4) were a mixture of ESCs and contaminating immortalized MEFs (**Figure**
194 **S3d**). We confirmed that the iPSC lines could form teratomas with tissues representing
195 all three germ layers (**Figure S3e**). A mark of complete reprogramming is the silencing
196 of the OSKM transgenes, and the OSKM transgene was not detected in RNA-seq data
197 (**Figure S3f**). Interestingly, the immortalization factors were only partially silenced. All
198 were silenced except E1A in the Bcl2.E1A GFP+ line and Myc in the Bcl2.Myc GFP+
199 lines (**Figure S3f**). We confirmed Hdac7SA and Mef2d were silenced by qRT-PCR
200 (**Figure S3g**). Overall, these data confirmed the successful generation of iPSC-like
201 lines from some transformed tumorigenic normal karyotype MEF lines.
202

203 **Transformed MEFs have starting problems that impact reprogramming**

204 We next explored the properties that made some immortal MEFs reprogramming-
205 capable and others –incapable. We focused on two (non-exclusive) models: the

206 inability to reprogram occurs due to problems in the originating MEFs, or it is caused
207 by failures to traverse the correct sequence of events required for successful
208 reprogramming.

209 Immortalization of somatic adult cells occurs by a range of mechanisms and is
210 accompanied by changes in the epigenetic state and gene expression patterns. We
211 analyzed the gene expression of the starting MEF cells and observed a wide range of
212 gene expression changes, with some MEF lines adopting substantial changes in gene
213 expression, whilst other cell lines were less affected (**Figure 3a**). There was no simple
214 correlation between the number of gene changes and reprogramming capability
215 (**Figure 3a**). Despite a large number of gene expression changes, there was no
216 evidence that the MEFs were transdifferentiating as their global gene expression
217 patterns continued to correlate well against wildtype MEFs (**Figure 3b**), and DPre
218 continued to identify the cells as MEF-like (**Figure S4a**). Gene ontology of the DE
219 genes suggested changes were associated with basic cellular processes such as up-
220 regulation of metabolic processes, and extracellular matrix genes, and down-
221 regulation of ribosomes, cell cycle, and DNA repair processes (**Figure 3c, S4b**), which
222 is the opposite of the senescent-like phenotype up-regulated in human arrested
223 embryos [49]. Indeed, the MEF-senescent gene signature (defined in **Figure S1c**) was
224 not up-regulated in the transformed lines, and the majority of lines had a significantly
225 reduced MEF-senescent signature, reminiscent of the level in ESCs (**Figure 3d**). This
226 indicates the transformed lines are avoiding senescence. Interestingly, the
227 reprogramming-competent signature (defined in **Figure S1c**) declines in all lines
228 (except SV40T), which suggests a reduced capability to reprogram (**Figure 3d**).

229 Surprisingly, the patterns of the gene expression changes were uniform across
230 cell lines; genes up or down-regulated in one transformed line tended to be either
231 unchanged or similarly deregulated in other lines (**Figure S4c**). This implies a partially
232 shared transformation signature that defines a spectrum of severity measured by the
233 total number of differentially expressed genes.

234 We next looked at specific genes that have been identified as key blocks or
235 promoters of the reprogramming process. Three factors related to cancer and cellular
236 transformation that also modulate reprogramming in wild-type cells are *Tp53*, *Cdkn1a,*
237 and *Rb1* (Retinoblastoma) [40, 50-53]. There was little change in *Rb1* mRNA levels,
238 and whilst *Cdkn1a* was relatively high in all cell lines except p53DD.E1A, SV40T, and
239 wildtype MEFs, its expression was not correlated with reprogramming capability
240 (**Figure S5a**). For *Tp53*, its expression levels were relatively consistent across the
241 MEF lines and did not correlate with the reprogramming ability of the lines (**Figure**
242 **S5a**). We inferred p53 activity by looking at known target genes of p53, and all lines
243 showed unaltered p53 activity except for the MEF line containing the dominant-
244 negative p53DD (p53DD.E1A), which had significantly down-regulated p53 target
245 genes (**Figure S5b, c**). Loss of p53 is beneficial for reprogramming in wildtype cells
246 [40, 53, 54], yet in transformed lines reduced p53 activity in the p53DD.E1A line did
247 not correlate with efficient reprogramming (**Figure 2b**). This suggests that, in contrast
248 to wild-type cells, reduced p53 activity is not beneficial for reprogramming in
249 transformed cells.

250

251 **Immortalized MEFs have diverse defects in reprogramming phases**

252 Reprogramming occurs in defined phases [14, 15, 55], so we explored if there are also
253 changes in the phases of reprogramming in the transformed lines. We performed
254 RNA-seq during reprogramming in the early phase (D3, D6), the mid-phase (D9, D12),
255 and the late phase (D15). These time points roughly correspond to three waves of
256 gene expression labeled initiation, maturation, and stabilization [15, 55]. The same
257 waves were observed in our data, based on sets of genes specific to the three phases
258 (**Figure 4a, S6a**). Interestingly, different transformed lines completed different phases
259 of reprogramming and appeared to fail at distinct stages. Lines that can successfully
260 be reprogrammed: Hdac7SA.Mef2d, Bcl2.Myc and Hdac7SA.Myc, closely resembled
261 the wild-type MEF reprogramming process (**Figure 4a**). Hdacs7SA.E1A and
262 p53DD.E1A progressed through maturation but struggled to up-regulate stabilization
263 phase genes (**Figure 4a**), although both can ultimately generate GFP+ reprogrammed
264 cells at very low efficiency (**Figure 2b**). For the transformed lines that failed to
265 reprogram, they failed at different stages. All cell lines failed stabilization, SV40T and
266 Hras.E1A failed maturation, but completed initiation, whilst Hras.Myc failed all three
267 stages (**Figure 4a, S6**).

268 The mesenchymal-epithelial transition (MET) is a key part of the early stage of
269 reprogramming [55, 56]. We checked that the transformed MEFs remained
270 mesenchymal-like, and they still expressed typical markers of the mesenchyme
271 (**Figure 4b**). All transformed lines managed to navigate the MET, except for Hras.Myc
272 (**Figure S6b**). Indeed, some lines, including two transformed lines that would
273 ultimately fail to reprogram; Hras.E1A and SV40T, up-regulated epithelial genes in
274 advance of wild-type reprogramming cells (**Figure S6b**), including two transformed

275 lines that would ultimately fail to reprogram, Hras.E1A and SV40T. This suggests that
276 failure to execute a MET affected only the Hras.Myc line.

277 We next looked at changes in the overall phases of reprogramming in the early,
278 middle, and late stages of reprogramming. We measured the number of significantly
279 deregulated genes at each time point and used this as a proxy for the divergence from
280 a typical reprogramming time course. In this analysis, down-regulated genes
281 represent genes that fail to be correctly up-regulated at the specific day of the time
282 course, whilst up-regulated genes those represent those genes that are erroneously
283 high during reprogramming (**Figure 4c**). A large number of differentially regulated
284 genes on a specific day represents a divergence from the expected reprogramming
285 stage. As expected, the lines that reprogrammed at the highest efficiency (p53DD.Myc)
286 also had the lowest overall divergence (**Figure 4c**). Lines that could reprogram, but at
287 low efficiency (Hdac7SA.Mef2d, Bcl2.Myc, Hdac7SA.E1A, and Hdac7SA.Myc) tended
288 to have high initial divergence from the MEF state (day 0), but would correctly regulate
289 reprogramming-associated genes at the later time points. Finally, lines that fail to
290 reprogram (Hras.E1A and Hras.Myc), or reprogram exceptionally rarely (Bcl2.E1A),
291 tended to have high divergence at all time points. The exception was the SV40T line,
292 which appeared to follow the reprogramming gene expression program closely, but
293 would diverge at day 12. Interestingly, GO analysis of the down-regulated genes
294 (genes that should be up-regulated at that specific timepoint) highlighted the regulation
295 of DNA repair genes, which was defective in both Hras.E1A and SV40T cell lines
296 (**Figure 4d**). However, overall, there was a surprising diversity of changes in the
297 transformed MEF lines that give rise to a range of defects at specific time points of the

298 reprogramming process. This data indicates that transformation-specific effects drive
299 line-specific blocks on reprogramming in a stage-dependent manner.
300

301 **Chemical intervention can convert some lines from reprogramming-incapable**
302 **to reprogramming-capable**

303 The analysis above suggested several pathways that could be manipulated,
304 specifically, the MET, ribosome biogenesis, cell proliferation, cell adhesion, DNA
305 repair, and apoptosis. We wondered if it was possible to intervene in the
306 reprogramming process and convert the reprogramming-incapable to -capable. To
307 investigate this we assayed the effect on reprogramming of a range of inhibitors on
308 two reprogramming-capable lines (WT and Hdac7SA.Mef2d MEFs) compared to two
309 reprogramming-incapable lines (Hras.E1A and SV40T MEFs). In total, we used 25
310 inhibitors targeting a range of pathways (**Figure S7a**). Whilst most inhibitors had no
311 effects or would ablate reprogramming, we observed that 5 inhibitors promoted the
312 reprogramming of one or more of the transformed reprogramming-incapable lines
313 (**Figure S7a**). The five small molecules target: MEK1/2 (PD; PD0325901), GSK3
314 (CHIR; CHIR99021), ROCK (Y; Y23637), G9a (BIX; BIX-01294), and histone
315 deacetylases (TSA). We tested the inhibitors in combinations and found that the most
316 efficient combination was CHIR, PD and Y (**Figure 5a, S7b**). However, only PD was
317 essential to convert SV40T and Hras.E1A to reprogramming-capable, as
318 serum+Vc+PD alone resulted in a small number of GFP+ cells. Conversely, CHIR, Y,
319 and BIX alone were incapable of converting Hras.E1A MEFs to reprogramming-
320 capable (**Figure S7b**). Interestingly the inhibitor cocktails had effects on
321 reprogramming in a transformed line-specific manner, as the addition of PD to the

322 reprogramming cocktail converted the normally reprogramming-capable Bcl2.E1A line
323 to incapable. This highlights the line-specific nature of the reprogramming barriers and
324 shows that overcoming a barrier in one line can potentially initiate a barrier in another
325 transformed line.

326 We confirmed the resulting iPSC-like cells derived from Hras.E1A or SV40T
327 MEFs reprogrammed with the PD, CHIR, Y and BIX adopted a normal morphology,
328 expressed pluripotent marker genes by qRT-PCR and Western blot (**Figure 5d, and**
329 **Figure S7c**. We sorted the GFP+ cells and established iPSC lines, and measured
330 their gene expression which closely correlated with ESCs by correlation and PCA
331 (**Figure 5e and Figure S8a, b**). It should be emphasized that the chemical cocktail
332 did not improve (or inhibit) wildtype MEF reprogramming, as wildtype cells
333 reprogrammed with similar rates in serum+Vc or the inhibitor cocktail (**Figure S7b**).
334 This emphasizes that the blocks seen in transformed cells are specific to
335 transformation and are not present in wild-type untransformed cells. Overall, the
336 results indicate that transformation-specific pathways are impairing the ability to
337 reprogram.

338

339 **Epigenetic defects underly the inability to reprogram**

340 We next focused on three specific cell lines: Hdac7SA.Mef2d, Hras.Myc and Hras.E1A.
341 Hdac7SA.Mef2d could reprogram in serum under normal conditions, whilst Hras.E1A
342 is initially resistant to reprogramming (**Figure 2b**), but the addition of PD converts it to
343 reprogramming-capable (**Figure 5a-d**). Hras.Myc, conversely, could not be converted
344 to reprogramming-capable with any of the conditions we tried (**Figure 5b**).

We explored the epigenetic regulation of the transformed MEFs. As a proxy for overall epigenetic activity, we looked at the expression levels of epigenetic factors involved in activation, repression, and the reading of epigenetic marks, as defined by the Epifactors database [57]. As expected, epigenetic-related factors were uniformly significantly up-regulated in ESCs compared to MEFs (**Figure S9a**), reflecting their more complex epigenetic regulation [58]. For the transformed MEFs, activators were more often significantly up-regulated, compared to untransformed MEFs (**Figure S9a**). Erasers and readers were only up-regulated in several transformed lines, all containing E1A: Hras.E1A and Bcl2.E1A and p53DD.E1A (**Figure S9a**). Nonetheless, whilst the different classes of epigenetic regulators varied across the transformed lines and did not discriminate reprogramming-capable from incapable, the general pattern for the majority of lines was increased expression of epigenetic regulators. Western blot of repressive histone modifications (e.g. H3K27me3, H3K9me3) tended to stay the same in the different lines (**Figure 6a**). However, histone acetylation inversely correlated with reprogramming capability as H3K27ac, H3ac or H4ac was high in reprogramming-incapable lines (Hras.Myc, Hras.E1A and SV40T). The results support epigenetic deregulation as a feature of the transformed lines and suggest increased epigenetic activation through histone acetylation is a feature of transformed reprogramming-incapable MEFs.

Transformed MEFs have a hyperactive chromatin state which MEK-inhibition partially reverts to wildtype

To explore changes in the epigenetic state of the MEFs we performed ATAC-seq on three transformed lines: Hdac7SA.Mef2d (reprogramming-capable), Hras.E1A

369 (initially incapable, but can be induced to reprogram with MEK inhibition), and
370 Hras.Myc (reprograming-incapable) and compared them to wildtype MEFs. Heatmaps
371 of open and closed chromatin indicated, surprisingly, that the Hdac7SA.Mef2d lines
372 were distal from all of the other transformed cell lines and had the greatest number of
373 specific peaks (**Figure S9b, Figure 6b**). The loci that were opened or closed in the
374 Hdac7SA.Mef2d lines were significantly correlated with matching changes in gene
375 expression (**Figure S9c**). Intriguingly, the overall ATAC signal was substantially higher
376 in all three transformed cell lines compared to wildtype MEFs (**Figure 6c**), and this
377 effect was despite fewer open loci in the Hras.Myc and Hras.E1A lines (**Figure 6b**).
378 This suggests that the transformed MEFs have increased levels of open chromatin.

379 We next focused on Hras.Myc and Hras.E1A lines, and removed the
380 Hdac7SA.Mef2d from further analysis as it is capable of reprogramming under normal
381 conditions and its ATAC-seq genome-wide pattern was distal from the other
382 transformed cell lines and wildtype MEFs. Hras.Myc and Hras.E1A are interesting as
383 they have relatively similar ATAC-seq patterns (**Figure S9b**), yet Hras.E1A can be
384 converted to reprogramming-capable, whilst Hras.Myc cannot. The two lines have
385 Hras, which does not impede reprogramming independently. Similarly, other lines co-
386 transformed with Myc are also reprogramming-capable (Bcl2.Myc and Hdac7SA.Myc),
387 as E1A transformed lines (Hdac7SA.E1A). Hence, the individual transgenes are not
388 incompatible with reprogramming, only the specific combinations of Hras.E1A and
389 Hras.Myc is important. One key difference between Hras.E1A and Hras.Myc is that
390 only Hras.E1A up-regulated initiation-stage genes and underwent an MET (**Figure 4a,**
391 **d**), suggesting the impairment of reprogramming is already primed in the starting
392 transformed MEFs.

393 To explore the underlying differences, we performed ATAC-seq on Hras.Myc
394 and Hras.E1A lines with and without PD treatment, and compared them to wildtype
395 MEFs. Interestingly, the global upregulation of the ATAC-seq signal at open loci was
396 lost in the PD-treated samples and returned to levels similar to the wild-type (**Figure**
397 **6d**). Western blot of H3K27ac, a mark of open active chromatin, supported this
398 observation as H3K27ac was higher in the transformed MEFs, and treatment with PD
399 for 3 days reduced H3K27ac to wildtype levels (**Figure 6e**). Interestingly the result is
400 transient, as treatment for 2 days followed by washing out reverted H3K27ac levels to
401 the hyperactive state (**Figure 6e**). This suggests that when PD inhibits MEK it leads
402 to the indirect closing of chromatin and reductions in the global levels of H3K27ac.
403 However, this effect was transient, and incomplete as the overall epigenetic landscape
404 did not revert to a wildtype state and 34,913 loci open in wildtype MEFs remained
405 closed in the transformed cell lines (**Figure 6f**).

406 To explore the mechanisms underlying the action of PD, we next divided the
407 loci into two types based upon their state in wildtype MEFs and the effect of PD: (1)
408 those loci that PD opens and reverts to a wildtype state, and (2) those loci that PD
409 closes, reverting to the wildtype state (**Figure 6g**). There was a roughly equal split
410 between the two classes (10,230 peaks versus 12,276 peaks; **Figure 6g**). GO analysis
411 of the peaks that were specifically reverted to wildtype state in the Hras.Myc cells
412 suggested that cell-cell adhesion gene loci were being remodeled (**Figure 6h**). Indeed,
413 chromatin at *Cdh1* and *Epcam* (two epithelial genes) were open in the Hras.E1A and
414 Hras.Myc MEFs, but were closed when the cells were treated with PD and reverted to
415 a wild-type MEF pattern at these genes (**Figure 6i**). Importantly, *Cdh1*, *Epcam,* and
416 other epithelial genes are not expressed in the transformed MEFs (**Figure 6j**),

417 indicating that epigenetic regulation at epithelial gene loci is deleterious for
418 reprogramming of transformed cell lines.

419 We next looked at possible downstream targets of MEK signaling that mediated
420 this indirect effect on chromatin. We used motif analysis to look for changes in
421 transcription factor occupancy at PD-modulated peaks (**Figure 6k**). Interestingly,
422 peaks that were closed in response to PD treatment were specifically enriched with
423 ETS, E2A, and KLF motifs. KLF proteins are major regulators of MET processes and
424 are pluripotent factors, whilst ETS factors have been identified as barriers for wild-type
425 untransformed reprogramming [59]. Loci that PD treatment opens were enriched with
426 REL, HOX, and TFE family TFs (**Figure 6k**). To explore deeper we looked at two small
427 groups of chromatin loci that distinguished reprogramming-capable from -incapable,
428 10001 (specifically open in WT and Hras.E1A+PD) and 01110 (specifically closed in
429 WT and Hras.E1A+PD) (**Figure 6g**). Interestingly, GO analysis of the 01110-group
430 suggested that enhancers near to genes related to cell differentiation were
431 erroneously opened (**Figure S10a**), for example at the developmental-related genes
432 *Isl1* and *Tcf7l2* (**Figure S10b**). This was a purely an effect on chromatin state as *Isl1*
433 and *Tcf7l2* were below the threshold for detection in the RNA-seq data. This pattern
434 extended to all of the genes in the vicinity of the 01110-chromatin state which were
435 not significantly up or down-regulated (**Figure S10c**). Conversely the genes
436 associated with the 10001 group were significantly down-regulated (**Figure S10c**),
437 although GO analysis suggests they are related to basic cellular processes such as
438 cytoskeleton organization (**Figure S10d**). Overall, this suggests the epigenetic
439 activation of developmental genes is deleterious for reprogramming, and MEK
440 inhibition can indirectly modulate the epigenetic state to close chromatin.

441

442 **Discussion**

443 Reprogramming cancerous/transformed cells to pluripotent cells, followed by their
444 differentiation, could form a valuable experimental model to recapitulate the earliest
445 stages of tumorigenesis [60, 61]. However, reprogramming of primary cancerous
446 tissues is often difficult or impossible for many cancers, and even in possible cases
447 the efficiency is considerably less than wild-type cells. Additionally, there remains
448 doubts about the *bona fide* reprogramming of the cancerous cells [28, 34, 35], and it
449 is not clear if reprogramming of primary human cells is complete. A significant problem
450 is the heterogeneity of primary cancer cells. To overcome this problem, we focused
451 on isogenic transformed MEF lines that allow us to control overall heterogeneity.
452 Although, even in this system there was a surprisingly large variation between the
453 different methods of MEF transformation.

454 There is an intimate but unclear link between reprogramming and
455 transformation [1, 61]. This link was recently explored in a novel system with multiple
456 inducible genes that could be manipulated to either transform MEFs or perform
457 reprogramming to iPSCs [17]. This elegant study showed that reprogramming and
458 transformation followed similar initial paths, but then later diverged. The system they
459 used for transformation was K-ras^{G12D} and *Myc* overexpression [17], which is similar
460 to the Hras.Myc (*Hras*^{G12V}/*Myc*) line described in this manuscript. Intriguingly, we show
461 that Hras.Myc is not compatible with reprogramming, and those cells fail very early in
462 the reprogramming process. However, several other lines described here can be
463 transformed and remain compatible with reprogramming. Potentially, transformation
464 strategies that remain compatible with reprogramming may recapitulate more of the

465 reprogramming process as they transform, and only diverge from reprogramming at
466 later stages.

467 *Ras* overexpression is beneficial for reprogramming wildtype MEFs, but is a
468 major barrier to reprogramming transformed cells [62]. Indeed, when $Hras^{G12V}$ was
469 overexpressed in p53-null immortalized MEFs reprogramming was blocked [62]. This
470 is reminiscent of the effect we observed in the Hras.Myc and Hras.E1A lines, which
471 are also resistant to reprogramming. However, we show that inhibition of MEK by PD
472 can only enable Hras.E1A cells to reprogram, and Hras.Myc cells remain
473 reprogramming-incapable. This indicates that $Hras^{G12V}$ is context-dependent when its
474 co-transformation factor is different.

475 Several studies have identified small molecules that assist in reprogramming
476 transformed cells to a more pluripotent-like state. mTOR was also identified as a
477 barrier to the reprogramming of sarcoma cells [63]. In our experiments, mTOR
478 inhibition had no effect, and ROCK inhibition worked only to boost reprogramming in
479 cooperation with MEK inhibition. This again highlights the context-specific nature of
480 the barriers in different transformed cell lines. Overall, MEK inhibition was the
481 dominant factor, which agrees with a previous study that showed that MEK inhibition
482 assisted in reprogramming human transformed cell lines [64]. PD is also a key
483 component of the 2iLIF medium used to maintain mouse ESCs in a naïve groundstate
484 [49, 65]. It is unclear if PD plays the same role in assisting in reprogramming
485 transformed MEFs and in promoting pluripotency. Indeed, reprogramming with PD is
486 deleterious for reprogramming wild-type MEFs, at least in the early stages of
487 reprogramming [62, 66].

488 Ultimately, these studies support a model for the context-specific
489 reprogramming of transformed cell lines. Different transformations lead to diverse
490 blocks at different stages of the reprogramming process. Inhibition of ERK by PD can
491 overcome some of these barriers, primarily in Ras-transformed cells [62, 64]. However,
492 each transformed cell line can harbor defects that block reprogramming at unique
493 stages. This perhaps reflects the diversity of regulatory features disrupted in cancer
494 and may relate to the 'stemness' of a cancer type or method of transformation [67].
495 Whilst there are many surprising similarities between tumorigenesis and
496 reprogramming, the nature of the relationship remains unclear [1]. Potentially, insights
497 into reprogramming can inform the process of transformation and provide new models
498 to address tumorigenesis and a clearer understanding of cell type plasticity in cancer
499 development.
500

501 **METHODS SUMMARY**

502 **Animal experimental approval**

503 Ethical approval for the generation of MEFs and teratoma assays was granted by the
504 Southern University of Science and Technology animal ethical committee, approval
505 number: SUSTC-2019-005.
506

507 **Cell culture and MEF transformation**

508 MEFs were derived from E13.5 embryos from OG2 mice [36], or a wild-type ICR
509 genetic background. MEF cells were cultured in DMEM high-glucose media containing
510 10% FBS (GIBCO), 1 × GlutaMAX (GIBCO), and 1x nonessential amino acids
511 (GIBCO), 0.5x Penicillin/Streptomycin (Hyclone). We generated immortalized MEF

512 cell lines by transfecting the following transgenes singly or in combination: P53DD,
513 PMM2-SV40T, PMM2-Hdac7SA, PMM2-Mef2d, Bcl2, PMM2-Myc, Ras, PMM2-E1A.
514 Immortalized MEFs were passaged for at least a month to remove wild-type
515 untransformed MEFs.
516

517 **Somatic cell reprogramming**

518 Wildtype MEFs and transformed MEFs in an OG2 reporter background were
519 reprogrammed as described in [36]. Briefly, OKSM in lentivirus was transfected into
520 15,000 MEFs in one well of a 12-well plate. One day after transfection the medium
521 was changed to reprogramming medium: DMEM high-glucose media containing 15%
522 FBS (GIBCO), 1× GlutaMAX (GIBCO), 1x non-essential amino acids (GIBCO), 0.5x
523 Penicillin/Streptomycin (Hyclone), 1mM sodium pyruvate (GIBCO), 0.1 mM 2-
524 mercaptoethanol (GIBCO), 1000 U/ml leukemia inhibitory factor (LIF) (Millipore) and
525 vitamin C (50 μg/ml, Sigma), as indicated in the figures. The medium was changed
526 daily. Inhibitors were added when the medium was changed to reprogramming
527 medium: PD0325901 (1 μM, Sigma), CHIR99021 (3 μM, Sigma), Y23637 (10 μM,
528 Sigma), BIX-01294 (200 nM, Selleck), and TSA (50 nM, Sigma).
529

530 **Karyotype analysis**

531 When the immortalized MEF cell growth density reached 80 to 90%, colchicine (0.2
532 μg/ml, Selleck), was added to the culture medium to a final concentration of 0.2 μg/ml.
533 The cells were incubated at 37°C for 120 min. After colchicine treatment, cells were
534 washed twice with PBS, and 0.5 ml of 0.25% trypsin was added for digestion. Then, 7
535 ml of 0.075 mol/L KCl solution preheated to 37°C was added, and the cell suspension

536 was blown with a straw and incubated at 37°C in a water bath for 25 mins. Collected
537 cells were fixed with a fixing solution (methanol glacial acetic acid 3:1 preparation) at
538 37°C, 3 min. After fixation, cells were centrifuged at 1200 revolutions per min (rpm) for
539 5 min, after which the supernatant was discarded. 7 ml of fresh fixing solution was
540 added, and the cells were gently beaten with a pipette tip, and fixed in a 37°C water
541 bath for 40 mins. Cells were centrifuged at 1200 rpm for 5 min, after which the
542 supernatant was discarded. Subsequently, 7 ml of fresh fixing solution was added,
543 followed by gentle beating with a filter tip. The cells were fixed in a 37°C water bath
544 for 40 min. After fixation the cells were centrifuged at 1200 rpm for 5 min, and then
545 most of the fixation solution was removed, and the cells were resuspended with part
546 of the fixation solution. Suspended cells were pipetted onto a slide from a distance of
547 about 30 cm. Immediately after dropping the cells on the slide, move the slide into the
548 oven at 75°C and bake for 3h. Next, 0.03 g trypsin powder to 55 ml normal saline was
549 added, followed by a gentle shaking, before adjusting the pH to about 7.2 with 3%
550 Tris-HCl. The preparation was put into trypsin digestion solution for 8-10 seconds,
551 followed by the addition of saline to stop digestion, then add Giemsa staining (C0133,
552 Beyotime Biotech) solution was added for 5-10 minutes, and then pick out the glass
553 slide with tweezers. Both sides were gently rinsed with running water and allowed to
554 dry at room temperature. After the slides dried, they were examined under a
555 microscope to look for a good chromosome split.
556

557 **Teratoma formation**

558 5-10 million mouse PSCs in a slurry of Matrigel and mTeSR (1:1) medium were
559 injected into 8-week-old female nude (BALB/cNj-Foxn1^{nu}/Gpt) immunodeficient mice.

560 Teratoma growth was quantified by measuring approximate elliptical area (mm^2) with
561 calipers measuring the outward width and height after growth for 60 days.
562 Representative tumors were dissected and sectioned and slices were stained with
563 hematoxylin and eosin.
564

565 **Alkaline phosphatase staining**

566 Reprogrammed cells were stained with alkaline phosphatase according to the
567 manufacturer's protocols. In brief, reprogramming cells were fixed in 1% (w/v)
568 formaldehyde, and then cells were stained with BCIP/NBT Alkaline Phosphatase Color
569 Development Kit (C3206, Beyotime Biotech) according to the kit's instructions.
570

571 **Flow cytometry**

572 Reprogramming cells were digested with trypsin and washed with DPBS once, and
573 analyzed or sorted with BD FACS Aria III flow cytometer. To monitor cellular
574 reprogramming status, reprogrammed cells were stained with antibodies: anti-CD44-
575 PE (25-0441-82, 1:500, Thermo Fisher) and anti-ICAM-1-FITC (sc-8439 FITC, 1:200,
576 Santa Cruz Biotechnology). Cells were sorted and analyzed on a BD FACS Aria III
577 instrument.
578

579 **Western blot**

580 Cells were lysed with RIPA buffer (P0013B, Beyotime Biotech). Proteins were
581 resolved in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
582 and transferred onto pre-activated polyvinylidene fluoride (PVDF) membranes
583 (IPVH00010, Millipore, MA, USA). The PVDF Membranes were incubated with anti-

584 H3 (ab1791, 1:3000, Abcam), anti-H3K27ac (ab4279, 1:2000, Abcam), anti-H3ac (06-
585 599,1:2000, Millipore), anti-H4ac (ab46983, 1:2000, Abcam), anti-H3K4me3 (ab8580,
586 1:2000, Abcam), anti-H3K4me (ab176877,1:2000, Abcam), anti-H3K27me3 (07-
587 449,1:2000, Millipore), anti-H3K9me3 (ab8898, 1:2000, Abcam), anti-NANOG (8822S,
588 1:1000, cell signaling), anti-OCT4 (sc-5279, 1:1000, Santa Cruz Biotechnology), anti-
589 P53 (sc-126, 1:1000, Santa Cruz Biotechnology), anti-HDAC7 (ab12174,1:2000,
590 Abcam), anti-MEF2D (ab32845,1:2000, Abcam), anti-ACTIN (ab8227, 1:3000,
591 Abcam), anti-MYC (ab32074,1:2000, Abcam), anti-SOX2 (4900S, 1:1000, cell
592 signaling). Afterward, the membranes were incubated with HRP-conjugated goat
593 antirabbit IgG (ab205718, 1:3000, Abcam) and visualized using an enhanced
594 chemiluminescence BeyoECL method (P0018AS, Beyotime Biotech).
595
596 **Immunofluorescence**
597 Cells were washed with cold PBS and fixed using 4% (w/v) paraformaldehyde in PBS
598 for 10 minutes at room temperature. Then cells were washed with PBS and
599 permeabilized with 0.25% (v/v) Triton X-100 for 10 minutes at room temperature. Fixed
600 cells were then washed with PBS three times and blocked with 3% (w/v) bovine serum
601 albumin for 1 hour at room temperature. Then cells were incubated with primary
602 antibodies: anti-NANOG (8822S, 1:1000, cell signaling), at 4°C overnight in primary
603 antibody. Following overnight incubation, cells were washed with wash buffer (0.1%
604 (v/v) Tween-20 in PBS) and incubated with secondary antibodies Alexa Fluor series
605 (Life Technologies) in wash buffer for 1 hour at room temperature. Cells were washed
606 again and nuclei were stained using DAPI (5 μg/ml, Thermo).
607

608 **RT-qPCR**

609 Total RNA from cells was isolated using RNAzol RT (MRC, RN190) according to the
610 manufacturer's protocols. cDNA synesis by using a PrimeScript RT Master Mix
611 (Takara, RR036A). Real-time PCR was performed in triplicate using SYBR Premix Ex
612 Taq (Takara, RR820A) and using a Biorad Real-time PCR system. The primers used
613 are listed in **Supplementary Table S2**.

614

615 **RNA-seq preparation and analysis**

616 RNA was isolated using RNAzol RT (MRC, RN190) according to the manufacturer's
617 protocol and prepared for sequencing with RNA-seq NEB Next Ultra RNA Library Prep
618 Kit (NEB, #7530). Samples were sequenced on an Illumina Novaseq 6000.

619 RNA-seq was analyzed essentially as described in [47], except scTE/te_counts
620 was used to assign mapped reads to genes and TEs [68], and the UCSC genome
621 browser repeatmask track and GENCODE vM23 annotation were used for TE and
622 gene assignment. Differential expression was determined using DESeq2 [69]. A gene
623 was considered differentially expressed had an absolute fold-change of at least 2, and
624 a Bonferroni-Hochberg corrected p-value of 0.01. GSEA was performed using fgsea
625 [70], and GO was done with goseq [71]. Cell type determination was scored with DPre
626 [48]. Other analyses were performed using glbase3 [72].

627

628 **ATAC-seq preparation and analysis**

629 ATAC-seq library was generated using the Tn5 enzyme from the
630 TruePrep DNA Library Prep Kit V2 (Vazyme, TD501-02) as previously described [73].
631 Briefly, a total of ~50,000 cells were washed once with 50 µl of cold PBS and

632 resuspended in 50 µl lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2,
633 0.2% (v/v) IGEPAL CA-630). The nuclei were centrifuged for 10 min at 500 × g at 4 °C,
634 followed by the addition of 50 µl transposition reaction mix (25 µl TD buffer, 2.5 µl Tn5
635 transposase (Vazyme), and 22.5 µl nuclease-free water). Samples were PCR-
636 amplified and purified using a MinElute kit (Qiagen). After selecting an appropriate
637 PCR cycle number (See [73]) samples were sequenced on an Illumina sequencer.
638 ATAC-seq data was analyzed essentially as described in reference: [59]. Briefly, reads
639 were aligned to the mm10 mouse genome with bowtie2 [74], peaks were called with
640 MACS2 [75], and then the *redefine_peaks* function, which is a generalized
641 reimplementation of the algorithm in [59], was used to recover low-scoring peaks by
642 sharing peak information across samples [76]. DNA binding motifs were detected
643 using HOMER [77]. All other analyses were performed using glbase3 [72]. ATAC-seq
644 data from GSE93029 [59] and GSE103980 [78] were reanalyzed as part of this study.
645

646 **Statistical analysis and biological replication**

647 Differential gene expression was calculated using DESeq2 (v1.36.0). A gene was
648 considered significantly differentially regulated if it had an absolute fold-change of at
649 least 2 and a Bonferroni-Hochberg corrected p-value (q-value) of <0.01. This criterion
650 was used in **Figures 3a, 4c, Figures S1c, and d**. RNA-seq experiments were
651 performed in at least biological duplicate (different samples on different days, or
652 independent cell lines). Gene ontology analysis was performed using goseq (v1.48.0)
653 and statistics were calculated using goseq's internal statistical model. A gene ontology
654 category was considered significantly enriched if there were at least 50 genes in that
655 GO term and a Bonferroni-Hochberg corrected p-value (q-value) of <0.01. GSEA was

656 performed using fgsea (v1.22.0). Gene sets were considered enriched or depleted if
657 they had an absolute NES (normalized enrichment score) of at least 1.5 and a
658 Bonferroni-Hochberg corrected p-value (q-value) of <0.01. Western blots were
659 repeated at least twice with similar results, with the exception of **Figure S3b** and **S7c**
660 which were performed once. **Figure 6a** was repeated three times. FACS analysis
661 (**Figure 1c**) was performed twice. qRT-PCR was performed on at least three biological
662 replicates with three technical replicates each. Transcription factor motif analysis was
663 performed using HOMER. A motif was considered significantly enriched if the
664 uncorrected p-value was <0.00001.

665

666 **Conflict of Interest**

667 The authors declare no conflict of interest.

668

669 **Acknowledgments**

670 This work was supported by the National Natural Science Foundation of China
671 (32150710521), and the Shenzhen Innovation Committee of Science and Technology
672 (ZDSYS20200811144002008). Additional supported was rendered by the Center for
673 Computational Science and Engineering of Southern University of Science and
674 Technology.

675

676 **Author contributions**

677 X.F. designed and performed the majority of the experiments, prepared the manuscript
678 figures and revised the text. Q.Z. designed the project and planned the experiments.
679 G.M., H.H., Y.L., J.C., Z.X., B.D., S.L. performed some experiments or assisted in

experimental procedures. I.A.B., L.S., assisted with the bioinformatic analysis. R.J. assisted in manuscript writing, acquired funding and supervised H.H.. A.P.H. performed some of the bioinformatic analysis, drafted the manuscript. All authors read and revised the manuscript.

Author information

RNA-seq data are available in the Gene expression omnibus (GEO) public database under accession number GSE213225.

References

1. Iglesias JM, Gumuzio J, Martin AG: **Linking Pluripotency Reprogramming and Cancer**. *Stem Cells Transl Med* 2017, **6**(2):335-339.

2. Sun L, Fu X, Ma G, Hutchins AP: **Chromatin and Epigenetic Rearrangements in Embryonic Stem Cell Fate Transitions**. *Front Cell Dev Biol* 2021, **9**:637309.

3. Folmes CD, Nelson TJ, Martinez-Fernandez A, Arrell DK, Lindor JZ, Dzeja PP, Ikeda Y, Perez-Terzic C, Terzic A: **Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming**. *Cell Metab* 2011, **14**(2):264-271.

4. Vander Heiden MG, Cantley LC, Thompson CB: **Understanding the Warburg effect: the metabolic requirements of cell proliferation**. *Science* 2009, **324**(5930):1029-1033.

5. Li Q, Hutchins AP, Chen Y, Li S, Shan Y, Liao B, Zheng D, Shi X, Li Y, Chan WY *et al*: **A sequential EMT-MET mechanism drives the differentiation of human embryonic stem cells towards hepatocytes**. *Nat Commun* 2017, **8**:15166.

6. Blum B, Benvenisty N: **The tumorigenicity of human embryonic stem cells**. *Adv Cancer Res* 2008, **100**:133-158.

7. Niu N, Mercado-Uribe I, Liu J: **Dedifferentiation into blastomere-like cancer stem cells via formation of polyploid giant cancer cells**. *Oncogene* 2017, **36**(34):4887-4900.

8. Hepburn AC, Steele RE, Veeratterapillay R, Wilson L, Kounatidou EE, Barnard A, Berry P, Cassidy JR, Moad M, El-Sherif A *et al*: **The induction of core pluripotency master regulators in cancers defines poor clinical outcomes and treatment resistance**. *Oncogene* 2019, **38**(22):4412-4424.

9. Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM, Zhang D, Li D, Xing X, Kim S *et al*: **Transposable elements drive widespread expression of oncogenes in human cancers**. *Nat Genet* 2019, **51**(4):611-617.

10. Hochedlinger K, Yamada Y, Beard C, Jaenisch R: **Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues**. *Cell* 2005, **121**(3):465-477.

11. Zhang J, Espinoza LA, Kinders RJ, Lawrence SM, Pfister TD, Zhou M, Veenstra TD, Thorgeirsson SS, Jessup JM: **NANOG modulates stemness in human colorectal cancer**. *Oncogene* 2013, **32**(37):4397-4405.

12. Jeter CR, Liu B, Liu X, Chen X, Liu C, Calhoun-Davis T, Repass J, Zaehres H, Shen JJ, Tang DG: **NANOG promotes cancer stem cell characteristics and prostate cancer resistance to androgen deprivation**. *Oncogene* 2011, **30**(36):3833-3845.

13. Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors**. *Cell* 2006, **126**(4):663-676.

14. Xu Y, Zhang M, Li W, Zhu X, Bao X, Qin B, Hutchins AP, Esteban MA: **Transcriptional Control of Somatic Cell Reprogramming**. *Trends Cell Biol* 2016, **26**(4):272-288.

15. Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J *et al*: **A molecular roadmap of reprogramming somatic cells into iPS cells**. *Cell* 2012, **151**(7):1617-1632.

16. Hussein SM, Puri MC, Tonge PD, Benevento M, Corso AJ, Clancy JL, Mosbergen R, Li M, Lee DS, Cloonan N *et al*: **Genome-wide characterization of the routes to pluripotency**. *Nature* 2014, **516**(7530):198-206.

17. Huyghe A, Furlan G, Schroeder J, Cascales E, Trajkova A, Ruel M, Studer F, Larcombe M, Yang Sun YB, Mugnier F *et al*: **Comparative roadmaps of reprogramming and oncogenic transformation identify Bcl11b and Atoh8 as broad regulators of cellular plasticity**. *Nat Cell Biol* 2022.

18. Abad M, Mosteiro L, Pantoja C, Canamero M, Rayon T, Ors I, Grana O, Megias D, Dominguez O, Martinez D *et al*: **Reprogramming in vivo produces teratomas and iPS cells with totipotency features**. *Nature* 2013, **502**(7471):340-345.

19. Ohnishi K, Semi K, Yamamoto T, Shimizu M, Tanaka A, Mitsunaga K, Okita K, Osafune K, Arioka Y, Maeda T *et al*: **Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation**. *Cell* 2014, **156**(4):663-677.

20. Srivastava Y, Tan DS, Malik V, Weng M, Javed A, Cojocaru V, Wu G, Veerapandian V, Cheung LWT, Jauch R: **Cancer-associated missense mutations enhance the pluripotency reprogramming activity of OCT4 and SOX17**. *FEBS J* 2020, **287**(1):122-144.

21. Miyoshi N, Ishii H, Nagai K, Hoshino H, Mimori K, Tanaka F, Nagano H, Sekimoto M, Doki Y, Mori M: **Defined factors induce reprogramming of gastrointestinal cancer cells**. *Proc Natl Acad Sci U S A* 2010, **107**(1):40-45.

22. Kim HJ, Jeong J, Park S, Jin YW, Lee SS, Lee SB, Choi D: **Establishment of Hepatocellular Cancer Induced Pluripotent Stem Cells Using a Reprogramming Technique**. *Gut Liver* 2017, **11**(2):261-269.

23. Zhang X, Cruz FD, Terry M, Remotti F, Matushansky I: **Terminal differentiation and loss of tumorigenicity of human cancers via pluripotency-based reprogramming**. *Oncogene* 2013, **32**(18):2249-2260, 2260 e2241-2221.

24. Carette JE, Pruszak J, Varadarajan M, Blomen VA, Gokhale S, Camargo FD, Wernig M, Jaenisch R, Brummelkamp TR: **Generation of iPSCs from cultured human malignant cells**. *Blood* 2010, **115**(20):4039-4042.

757 25. Li T, Zhang Y, Li Y, Wang X, Bao W, Huang J, Ma Y, Li S, Wang S, Yang Y *et al*: **Modeling**
758 **leukemia with pediatric acute leukemia patient-derived iPSCs**. *Stem Cell Res* 2021,
759 **54**:102404.

760 26. Dannenmann B, Klimiankou M, Oswald B, Solovyeva A, Mardan J, Nasri M, Ritter M, Zahabi
761 A, Arreba-Tutusaus P, Mir P *et al*: **iPSC modeling of stage-specific leukemogenesis reveals**
762 **BAALC as a key oncogene in severe congenital neutropenia**. *Cell Stem Cell* 2021,
763 **28**(5):906-922 e906.

764 27. Liu Y, Cheng H, Gao S, Lu X, He F, Hu L, Hou D, Zou Z, Li Y, Zhang H *et al*: **Reprogramming**
765 **of MLL-AF9 leukemia cells into pluripotent stem cells**. *Leukemia* 2014, **28**(5):1071-1080.

766 28. Kumar S, Curran JE, Glahn DC, Blangero J: **Utility of Lymphoblastoid Cell Lines for**
767 **Induced Pluripotent Stem Cell Generation**. *Stem Cells Int* 2016, **2016**:2349261.

768 29. Barrett R, Ornelas L, Yeager N, Mandefro B, Sahabian A, Lenaeus L, Targan SR, Svendsen
769 CN, Sareen D: **Reliable generation of induced pluripotent stem cells from human**
770 **lymphoblastoid cell lines**. *Stem Cells Transl Med* 2014, **3**(12):1429-1434.

771 30. Bedel A, Pasquet JM, Lippert E, Taillepierre M, Lagarde V, Dabernat S, Dubus P, Charaf L,
772 Beliveau F, de Verneuil H *et al*: **Variable behavior of iPSCs derived from CML patients for**
773 **response to TKI and hematopoietic differentiation**. *PLoS One* 2013, **8**(8):e71596.

774 31. Hu K, Yu J, Suknuntha K, Tian S, Montgomery K, Choi KD, Stewart R, Thomson JA, Slukvin,
775 II: **Efficient generation of transgene-free induced pluripotent stem cells from normal and**
776 **neoplastic bone marrow and cord blood mononuclear cells**. *Blood* 2011, **117**(14):e109-
777 119.

778 32. Kumano K, Arai S, Hosoi M, Taoka K, Takayama N, Otsu M, Nagae G, Ueda K, Nakazaki K,
779 Kamikubo Y *et al*: **Generation of induced pluripotent stem cells from primary chronic**
780 **myelogenous leukemia patient samples**. *Blood* 2012, **119**(26):6234-6242.

781 33. Bang JS, Choi NY, Lee M, Ko K, Park YS, Ko K: **Reprogramming of Cancer Cells into**
782 **Induced Pluripotent Stem Cells Questioned**. *Int J Stem Cells* 2019, **12**(3):430-439.

783 34. Stricker SH, Feber A, Engstrom PG, Caren H, Kurian KM, Takashima Y, Watts C, Way M, Dirks
784 P, Bertone P *et al*: **Widespread resetting of DNA methylation in glioblastoma-initiating**
785 **cells suppresses malignant cellular behavior in a lineage-dependent manner**. *Genes Dev*
786 2013, **27**(6):654-669.

787 35. Lee JH, Salci KR, Reid JC, Orlando L, Tanasijevic B, Shapovalova Z, Bhatia M: **Brief Report:**
788 **Human Acute Myeloid Leukemia Reprogramming to Pluripotency Is a Rare Event and**
789 **Selects for Patient Hematopoietic Cells Devoid of Leukemic Mutations**. *Stem Cells* 2017,
790 **35**(9):2095-2102.

791 36. Zhuang Q, Li W, Benda C, Huang Z, Ahmed T, Liu P, Guo X, Ibanez DP, Luo Z, Zhang M *et al*:
792 **NCoR/SMRT co-repressors cooperate with c-MYC to create an epigenetic barrier to**
793 **somatic cell reprogramming**. *Nat Cell Biol* 2018, **20**(4):400-412.

794 37. O'Malley J, Skylaki S, Iwabuchi KA, Chantzoura E, Ruetz T, Johnsson A, Tomlinson SR,
795 Linnarsson S, Kaji K: **High-resolution analysis with novel cell-surface markers identifies**
796 **routes to iPS cells**. *Nature* 2013, **499**(7456):88-91.

797 38. Huang Y, Zhang H, Wang L, Tang C, Qin X, Wu X, Pan M, Tang Y, Yang Z, Babarinde IA *et al*:
798 **JMJD3 acts in tandem with KLF4 to facilitate reprogramming to pluripotency**. *Nat*
799 *Commun* 2020, **11**(1):5061.

39. Seeburg PH, Colby WW, Capon DJ, Goeddel DV, Levinson AD: **Biological properties of human c-Ha-ras1 genes mutated at codon 12**. *Nature* 1984, **312**(5989):71-75.

40. Hong H, Takahashi K, Ichisaka T, Aoi T, Kanagawa O, Nakagawa M, Okita K, Yamanaka S: **Suppression of induced pluripotent stem cell generation by the p53-p21 pathway**. *Nature* 2009, **460**(7259):1132-1135.

41. Luo Z, Qing X, Benda C, Huang Z, Zhang M, Huang Y, Zhang H, Wang L, Lai Y, Ward C *et al*: **Nuclear-cytoplasmic shuttling of class IIa histone deacetylases regulates somatic cell reprogramming**. *Cell Regen* 2019, **8**(1):21-29.

42. Chunduri NK, Storchova Z: **The diverse consequences of aneuploidy**. *Nat Cell Biol* 2019, **21**(1):54-62.

43. Shahbazi MN, Wang T, Tao X, Weatherbee BAT, Sun L, Zhan Y, Keller L, Smith GD, Pellicer A, Scott RT, Jr. *et al*: **Developmental potential of aneuploid human embryos cultured beyond implantation**. *Nat Commun* 2020, **11**(1):3987.

44. Yang Y, Shi L, Fu X, Ma G, Yang Z, Li Y, Zhou Y, Yuan L, Xia Y, Zhong X *et al*: **Metabolic and epigenetic dysfunctions underlie the arrest of in vitro fertilized human embryos in a senescent-like state**. *PLoS Biol* 2022, **20**(6):e3001682.

45. Taapken SM, Nisler BS, Newton MA, Sampsell-Barron TL, Leonhard KA, McIntire EM, Montgomery KD: **Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells**. *Nat Biotechnol* 2011, **29**(4):313-314.

46. Gaztelumendi N, Nogues C: **Chromosome instability in mouse embryonic stem cells**. *Sci Rep* 2014, **4**:5324.

47. Hutchins AP, Yang Z, Li Y, He F, Fu X, Wang X, Li D, Liu K, He J, Wang Y *et al*: **Models of global gene expression define major domains of cell type and tissue identity**. *Nucleic Acids Res* 2017, **45**(5):2354-2367.

48. Steffens S, Fu X, He F, Li Y, Babarinde IA, Hutchins AP: **DPre: computational identification of differentiation bias and genes underlying cell type conversions**. *Bioinformatics* 2020, **36**(5):1637-1639.

49. Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A: **The ground state of embryonic stem cell self-renewal**. *Nature* 2008, **453**(7194):519-523.

50. Kareta MS, Gorges LL, Hafeez S, Benayoun BA, Marro S, Zmoos AF, Cecchini MJ, Spacek D, Batista LF, O'Brien M *et al*: **Inhibition of pluripotency networks by the Rb tumor suppressor restricts reprogramming and tumorigenesis**. *Cell Stem Cell* 2015, **16**(1):39-50.

51. Tapia N, Scholer HR: **p53 connects tumorigenesis and reprogramming to pluripotency**. *J Exp Med* 2010, **207**(10):2045-2048.

52. Kawamura T, Suzuki J, Wang YV, Menendez S, Morera LB, Raya A, Wahl GM, Izpisua Belmonte JC: **Linking the p53 tumour suppressor pathway to somatic cell reprogramming**. *Nature* 2009, **460**(7259):1140-1144.

53. Marion RM, Strati K, Li H, Murga M, Blanco R, Ortega S, Fernandez-Capetillo O, Serrano M, Blasco MA: **A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity**. *Nature* 2009, **460**(7259):1149-1153.

54. Utikal J, Polo JM, Stadtfeld M, Maherali N, Kulalert W, Walsh RM, Khalil A, Rheinwald JG, Hochedlinger K: **Immortalization eliminates a roadblock during cellular reprogramming into iPS cells**. *Nature* 2009, **460**(7259):1145-1148.

55. Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL: **Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming**. *Cell Stem Cell* 2010, **7**(1):64-77.

56. Li R, Liang J, Ni S, Zhou T, Qing X, Li H, He W, Chen J, Li F, Zhuang Q *et al*: **A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts**. *Cell Stem Cell* 2010, **7**(1):51-63.

57. Medvedeva YA, Lennartsson A, Ehsani R, Kulakovskiy IV, Vorontsov IE, Panahandeh P, Khimulya G, Kasukawa T, Consortium F, Drablos F: **EpiFactors: a comprehensive database of human epigenetic factors and complexes**. *Database (Oxford)* 2015, **2015**:bav067.

58. He J, Fu X, Zhang M, He F, Li W, Abdul MM, Zhou J, Sun L, Chang C, Li Y *et al*: **Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells**. *Nat Commun* 2019, **10**(1):34.

59. Li D, Liu J, Yang X, Zhou C, Guo J, Wu C, Qin Y, Guo L, He J, Yu S *et al*: **Chromatin Accessibility Dynamics during iPSC Reprogramming**. *Cell Stem Cell* 2017, **21**(6):819-833 e816.

60. Suknuntha K, Ishii Y, Tao L, Hu K, McIntosh BE, Yang D, Swanson S, Stewart R, Wang JYJ, Thomson J *et al*: **Discovery of survival factor for primitive chronic myeloid leukemia cells using induced pluripotent stem cells**. *Stem Cell Res* 2015, **15**(3):678-693.

61. Kim J, Zaret KS: **Reprogramming of human cancer cells to pluripotency for models of cancer progression**. *EMBO J* 2015, **34**(6):739-747.

62. Ferreiros A, Pedrosa P, Da Silva-Alvarez S, Triana-Martinez F, Vilas JM, Picallos-Rabina P, Gonzalez P, Gomez M, Li H, Garcia-Caballero T *et al*: **Context-Dependent Impact of RAS Oncogene Expression on Cellular Reprogramming to Pluripotency**. *Stem Cell Reports* 2019, **12**(5):1099-1112.

63. Ito K, Nagata K, Ohta S, Matsuda Y, Ukai T, Yasuda I, Ota A, Kobayashi R, Kabata M, Sankoda N *et al*: **The oncogene-dependent resistance to reprogramming unveils cancer therapeutic targets**. *Cell Rep* 2022, **39**(4):110721.

64. Kong Y, Gimple RC, McVicar RN, Hodges AP, Yin J, Liu Y, Zhan W, Snyder EY: **"Reprogram Enablement" as an Assay for Identifying Early Oncogenic Pathways by Their Ability to Allow Neoplastic Cells to Reacquire an Epiblast State**. *Stem Cell Reports* 2020, **15**(3):761-775.

65. Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A: **Promotion of reprogramming to ground state pluripotency by signal inhibition**. *PLoS Biol* 2008, **6**(10):e253.

66. Shi Y, Do JT, Desponts C, Hahm HS, Scholer HR, Ding S: **A combined chemical and genetic approach for the generation of induced pluripotent stem cells**. *Cell Stem Cell* 2008, **2**(6):525-528.

67. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O *et al*: **Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation**. *Cell* 2018, **173**(2):338-354 e315.

68. He J, Babarinde IA, Sun L, Xu S, Chen R, Shi J, Wei Y, Li Y, Ma G, Zhuang Q *et al*: **Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE**. *Nat Commun* 2021, **12**(1):1456.

886 69. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for**
887 **RNA-seq data with DESeq2**. *Genome Biol* 2014, **15**(12):550.

888 70. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A: **Fast gene set**
889 **enrichment analysis**. *bioRxiv* 2021:060012.

890 71. Young MD, Wakefield MJ, Smyth GK, Oshlack A: **Gene ontology analysis for RNA-seq:**
891 **accounting for selection bias**. *Genome Biol* 2010, **11**(2):R14.

892 72. Hutchins AP, Jauch R, Dyla M, Miranda-Saavedra D: **glbase: a framework for combining,**
893 **analyzing and displaying heterogeneous genomic and high-throughput sequencing data**.
894 *Cell Regeneration* 2014, **3**:1.

895 73. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **Transposition of native**
896 **chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding**
897 **proteins and nucleosome position**. *Nat Methods* 2013, **10**(12):1213-1218.

898 74. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nat Methods* 2012,
899 **9**(4):357-359.

900 75. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM,
901 Brown M, Li W *et al*: **Model-based analysis of ChIP-Seq (MACS)**. *Genome Biol* 2008,
902 **9**(9):R137.

903 76. Ma G, Babarinde IA, Zhuang Q, Hutchins AP: **Unified Analysis of Multiple ChIP-Seq**
904 **Datasets**. *Methods Mol Biol* 2021, **2198**:451-465.

905 77. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass
906 CK: **Simple combinations of lineage-determining transcription factors prime cis-**
907 **regulatory elements required for macrophage and B cell identities**. *Mol Cell* 2010,
908 **38**(4):576-589.

909 78. Malik V, Glaser LV, Zimmer D, Velychko S, Weng M, Holzner M, Arend M, Chen Y, Srivastava
910 Y, Veerapandian V *et al*: **Pluripotency reprogramming by competent and incompetent**
911 **POU factors uncovers temporal dependency for Oct4 and Sox2**. *Nat Commun* 2019,
912 **10**(1):3477.
913

914

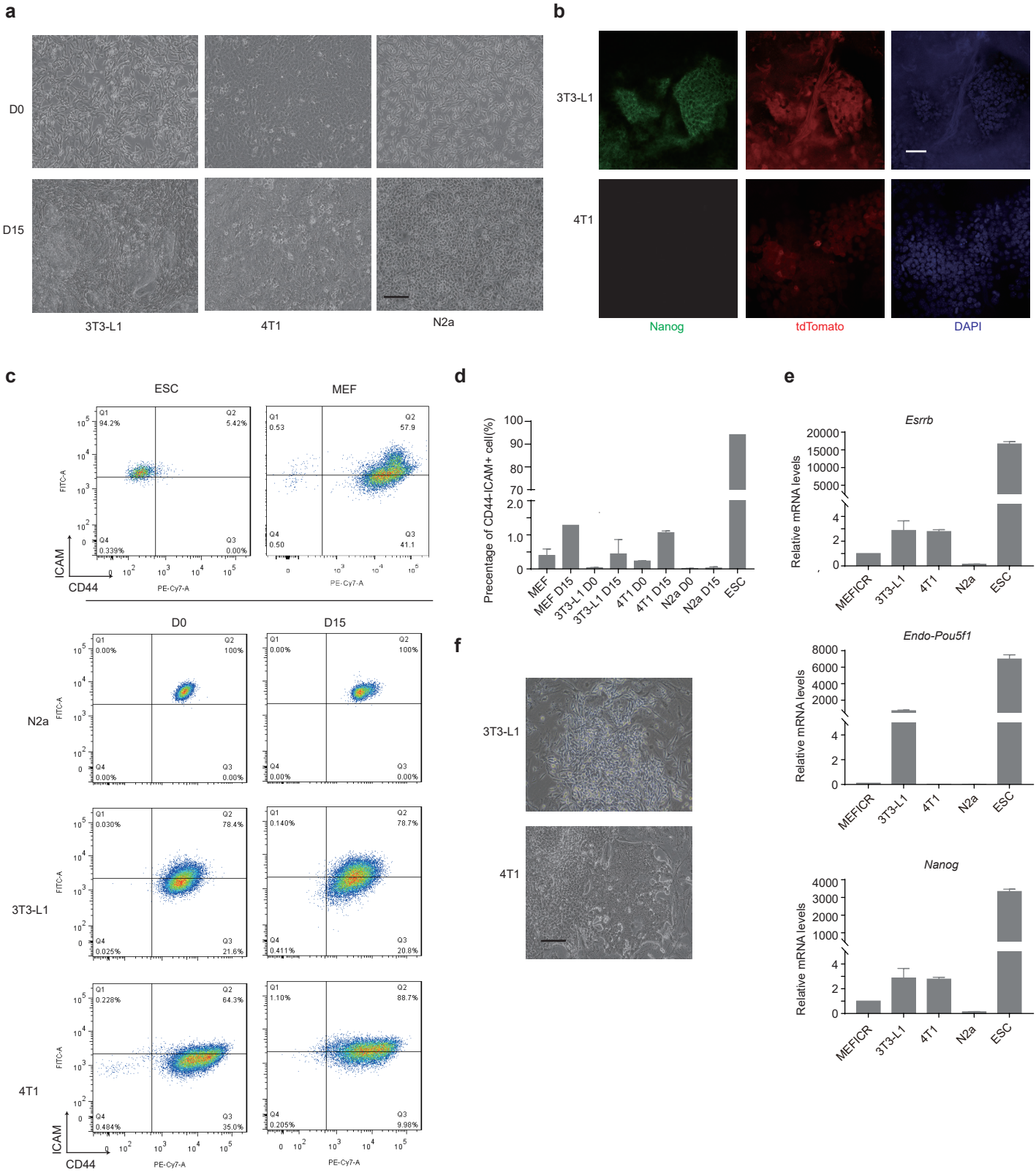
a
D0
D15
3T3-L1
4T1
N2a
b
3T3-L1
4T1
Nanog
tdTomato
DAPI
c
ESC
MEF
D0
D15
N2a
3T3-L1
4T1
ICAM
CD44
FITC-A
PE-Cy7-A
d
Precentage of CD44-ICAM+ cell(%)
MEF
MEF D15
3T3-L1 D0
3T3-L1 D15
4T1 D0
4T1 D15
N2a D0
N2a D15
ESC
e
Esrrb
Endo-Pou5f1
Nanog
Relative mRNA levels
MEFICR
3T3-L1
4T1
N2a
ESC
f
3T3-L1
4T1

Figure 1

915 **Figure 1. Transformed cell lines are highly resistant to reprogramming.**

916 **a** Bright-field images of the reprogramming of 3T3-L1, 4T1, and N2a cell lines in
917 serum+Vc reprogramming conditions. Images are shown from day 0 (D0) or day
918 15 (D15) of the reprogramming experiment. Morphologically iPSC-like cells can
919 typically be observed at D15 and onwards in wild-type reprogramming. Scale
920 bar=100 μm.

921 **b** Fluorescent microscopy images for NANOG and tdTomato expressed from the
922 OSKM transgene cassette for the reprogramming of 3T3-L1 and 4T1 cells. Scale
923 bar=100 μm.

924 **c** FACS (fluorescent activated cell sorting) plots for wildtype MEFs and mESCs and
925 cells from a reprogramming time course at day 0 or 15 from 3T3-L1, 4T1, and N2a
926 cells.

927 **d** Bar chart of the percentage of ICAM+/CD44- sorted cells in the indicated cell lines
928 at day 0 or day 15 of reprogramming. Data is from the mean of three biological
929 replicates.

930 **e** RT-qPCR showed the indicated gene expression from pooled cultures of the sorted
931 ICAM+ and CD44- cells from the indicated cell lines five days after replating. The
932 mean of two biological replicates ± s.d. is shown.

933 **f** Bright-field images of the sorted ICAM+ and CD44- cells from the indicated cell
934 lines five days after replating. Scale bar=100 μm.

935

936

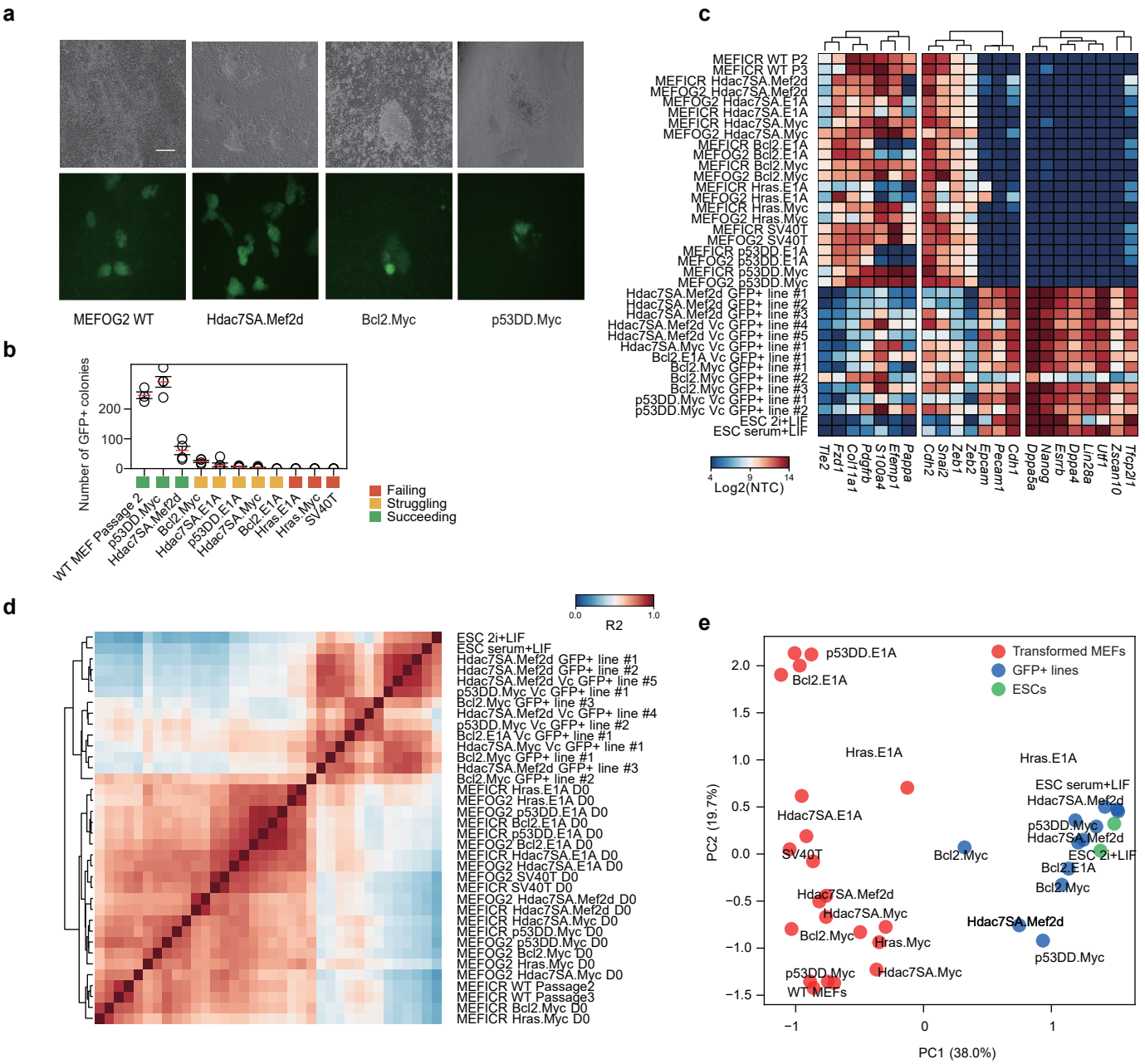
a
MEFOG2 WT
Hdac7SA.Mef2d
Bcl2.Myc
p53DD.Myc
b
Number of GFP+ colonies
200
100
0
WT MEF Passage 2
p53DD.Myc
Hdac7SA.Mef2d
Bcl2.Myc
Hdac7SA.E1A
p53DD.E1A
Hdac7SA.Myc
Bcl2.E1A
Hras.E1A
Hras.Myc
SV40T
Failing
Struggling
Succeeding
c
MEFICR WT P2
MEFICR WT P3
MEFICR Hdac7SA.Mef2d
MEFOG2 Hdac7SA.Mef2d
MEFOG2 Hdac7SA.E1A
MEFICR Hdac7SA.E1A
MEFICR Hdac7SA.Myc
MEFOG2 Hdac7SA.Myc
MEFICR Bcl2.E1A
MEFOG2 Bcl2.E1A
MEFICR Bcl2.Myc
MEFOG2 Bcl2.Myc
MEFICR Hras.E1A
MEFOG2 Hras.E1A
MEFICR Hras.Myc
MEFOG2 Hras.Myc
MEFICR SV40T
MEFOG2 SV40T
MEFICR p53DD.E1A
MEFOG2 p53DD.E1A
MEFICR p53DD.Myc
MEFOG2 p53DD.Myc
Hdac7SA.Mef2d GFP+ line #1
Hdac7SA.Mef2d GFP+ line #2
Hdac7SA.Mef2d GFP+ line #3
Hdac7SA.Mef2d Vc GFP+ line #4
Hdac7SA.Mef2d Vc GFP+ line #5
Hdac7SA.Myc Vc GFP+ line #1
Bcl2.E1A Vc GFP+ line #1
Bcl2.Myc GFP+ line #1
Bcl2.Myc GFP+ line #2
Bcl2.Myc GFP+ line #3
p53DD.Myc Vc GFP+ line #1
p53DD.Myc Vc GFP+ line #2
ESC 2i+LIF
ESC serum+LIF
Tle2
Fzd1
Col11a1
Pdgfrb
S100a4
Efemp1
Pappa
Cdh2
Snai2
Zeb1
Zeb2
Epcam
Pecam1
Cdh1
Dppa5a
Nanog
Esrrb
Dppa4
Lin28a
Utf1
Zscan10
Tfcp2l1
4
9
14
Log2(NTC)
d
0.0
1.0
R2
ESC 2i+LIF
ESC serum+LIF
Hdac7SA.Mef2d GFP+ line #1
Hdac7SA.Mef2d GFP+ line #2
Hdac7SA.Mef2d Vc GFP+ line #5
p53DD.Myc Vc GFP+ line #1
Bcl2.Myc GFP+ line #3
Hdac7SA.Mef2d Vc GFP+ line #4
p53DD.Myc Vc GFP+ line #2
Bcl2.E1A Vc GFP+ line #1
Hdac7SA.Myc Vc GFP+ line #1
Bcl2.Myc GFP+ line #1
Hdac7SA.Mef2d GFP+ line #3
Bcl2.Myc GFP+ line #2
MEFICR Hras.E1A D0
MEFOG2 Hras.E1A D0
MEFOG2 p53DD.E1A D0
MEFICR Bcl2.E1A D0
MEFICR p53DD.E1A D0
MEFOG2 Bcl2.E1A D0
MEFICR Hdac7SA.E1A D0
MEFOG2 Hdac7SA.E1A D0
MEFOG2 SV40T D0
MEFICR SV40T D0
MEFOG2 Hdac7SA.Mef2d D0
MEFICR Hdac7SA.Mef2d D0
MEFICR Hdac7SA.Myc D0
MEFICR p53DD.Myc D0
MEFOG2 p53DD.Myc D0
MEFOG2 Bcl2.Myc D0
MEFOG2 Hras.Myc D0
MEFOG2 Hdac7SA.Myc D0
MEFICR WT Passage2
MEFICR WT Passage3
MEFICR Bcl2.Myc D0
MEFICR Hras.Myc D0
e
Transformed MEFs
GFP+ lines
ESCs
p53DD.E1A
Bcl2.E1A
Hras.E1A
Hdac7SA.E1A
SV40T
Bcl2.Myc
Hdac7SA.Mef2d
Hdac7SA.Myc
Hras.Myc
p53DD.Myc
WT MEFs
ESC serum+LIF
ESC 2i+LIF
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
PC2 (19.7%)
−1
0
1
PC1 (38.0%)

Figure 2

937 **Figure 2. Characterization of transformed fibroblast cell lines capable and**
938 **incapable of reprogramming to iPSC-like cells.**

939 **a** Representative brightfield and GFP images from reprogrammed transformed
940 MEFs for the indicated cell lines. Scale bar=100 μm.

941 **b** Chart showing the number of Oct4-GFP+ colonies from the indicated transformed
942 MEF lines. Each MEF line was at least 1 month old, the experiment was performed
943 at least three times for each cell line. Circles represent biological replicates; the
944 red line indicates the mean and the error bars indicate the standard error of the
945 mean (SEM).

946 **c** Heatmap from the RNA-seq data showing the expression of a representative set
947 of genes specific to somatic cells, pluripotent stem cells, or from the MET
948 (mesenchymal-epithelial transition). NTC=normalized tag count.

949 **d** Correlation heatmap (R^2) of the original MEF samples and the reprogrammed
950 GFP+ cells and ESCs.

951 **e** PCA showing the first two dimensions of the RNA-seq data from the original
952 wildtype or transformed MEF lines (red) and the GFP+ cell lines (blue) and ESCs
953 (green) maintained in serum+LIF or 2i+LIF.

954

955

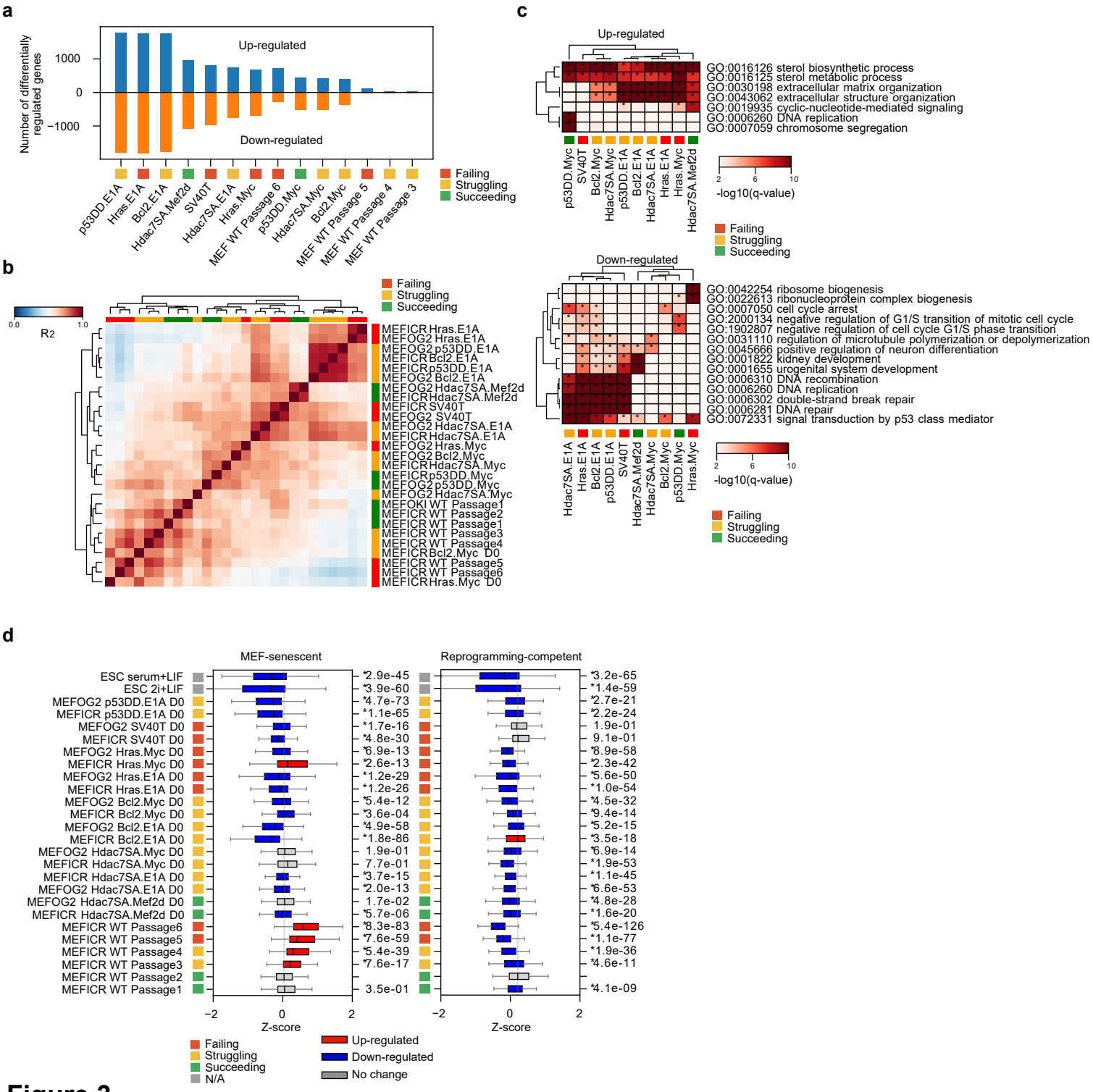
a
Number of differentially regulated genes
1000
0
−1000
Up-regulated
Down-regulated
p53DD.E1A
Hras.E1A
Bcl2.E1A
Hdac7SA.Mef2d
SV40T
Hdac7SA.E1A
Hras.Myc
MEF WT Passage 6
p53DD.Myc
Hdac7SA.Myc
Bcl2.Myc
MEF WT Passage 5
MEF WT Passage 4
MEF WT Passage 3
Failing
Struggling
Succeeding
b
R2
0.0
1.0
Failing
Struggling
Succeeding
MEFICR Hras.E1A
MEFOG2 Hras.E1A
MEFOG2 p53DD.E1A
MEFICR Bcl2.E1A
MEFICR p53DD.E1A
MEFOG2 Bcl2.E1A
MEFOG2 Hdac7SA.Mef2d
MEFICR Hdac7SA.Mef2d
MEFICR SV40T
MEFOG2 SV40T
MEFOG2 Hdac7SA.E1A
MEFICR Hdac7SA.E1A
MEFOG2 Hras.Myc
MEFOG2 Bcl2.Myc
MEFICR Hdac7SA.Myc
MEFICR p53DD.Myc
MEFOG2 p53DD.Myc
MEFOG2 Hdac7SA.Myc
MEFOKI WT Passage1
MEFICR WT Passage2
MEFICR WT Passage1
MEFICR WT Passage3
MEFICR WT Passage4
MEFICR Bcl2.Myc D0
MEFICR WT Passage5
MEFICR WT Passage6
MEFICR Hras.Myc D0
c
Up-regulated
GO:0016126 sterol biosynthetic process
GO:0016125 sterol metabolic process
GO:0030198 extracellular matrix organization
GO:0043062 extracellular structure organization
GO:0019935 cyclic-nucleotide-mediated signaling
GO:0006260 DNA replication
GO:0007059 chromosome segregation
p53DD.Myc
SV40T
Bcl2.Myc
Hdac7SA.Myc
p53DD.E1A
Bcl2.E1A
Hdac7SA.E1A
Hras.E1A
Hras.Myc
Hdac7SA.Mef2d
2
6
10
-log10(q-value)
Failing
Struggling
Succeeding
Down-regulated
GO:0042254 ribosome biogenesis
GO:0022613 ribonucleoprotein complex biogenesis
GO:0007050 cell cycle arrest
GO:2000134 negative regulation of G1/S transition of mitotic cell cycle
GO:1902807 negative regulation of cell cycle G1/S phase transition
GO:0031110 regulation of microtubule polymerization or depolymerization
GO:0045666 positive regulation of neuron differentiation
GO:0001822 kidney development
GO:0001655 urogenital system development
GO:0006310 DNA recombination
GO:0006260 DNA replication
GO:0006302 double-strand break repair
GO:0006281 DNA repair
GO:0072331 signal transduction by p53 class mediator
Hdac7SA.E1A
Hras.E1A
Bcl2.E1A
p53DD.E1A
SV40T
Hdac7SA.Mef2d
Hdac7SA.Myc
Bcl2.Myc
p53DD.Myc
Hras.Myc
2
6
10
-log10(q-value)
Failing
Struggling
Succeeding
d
MEF-senescent
Reprogramming-competent
ESC serum+LIF
ESC 2i+LIF
MEFOG2 p53DD.E1A D0
MEFICR p53DD.E1A D0
MEFOG2 SV40T D0
MEFICR SV40T D0
MEFOG2 Hras.Myc D0
MEFICR Hras.Myc D0
MEFOG2 Hras.E1A D0
MEFICR Hras.E1A D0
MEFOG2 Bcl2.Myc D0
MEFICR Bcl2.Myc D0
MEFOG2 Bcl2.E1A D0
MEFICR Bcl2.E1A D0
MEFOG2 Hdac7SA.Myc D0
MEFICR Hdac7SA.Myc D0
MEFICR Hdac7SA.E1A D0
MEFOG2 Hdac7SA.E1A D0
MEFOG2 Hdac7SA.Mef2d D0
MEFICR Hdac7SA.Mef2d D0
MEFICR WT Passage6
MEFICR WT Passage5
MEFICR WT Passage4
MEFICR WT Passage3
MEFICR WT Passage2
MEFICR WT Passage1
*2.9e-45
*3.9e-60
*4.7e-73
*1.1e-65
*1.7e-16
*4.8e-30
*6.9e-13
*2.6e-13
*1.2e-29
*1.2e-26
*5.4e-12
*3.6e-04
*4.9e-58
*1.8e-86
1.9e-01
7.7e-01
*3.7e-15
*2.0e-13
1.7e-02
*5.7e-06
*8.3e-83
*7.6e-59
*5.4e-39
*7.6e-17
3.5e-01
*3.2e-65
*1.4e-59
*2.7e-21
*2.2e-24
1.9e-01
9.1e-01
*8.9e-58
*2.3e-42
*5.6e-50
*1.0e-54
*4.5e-32
*9.4e-14
*5.2e-15
*3.5e-18
*6.9e-14
*1.9e-53
*1.1e-45
*6.6e-53
*4.8e-28
*1.6e-20
*5.4e-126
*1.1e-77
*1.9e-36
*4.6e-11
*4.1e-09
−2
0
2
Z-score
Failing
Struggling
Succeeding
N/A
Up-regulated
Down-regulated
No change

Figure 3

956 **Figure 3. The immortalized MEFs are widely divergent from wild-type MEFs.**

957 **a** Chart showing the number of significantly differentially regulated genes up (+) or
958 down (-) regulated (Bonferroni-Hochberg corrected p-value of <0.01 and a
959 minimum fold-change of 2).

960 **b** Correlation (R^2) of the immortalized MEF lines compared to the wild-type MEFs.

961 **c** Selected GO terms from the significantly up or down-regulated genes from the
962 transformed MEF lines. A term was considered significant if it had a Bonferroni-
963 Hochberg corrected p-value of <0.01). See also **Figure S4b** for the full table.

964 **d** Boxplots for the genes defined as the MEF-senescent signature or the
965 reprogramming-capable signature from **Figure S1c**. Significance is from a two-
966 sided Welch's t-test for pair-wise comparisons versus MEFICR wildtype Passage
967 2 as the reference. A change was considered significant if the p-value was < 0.01.
968 Significant changes are colored red or blue.

969

970

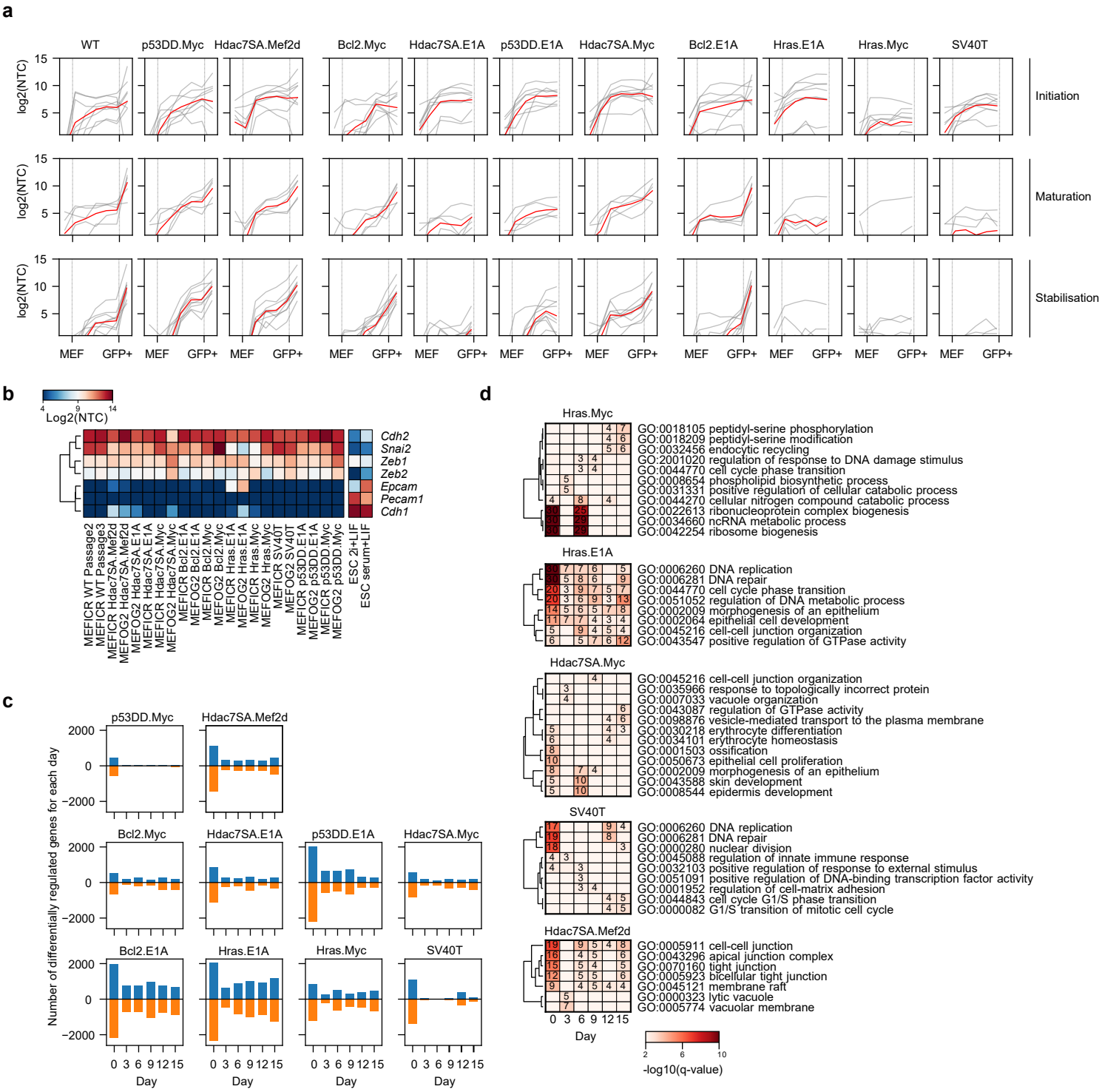
a
WT
p53DD.Myc
Hdac7SA.Mef2d
Bcl2.Myc
Hdac7SA.E1A
p53DD.E1A
Hdac7SA.Myc
Bcl2.E1A
Hras.E1A
Hras.Myc
SV40T
log2(NTC)
Initiation
Maturation
Stabilisation
MEF
GFP+
b
Log2(NTC)
Cdh2
Snai2
Zeb1
Zeb2
Epcam
Pecam1
Cdh1
MEFICR WT Passage2
MEFICR WT Passage3
MEFICR Hdac7SA.Mef2d
MEFOG2 Hdac7SA.Mef2d
MEFOG2 Hdac7SA.E1A
MEFICR Hdac7SA.E1A
MEFICR Hdac7SA.Myc
MEFOG2 Hdac7SA.Myc
MEFICR Bcl2.E1A
MEFOG2 Bcl2.E1A
MEFICR Bcl2.Myc
MEFOG2 Bcl2.Myc
MEFICR Hras.E1A
MEFOG2 Hras.E1A
MEFICR Hras.Myc
MEFOG2 Hras.Myc
MEFICR SV40T
MEFOG2 SV40T
MEFICR p53DD.E1A
MEFOG2 p53DD.E1A
MEFICR p53DD.Myc
MEFOG2 p53DD.Myc
ESC 2i+LIF
ESC serum+LIF
c
Number of differentially regulated genes for each day
Day
d
Hras.Myc
GO:0018105 peptidyl-serine phosphorylation
GO:0018209 peptidyl-serine modification
GO:0032456 endocytic recycling
GO:2001020 regulation of response to DNA damage stimulus
GO:0044770 cell cycle phase transition
GO:0008654 phospholipid biosynthetic process
GO:0031331 positive regulation of cellular catabolic process
GO:0044270 cellular nitrogen compound catabolic process
GO:0022613 ribonucleoprotein complex biogenesis
GO:0034660 ncRNA metabolic process
GO:0042254 ribosome biogenesis
Hras.E1A
GO:0006260 DNA replication
GO:0006281 DNA repair
GO:0044770 cell cycle phase transition
GO:0051052 regulation of DNA metabolic process
GO:0002009 morphogenesis of an epithelium
GO:0002064 epithelial cell development
GO:0045216 cell-cell junction organization
GO:0043547 positive regulation of GTPase activity
Hdac7SA.Myc
GO:0045216 cell-cell junction organization
GO:0035966 response to topologically incorrect protein
GO:0007033 vacuole organization
GO:0043087 regulation of GTPase activity
GO:0098876 vesicle-mediated transport to the plasma membrane
GO:0030218 erythrocyte differentiation
GO:0034101 erythrocyte homeostasis
GO:0001503 ossification
GO:0050673 epithelial cell proliferation
GO:0002009 morphogenesis of an epithelium
GO:0043588 skin development
GO:0008544 epidermis development
SV40T
GO:0006260 DNA replication
GO:0006281 DNA repair
GO:0000280 nuclear division
GO:0045088 regulation of innate immune response
GO:0032103 positive regulation of response to external stimulus
GO:0051091 positive regulation of DNA-binding transcription factor activity
GO:0001952 regulation of cell-matrix adhesion
GO:0044843 cell cycle G1/S phase transition
GO:0000082 G1/S transition of mitotic cell cycle
Hdac7SA.Mef2d
GO:0005911 cell-cell junction
GO:0043296 apical junction complex
GO:0070160 tight junction
GO:0005923 bicellular tight junction
GO:0045121 membrane raft
GO:0000323 lytic vacuole
GO:0005774 vacuolar membrane
Day
-log10(q-value)

Figure 4

971 **Figure 4. Transformed MEFs fail at different phases of reprogramming**

972 **a** Line charts showing the expression of initiation, maturation, and stabilization genes
973 [55], at the indicated time points or in MEFs and GFP+ cells. See **Figure S6** for the
974 expression of the individual genes in each class. The red line is the mean of all
975 genes, the grey lines are the individual genes.

976 **b** Heatmap of the expression of mesenchymal and epithelial marker genes in starting
977 lines and ESCs.

978 **c** Bar charts showing the number of significantly differentially regulated genes at
979 each time point for the indicated transformed lines versus a wildtype
980 reprogramming time course. Differential expression was compared for each line
981 and time point versus the corresponding wildtype timepoint. A gene was
982 considered differentially regulated if the q-value (Bonferroni-Hochberg corrected p-
983 value) was <0.01 and the fold-change >2.

984 **d** Heatmaps of significantly overrepresented gene ontology (biological process)
985 terms for down-regulated genes from the indicated time points and cell lines.

986

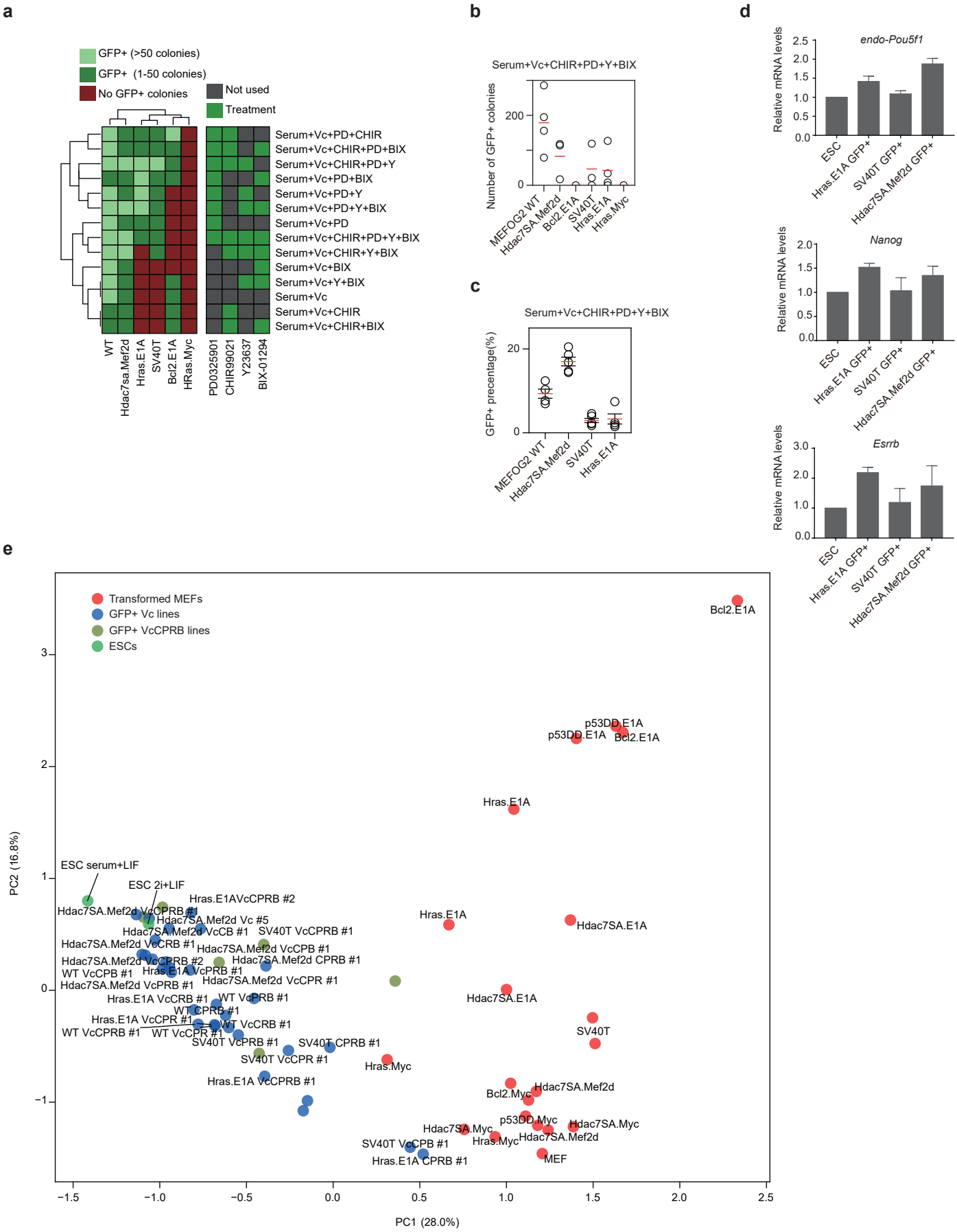
a
GFP+ (>50 colonies)
GFP+ (1-50 colonies)
No GFP+ colonies
Not used
Treatment
Serum+Vc+PD+CHIR
Serum+Vc+CHIR+PD+BIX
Serum+Vc+CHIR+PD+Y
Serum+Vc+PD+BIX
Serum+Vc+PD+Y
Serum+Vc+PD+Y+BIX
Serum+Vc+PD
Serum+Vc+CHIR+PD+Y+BIX
Serum+Vc+CHIR+Y+BIX
Serum+Vc+BIX
Serum+Vc+Y+BIX
Serum+Vc
Serum+Vc+CHIR
Serum+Vc+CHIR+BIX
WT
Hdac7sa.Mef2d
Hras.E1A
SV40T
Bcl2.E1A
HRas.Myc
PD0325901
CHIR99021
Y23637
BIX-01294
b
Serum+Vc+CHIR+PD+Y+BIX
Number of GFP+ colonies
200
0
MEFOG2 WT
Hdac7SA.Mef2d
Bcl2.E1A
SV40T
Hras.E1A
Hras.Myc
c
Serum+Vc+CHIR+PD+Y+BIX
GFP+ precentage(%)
20
0
MEFOG2 WT
Hdac7SA.Mef2d
SV40T
Hras.E1A
d
endo-Pou5f1
Relative mRNA levels
2.5
2.0
1.5
1.0
0.5
0.0
ESC
Hras.E1A GFP+
SV40T GFP+
Hdac7SA.Mef2d GFP+
Nanog
Relative mRNA levels
2.0
1.5
1.0
0.5
0.0
ESC
Hras.E1A GFP+
SV40T GFP+
Hdac7SA.Mef2d GFP+
Esrrb
Relative mRNA levels
3.0
2.0
1.0
0.0
ESC
Hras.E1A GFP+
SV40T GFP+
Hdac7SA.Mef2d GFP+
e
Transformed MEFs
GFP+ Vc lines
GFP+ VcCPRB lines
ESCs
PC2 (16.8%)
PC1 (28.0%)
3
2
1
0
−1
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
Bcl2.E1A
p53DD.E1A
p53DD.E1A
Bcl2.E1A
Hras.E1A
ESC serum+LIF
ESC 2i+LIF
Hras.E1AVcCPRB #2
Hdac7SA.Mef2d VcCPRB #1
Hdac7SA.Mef2d Vc #5
Hdac7SA.Mef2d VcCB #1
SV40T VcCPRB #1
Hdac7SA.Mef2d VcCRB #1
Hdac7SA.Mef2d VcCPB #1
Hdac7SA.Mef2d VcCPRB #2
Hdac7SA.Mef2d CPRB #1
WT VcCPB #1
Hras.E1A VcPRB #1
Hdac7SA.Mef2d VcCPR #1
Hdac7SA.Mef2d VcPRB #1
Hras.E1A VcCRB #1
WT VcPRB #1
WT CPRB #1
Hras.E1A VcCPR #1
WT VcCRB #1
WT VcCPRB #1
WT VcCPR #1
SV40T VcPRB #1
SV40T CPRB #1
SV40T VcCPR #1
Hras.E1A VcCPRB #1
Hras.E1A
Hdac7SA.E1A
Hdac7SA.E1A
SV40T
SV40T
Hras.Myc
Bcl2.Myc
Hdac7SA.Mef2d
p53DD.Myc
Hdac7SA.Myc
Hdac7SA.Myc
Hdac7SA.Mef2d
Hras.Myc
SV40T VcCPB #1
Hras.E1A CPRB #1
MEF

Figure 5

987 **Figure 5. Chemical inhibition can convert some reprogramming-incapable**
988 **immortalized MEFs to reprogramming-capable**

989 **a** Heatmap showing the number of GFP colonies produced depending upon the
990 reprogramming conditions used, for the indicated transformed lines.

991 **b** Chart showing the number of GFP+ colonies in the indicated immortal cell lines
992 with the inhibitor cocktail. Data is from four biological replicates (MEFOGF WT) or
993 from three biological replicates (all other transformed cell lines).

994 **c** Percentage of cells sorted by FACS at day 12 of the reprogramming assay in the
995 indicated transformed cell lines or wildtype MEFs. Data is from four biological
996 replicates.

997 **d** RT-qPCR at day 15 of a reprogramming experiment using primers targeting
998 endogenous Oct4 (endo-*Pou5f1*), *Nanog, and Esrrb*. The mean of three biological
999 replicates ± s.d. is shown.

1000 **e** PCA of the RNA-seq from the immortalized MEFs, and the reprogrammed iPSCs
1001 and ESCs. The various chemical cocktails used during the reprogramming are
1002 abbreviated for clarity. Key: Vc=Vitamin C; P=PD0325901; C=CHIR99021;
1003 R=Y23637; B=BIX-01294.

1004

1005

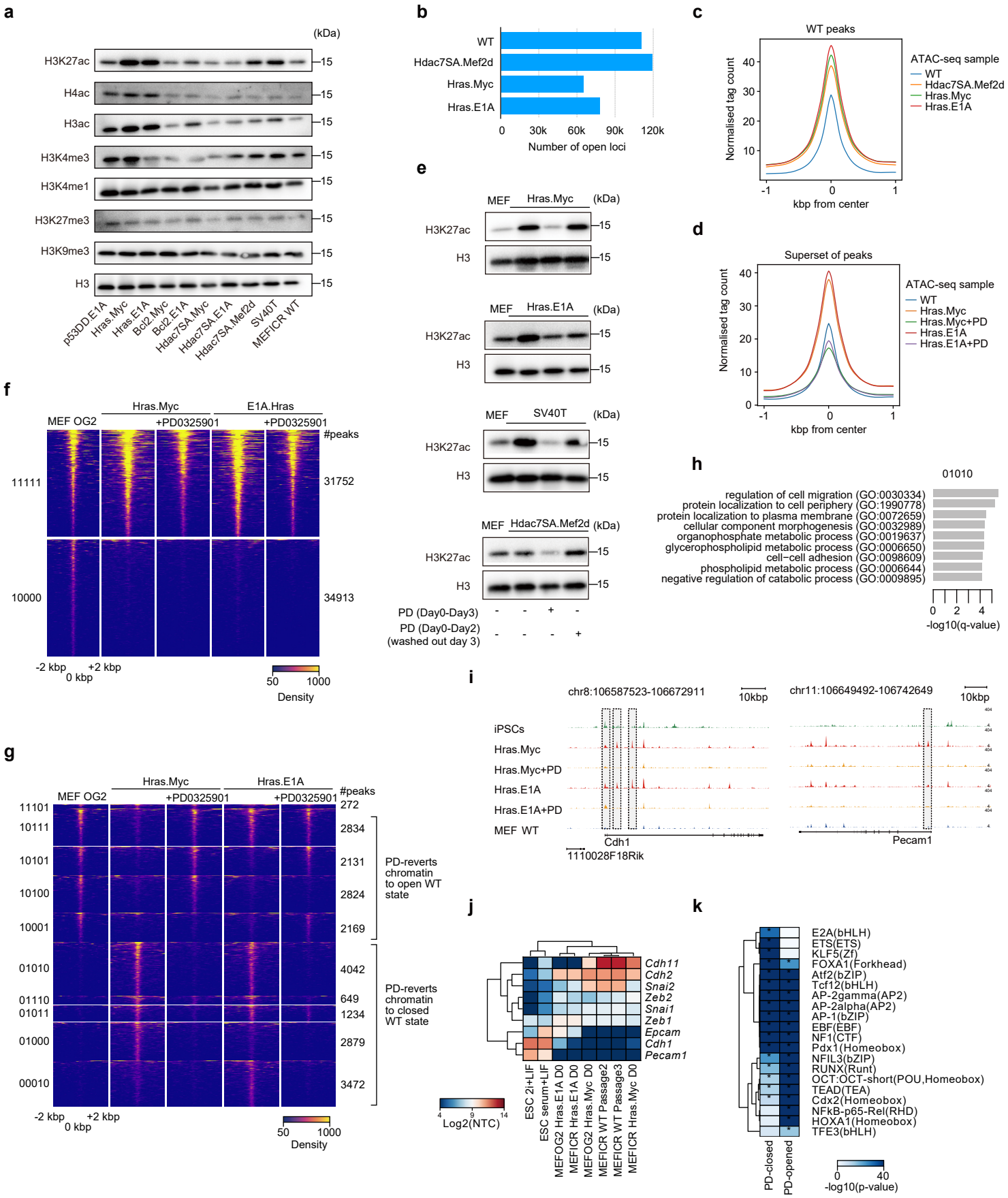
a
(kDa)
H3K27ac
H4ac
H3ac
H3K4me3
H3K4me1
H3K27me3
H3K9me3
H3
15
p53DD E1A
Hras.Myc
Hras.E1A
Bcl2.Myc
Bcl2.E1A
Hdac7SA.Myc
Hdac7SA.E1A
Hdac7SA.Mef2d
SV40T
MEFICR WT
b
WT
Hdac7SA.Mef2d
Hras.Myc
Hras.E1A
0
30k
60k
90k
120k
Number of open loci
c
WT peaks
Normalised tag count
kbp from center
ATAC-seq sample
WT
Hdac7SA.Mef2d
Hras.Myc
Hras.E1A
d
Superset of peaks
Normalised tag count
kbp from center
ATAC-seq sample
WT
Hras.Myc
Hras.Myc+PD
Hras.E1A
Hras.E1A+PD
e
MEF
Hras.Myc
Hras.E1A
SV40T
Hdac7SA.Mef2d
(kDa)
H3K27ac
H3
PD (Day0-Day3)
PD (Day0-Day2)
(washed out day 3)
f
Hras.Myc
E1A.Hras
MEF OG2
+PD0325901
#peaks
11111
31752
10000
34913
-2 kbp
+2 kbp
0 kbp
50
1000
Density
g
Hras.Myc
Hras.E1A
MEF OG2
+PD0325901
#peaks
11101
272
10111
2834
10101
2131
10100
2824
10001
2169
01010
4042
01110
649
01011
1234
01000
2879
00010
3472
PD-reverts chromatin to open WT state
PD-reverts chromatin to closed WT state
-2 kbp
+2 kbp
0 kbp
50
1000
Density
h
01010
regulation of cell migration (GO:0030334)
protein localization to cell periphery (GO:1990778)
protein localization to plasma membrane (GO:0072659)
cellular component morphogenesis (GO:0032989)
organophosphate metabolic process (GO:0019637)
glycerophospholipid metabolic process (GO:0006650)
cell−cell adhesion (GO:0098609)
phospholipid metabolic process (GO:0006644)
negative regulation of catabolic process (GO:0009895)
0
2
4
-log10(q-value)
i
chr8:106587523-106672911
10kbp
chr11:106649492-106742649
10kbp
iPSCs
Hras.Myc
Hras.Myc+PD
Hras.E1A
Hras.E1A+PD
MEF WT
Cdh1
1110028F18Rik
Pecam1
j
Cdh11
Cdh2
Snai2
Zeb2
Snai1
Zeb1
Epcam
Cdh1
Pecam1
ESC 2i+LIF
ESC serum+LIF
MEFOG2 Hras.E1A D0
MEFICR Hras.E1A D0
MEFOG2 Hras.Myc D0
MEFICR WT Passage2
MEFICR WT Passage3
MEFICR Hras.Myc D0
4
9
14
Log2(NTC)
k
E2A(bHLH)
ETS(ETS)
KLF5(Zf)
FOXA1(Forkhead)
Atf2(bZIP)
Tcf12(bHLH)
AP-2gamma(AP2)
AP-2alpha(AP2)
AP-1(bZIP)
EBF(EBF)
NF1(CTF)
Pdx1(Homeobox)
NFIL3(bZIP)
RUNX(Runt)
OCT:OCT-short(POU,Homeobox)
TEAD(TEA)
Cdx2(Homeobox)
NFkB-p65-Rel(RHD)
HOXA1(Homeobox)
TFE3(bHLH)
PD-closed
PD-opened
0
40
-log10(p-value)

Figure 6

1006 **Figure 6. Inhibition of MEK repairs some epigenetic defects in the transformed**
1007 **MEFs**

1008 **a** Western blot for a selection of epigenetic marks and total histone H3 as a loading
1009 control.

1010 **b** Bar chart showing the number of ATAC-seq loci identified as open/accessible in
1011 the indicated wildtype or transformed MEFs.

1012 **c** Pileups of the ATAC-seq data centered on all wildtype MEF peaks and showing
1013 the flanking 1 kbp (kilo base pairs). Normalized tag counts (reads per million per
1014 bin) are shown for the indicated samples.

1015 **d** Pileups of the ATAC-seq data centered on a non-redundant superset of all peaks
1016 from all of the ATAC-seq samples. The plot is centered on the peak summit and
1017 shows the flanking 1 kbp. Normalized tag counts are shown for the indicated
1018 samples.

1019 **e** Western blot of activatory histone marks upon addition of PD to the MEFs for 3
1020 days, or cells treated with PD for two days and then washed out on the third data
1021 (washed out on day 3).

1022 **f** Heatmap of the ATAC-seq data showing the permanently open (11111), and open
1023 only in the wildtype MEFs (10000). The number of peaks in each class is indicated
1024 on the right-hand side. Read density is shown centered on the peak summit and 2
1025 kbp on each side. The full heatmap containing all groups is in **Figure S9b**.

1026 **g** Heatmap of selected peak groups for the indicated ATAC-seq samples. The full
1027 heatmap containing all groups is in **Figure S9b**.

1028 **h** Gene ontology analysis for nearby genes (a TSS within 10 kbp of an open locus)
1029 from the 01010 group (closed in wildtype MEFs, PD causes these loci to close in
1030 the transformed MEFs) from **Figure 6g**.

1031 **i** Genome pileup view of the ATAC-seq data showing the *Cdh1* and *Epcam* loci in
1032 wildtype MEFs, Hras.E1A, and Hras.Myc MEFs with or without treatment with PD.
1033 Genome assembly is mm10. Regions of open chromatin that are modified in the
1034 PD-treated cells are indicated by vertical grey bars.

1035 **j** Expression heatmap of selected epithelial and mesenchymal genes in the
1036 indicated MEF lines and ESCs.

1037 **k** Motif analysis of selected PD-close and PD-open genes.

1038