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 (Iglesias *et al*, 2017). Both cell types have a relaxed chromatin state (Sun *et al*, 2021), adopt a glycolysis-biased metabolism despite the availability of oxygen (Folmes *et al*, 2011; Vander Heiden *et al*, 2009), can undergo an epithelial-mesenchymal transition (EMT) (Li *et al*, 2017b), and form teratomas when injected into immunocompromised mice (Blum & Benvenisty, 2008). 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* (Niu *et al*, 2017). Indeed, expression of pluripotent-specific genes in patient tumor samples correlate with poor clinical outcomes (Hepburn *et al*, 2019; Jang *et al*, 2019). For example, overexpression of *Oct4* led to epithelial lesions (Hochedlinger *et al*, 2005), and *NANOG* expression is associated with colorectal (Zhang *et al*, 2013a) and prostate cancer (Jeter *et al*, 2011).

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* (Takahashi & Yamanaka, 2006). Curiously, despite being an artificial process, the reprogramming of somatic cells to iPSCs passes through distinct phases, reminiscent of a developmental program (Hussein *et al*, 2014; Polo *et al*, 2012; Xu *et al*, 2016). Tumorigenic transformation also passes through a series of distinct phases in a process that has similarities to somatic cell reprogramming (Huyghe *et al*, 2022; Iglesias *et al.*, 2017). Some studies suggest direct links between reprogramming and cancer. For example, transient *in vivo* activation of reprogramming factors leads to tumor formation (Abad *et al*, 2013; Ohnishi *et al*, 2014), and cancer-associated

mutations in the transcription factor *SOX17* can confer reprogramming capability to the normally incapable SOX17 (Srivastava *et al*, 2020).

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 (Carette *et al*, 2010; Kim *et al*, 2017; Miyoshi *et al*, 2010; Zhang *et al*, 2013b). 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 (Dannenmann *et al*, 2021; Li *et al*, 2021; Liu *et al*, 2014), lymphoblastoid cells (Barrett *et al*, 2014; Kumar *et al*, 2016) and chronic myeloid leukemia cells (Bedel *et al*, 2013; Carette *et al.*, 2010; Hu *et al*, 2011; Kumano *et al*, 2012). However, these blood cells seem to be the exception, since in the general transformed cells are resistant to reprogramming (Bang *et al*, 2019). 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 those cells is closer to iPSCs than the parental lines, it remains distinct from genuine ESCs/iPSCs (Kumar *et al.*, 2016; Stricker *et al*, 2013; Zhang *et al.*, 2013b). It is possible that reprogramming from primary tissues is selecting for bystander untransformed cells (Lee *et al*, 2017). 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 (Zhuang *et al*, 2018). 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

109 detected by immunofluorescence in some 3T3-L1 and 4T1 colonies (**Figure 1b**). We
110 attempted to manually pick the iPSC-like colonies and establish iPSC lines. However,
111 contaminating 3T3-L1 and 4T1 cells would outcompete any iPSC-like colonies, and
112 based on morphology, the cultures would rapidly revert to homogenous 3T3-L1 or 4T1
113 cultures. Hence, we sorted reprogrammed cells positive for ICAM and negative for
114 CD44 (ICAM+/CD44-), which marks the late stage of reprogramming in wild-type
115 MEFs (O'Malley *et al*, 2013). ICAM+/CD44- cells were sorted from the reprogrammed
116 3T3-L1 and 4T1 cells, and we detected small numbers of ICAM+/CD44- cells at D15
117 of reprogramming in 3T3-L1 and 4T1, but not N2a (**Figure 1c, d**). qRT-PCR of the
118 ICAM+/CD44- cells indicated that *Esrrb* and *Nanog* were up-regulated in the 4T1, and
119 *Essrb*, *Nanog,* and endogenous-*Pou5f1* in 3T3-L1 (**Figure 1e**). However, when the
120 sorted cells were replated they rapidly reverted to the original cell type morphology
121 and there was no evidence of iPSC-like cells (**Figure 1f**). Overall, these results
122 highlight the difficulty in reprogramming transformed cell lines to iPSCs.
123

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

125 We first explored wildtype MEF reprogramming to understand the blocks on
126 reprogramming in the normal case so that we can rule them out in the transformed cell
127 lines. Wildtype MEFs can only reprogram at early passages (Huang *et al*, 2020), and
128 the efficiency declines rapidly before complete failure after passage 4 (**Figure S1a**),
129 when using a conventional reprogramming OSKM (*Oct4*, *Sox2*, *Klf4*, and *Myc*) system
130 (Zhuang *et al.*, 2018). RNA-seq of MEFs from passages 1-6 revealed changes in cell
131 cycle genes, including the downregulation of cyclins and other positive cell cycle
132 regulators and the activation of cell cycle inhibitors, particularly *Cdkn1a* (**Figure S1b**).

133 We divided the samples into MEFs that could reprogram successfully (P1, 2) or failed
134 (P5, 6) and measured significantly differentially expressed (DE) genes. In total, 337
135 genes were significantly up-regulated and 419 significantly down-regulated (**Figure**
136 **S1c, d**). Gene ontology (GO) and gene set enrichment analysis (GSEA) of the
137 differentially expressed genes indicated that the down-regulated genes were related
138 to the cell cycle and extracellular matrix. Up-regulated genes were related to cell
139 migration/adhesion, MAPK activity, apoptosis, and inflammation (**Figure S1e, f**),
140 possibly caused by reactive oxygen species (**Figure S1f**). The inflammatory response
141 is exemplified by the up-regulation of cytokines and chemokines in the P5,6 MEFs
142 (**Figure S1g**). We defined the up-regulated gene set as the 'MEF senescent signature'
143 and the down-regulated gene set as a 'reprogramming permissive signature'
144 (**Supplementary Table S1**).
145

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

147 As the above-transformed cell lines have been maintained in culture for an extended
148 time they have likely accumulated many genetic alterations to adapt to the cell culture
149 environment. These changes may permanently impair their ability to reprogram.
150 Consequently, to achieve a controlled system we generated immortalized MEFs using
151 9 different combinations of factors to represent a spectrum of transformation. The
152 factors were chosen to cover different methods of immortalization: oncogenic
153 transcription factors (*Myc*, Hras^{G12V}, Mef2d, p53DD), viral transforming factors (SV40T,
154 E1A), apoptotic factors (Bcl2), and an engineered epigenetic factor (Hdac7SA; serines
155 at positions 178, 344, and 479 substituted with alanine, to block nuclear export) (Hong
156 *et al*, 2009; Luo *et al*, 2019; Seeburg *et al*, 1984). The sequences of the transformation

157 factors are in **Supplementary Table S2**. Not all of these factors could successfully
158 transform MEFs, and, except for SV40T, at least two factors were required for
159 successful transformation (**Figure S2a, b**). The immortalized cell lines were
160 maintained for at least 1 month to remove any non-transformed background MEFs.
161 After 1 month resulting lines still expressed the appropriate transgenes, as confirmed
162 by RNA-seq, RT-qPCR, and Western blot (**Figure S2c-e**). The exception was the
163 p53DD.Myc lines, as the p53DD transgene was silenced at the protein level, whilst
164 *Myc* could be detected by both RT-qPCR and Western blot (**Figure S2c-e**). This
165 suggests that both transgenes are required to generate immortal MEFs, but once
166 immortalized, p53DD becomes dispensable.

167 We next looked at the features of the transformed cells. All of the immortalized
168 MEFs proliferated faster than the wild-type MEFs (**Figure S2f**). We tested five lines
169 for the capability to form tumors in nude mice, Hras.E1A, SV40T, and Hras.Myc could
170 form tumors, whilst Bcl2.Myc and Hdac7SA.Mef2d could not (**Figure S2g, h**). Hoechst
171 staining of cross sections through the tumors indicated various differentiation layers,
172 although mainly mesoderm (**Figure S2i**). Aneuploidy is a common feature of cancer,
173 although the role of aneuploidy in transformation remains unclear (Chunduri &
174 Storchova, 2019). A normal karyotype is required for post-implantation embryonic
175 development (Shahbazi *et al*, 2020), but may not be incompatible with pluripotency
176 (Gaztelumendi & Nogues, 2014; Taapken *et al*, 2011; Yang *et al*, 2022). To rule out
177 the impact of karyotype abnormalities on reprogramming capability we confirmed a
178 normal karyotype for four of the lines that the subsequent study will mainly focus on
179 (**Figure S2j**).

180

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

182 Using this panel of immortalized MEF cell lines we reprogrammed them using a
183 lentiviral OSKM system with vitamin C, which is known to promote reprogramming
184 (Zhuang *et al.*, 2018). We used transformed lines derived from OG2 MEFs, which
185 contain an Oct4-GFP reporter (Zhuang *et al.*, 2018). Surprisingly, several lines could
186 give rise to GFP+ colonies. We labeled the lines as 'succeeding' (Hdac7SA.Mef2d,
187 p53DD.Myc), 'struggling' (Bcl2.Myc, Hdac7SA.E1A, p53DD.E1A, Hdac7SA.Myc), and
188 'failing' (Bcl2.E1A, Hras.E1A, Hras.Myc, and SV40T) (**Figure 2a, b**). It should be noted
189 that reprogramming the transformed lines was less efficient than wild-type MEFs,
190 except for the p53DD.Myc line. Complete reprogramming requires a passage before
191 the pluripotency gene expression program can be stably established. This process
192 partially relies on the fast-dividing iPSCs outcompeting the slow-growing/senescent
193 wildtype MEFs, however, the transformed MEFs were also highly proliferative (**Figure**
194 **S2f**) and would compete with the reprogrammed iPSCs. Hence, we FACS sorted the
195 GFP+ cells and passaged them to establish iPSC lines in the absence of transformed
196 MEFs. Pluripotency marker genes were robustly expressed only in the sorted cells
197 (**Figure 2c**), and we validated the expression of these genes by RT-qPCR and
198 Western blot (**Figure S3a, b**). Comparison of the RNA-seq to a panel of pluripotency
199 genes and somatic genes indicated that the gene expression levels were close to
200 ESCs (**Figure 2c**), and cross-correlation and principal component analysis (PCA)
201 indicated the cells were similar to ESCs (**Figure 2d, e**). DPre was used to compare
202 the gene expression in the GFP+ sorted lines to a panel of cell types (Hutchins *et al*,
203 2017; Steffens *et al*, 2020), and identified most lines as ESC-like (**Figure S3c, d**).
204 DPre also suggested that some ESC lines (e.g. Hdac7SA.Mef2d GFP+ line #4) were

205 a mixture of ESCs and contaminating immortalized MEFs (**Figure S3d**). We confirmed
206 that the iPSC lines could form teratomas with tissues representing all three germ
207 layers (**Figure S3e**). A mark of complete reprogramming is the silencing of the OSKM
208 transgenes, and the OSKM transgene was not detected in RNA-seq data (**Figure S3f**).
209 Interestingly, the immortalization factors were only partially silenced. All were silenced
210 except E1A in the Bcl2.E1A GFP+ line and Myc in the Bcl2.Myc GFP+ lines (**Figure**
211 **S3f**). We confirmed Hdac7SA and Mef2d were silenced by qRT-PCR (**Figure S3g**).
212 Overall, these data confirmed the successful generation of iPSC-like lines from some
213 transformed tumorigenic normal karyotype MEF lines.
214

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

216 We next explored the properties that made some immortal MEFs reprogramming-
217 capable and others –incapable. We focused on two (non-exclusive) models: the
218 inability to reprogram occurs due to problems in the originating MEFs, or it is caused
219 by failures to traverse the correct sequence of events required for successful
220 reprogramming.

221 Immortalization of somatic adult cells occurs by a range of mechanisms and is
222 accompanied by changes in the epigenetic state and gene expression patterns. We
223 analyzed the gene expression of the starting MEF cells and observed a wide range of
224 gene expression changes, with some MEF lines adopting substantial changes in gene
225 expression, whilst other cell lines were less affected (**Figure 3a**). There was no simple
226 correlation between the number of gene changes and reprogramming capability
227 (**Figure 3a**). Despite a large number of gene expression changes, there was no
228 evidence that the MEFs were transdifferentiating as their global gene expression

229 patterns continued to correlate well against wildtype MEFs (**Figure 3b**), and DPre
230 continued to identify the cells as MEF-like (**Figure S4a**). Gene ontology of the DE
231 genes suggested changes were associated with basic cellular processes such as up-
232 regulation of metabolic processes, and extracellular matrix genes, and down-
233 regulation of ribosomes, cell cycle, and DNA repair processes (**Figure 3c, S4b**), which
234 is the opposite of the senescent-like phenotype up-regulated in human arrested
235 embryos (Ying *et al*, 2008). Indeed, the MEF-senescent gene signature (defined in
236 **Figure S1c**) was not up-regulated in the transformed lines, and the majority of lines
237 had a significantly reduced MEF-senescent signature, reminiscent of the level in ESCs
238 (**Figure 3d**). This indicates the transformed lines are avoiding senescence.
239 Interestingly, the reprogramming-competent signature (defined in **Figure S1c**)
240 declines in all lines (except SV40T), which suggests a reduced capability to reprogram
241 (**Figure 3d**).

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

247 We next looked at specific genes that have been identified as key blocks or
248 promoters of the reprogramming process. Three factors related to cancer and cellular
249 transformation that also modulate reprogramming in wild-type cells are *Tp53*, *Cdkn1a,*
250 and *Rb1* (Retinoblastoma) (Hong *et al.*, 2009; Kareta *et al*, 2015; Kawamura *et al*,
251 2009; Marion *et al*, 2009; Tapia & Scholer, 2010). There was little change in *Rb1*
252 mRNA levels, and whilst *Cdkn1a* was relatively high in all cell lines except p53DD.E1A,

253 SV40T, and wildtype MEFs, its expression was not correlated with reprogramming
254 capability (**Figure S5a**). For *Tp53*, its expression levels were relatively consistent
255 across the MEF lines and did not correlate with the reprogramming ability of the lines
256 (**Figure S5a**). We inferred p53 activity by looking at known target genes of p53, and
257 all lines showed unaltered p53 activity except for the MEF line containing the
258 dominant-negative p53DD (p53DD.E1A), which had significantly down-regulated p53
259 target genes (**Figure S5b, c**). Loss of p53 is beneficial for reprogramming in wildtype
260 cells (Hong *et al.*, 2009; Marion *et al.*, 2009; Utikal *et al*, 2009), yet in transformed lines
261 reduced p53 activity in the p53DD.E1A line did not correlate with efficient
262 reprogramming (**Figure 2b**). This suggests that, in contrast to wild-type cells, reduced
263 p53 activity is not beneficial for reprogramming in transformed cells.
264

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

266 Reprogramming occurs in defined phases (Polo *et al.*, 2012; Samavarchi-Tehrani *et*
267 *al*, 2010; Xu *et al.*, 2016), so we explored if there are also changes in the phases of
268 reprogramming in the transformed lines. We performed RNA-seq during
269 reprogramming in the early phase (D3, D6), the mid-phase (D9, D12), and the late
270 phase (D15). These time points roughly correspond to three waves of gene expression
271 labeled initiation, maturation, and stabilization (Polo *et al.*, 2012; Samavarchi-Tehrani
272 *et al.*, 2010). The same waves were observed in our data, based on sets of genes
273 specific to the three phases (**Figure 4a, S6a**). Interestingly, different transformed lines
274 completed different phases of reprogramming and appeared to fail at distinct stages.
275 Lines that can successfully be reprogrammed: Hdac7SA.Mef2d, Bcl2.Myc and
276 Hdac7SA.Myc, closely resembled the wild-type MEF reprogramming process (**Figure**

277 **4a**). Hdacs7SA.E1A and p53DD.E1A progressed through maturation but struggled to
278 up-regulate stabilization phase genes (**Figure 4a**), although both can ultimately
279 generate GFP+ reprogrammed cells at very low efficiency (**Figure 2b**). For the
280 transformed lines that failed to reprogram, they failed at different stages. All cell lines
281 failed stabilization, SV40T and Hras.E1A failed maturation, but completed initiation,
282 whilst Hras.Myc failed all three stages (**Figure 4a, S6**).

283 The mesenchymal-epithelial transition (MET) is a key part of the early stage of
284 reprogramming (Li *et al*, 2010; Samavarchi-Tehrani *et al.*, 2010). We checked that the
285 transformed MEFs remained mesenchymal-like, and they still expressed typical
286 markers of the mesenchyme (**Figure 4b**). All transformed lines managed to navigate
287 the MET, except for Hras.Myc (**Figure S6b**). Indeed, some lines, including two
288 transformed lines that would ultimately fail to reprogram; Hras.E1A and SV40T, up-
289 regulated epithelial genes in advance of wild-type reprogramming cells (**Figure S6b**),
290 including two transformed lines that would ultimately fail to reprogram, Hras.E1A and
291 SV40T. This suggests that failure to execute a MET affected only the Hras.Myc line.

292 We next looked at changes in the overall phases of reprogramming in the early,
293 middle, and late stages of reprogramming. We measured the number of significantly
294 deregulated genes at each time point and used this as a proxy for the divergence from
295 a typical reprogramming time course. In this analysis, down-regulated genes
296 represent genes that fail to be correctly up-regulated at the specific day of the time
297 course, whilst up-regulated genes those represent those genes that are erroneously
298 high during reprogramming (**Figure 4c**). A large number of differentially regulated
299 genes on a specific day represents a divergence from the expected reprogramming
300 stage. As expected, the lines that reprogrammed at the highest efficiency (p53DD.Myc)

also had the lowest overall divergence (**Figure 4c**). Lines that could reprogram, but at low efficiency (Hdac7SA.Mef2d, Bcl2.Myc, Hdac7SA.E1A, and Hdac7SA.Myc) tended to have high initial divergence from the MEF state (day 0), but would correctly regulate reprogramming-associated genes at the later time points. Finally, lines that fail to reprogram (Hras.E1A and Hras.Myc), or reprogram exceptionally rarely (Bcl2.E1A), tended to have high divergence at all time points. The exception was the SV40T line, which appeared to follow the reprogramming gene expression program closely, but would diverge at day 12. Interestingly, GO analysis of the down-regulated genes (genes that should be up-regulated at that specific timepoint) highlighted the regulation of DNA repair genes, which was defective in both Hras.E1A and SV40T cell lines (**Figure 4d**). However, overall, there was a surprising diversity of changes in the transformed MEF lines that give rise to a range of defects at specific time points of the reprogramming process. This data indicates that transformation-specific effects drive line-specific blocks on reprogramming in a stage-dependent manner.

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

The analysis above suggested several pathways that could be manipulated, specifically, the MET, ribosome biogenesis, cell proliferation, cell adhesion, DNA repair, and apoptosis. We wondered if it was possible to intervene in the reprogramming process and convert the reprogramming-incapable to -capable. To investigate this we assayed the effect on reprogramming of a range of inhibitors on two reprogramming-capable lines (WT and Hdac7SA.Mef2d MEFs) compared to two reprogramming-incapable lines (Hras.E1A and SV40T MEFs). In total, we used 25

325 inhibitors targeting a range of pathways (**Figure S7a**). Whilst most inhibitors had no
326 effects or would ablate reprogramming, we observed that 5 inhibitors promoted the
327 reprogramming of one or more of the transformed reprogramming-incapable lines
328 (**Figure S7a**). The five small molecules target: MEK1/2 (PD; PD0325901), GSK3
329 (CHIR; CHIR99021), ROCK (Y; Y23637), G9a (BIX; BIX-01294), and histone
330 deacetylases (TSA). We tested the inhibitors in combinations and found that the most
331 efficient combination was CHIR, PD and Y (**Figure 5a, S7b**). However, only PD was
332 essential to convert SV40T and Hras.E1A to reprogramming-capable, as
333 serum+Vc+PD alone resulted in a small number of GFP+ cells. Conversely, CHIR, Y,
334 and BIX alone were incapable of converting Hras.E1A MEFs to reprogramming-
335 capable (**Figure S7b**). Interestingly the inhibitor cocktails had effects on
336 reprogramming in a transformed line-specific manner, as the addition of PD to the
337 reprogramming cocktail converted the normally reprogramming-capable Bcl2.E1A line
338 to incapable. This highlights the line-specific nature of the reprogramming barriers and
339 shows that overcoming a barrier in one line can potentially initiate a barrier in another
340 transformed line.

341 We confirmed the resulting iPSC-like cells derived from Hras.E1A or SV40T
342 MEFs reprogrammed with the PD, CHIR, Y and BIX adopted a normal morphology,
343 expressed pluripotent marker genes by qRT-PCR and Western blot (**Figure 5d, and**
344 **Figure S7c**. We sorted the GFP+ cells and established iPSC lines, and measured
345 their gene expression which closely correlated with ESCs by correlation and PCA
346 (**Figure 5e and Figure S8a, b**). It should be emphasized that the chemical cocktail
347 did not improve (or inhibit) wildtype MEF reprogramming, as wildtype cells
348 reprogrammed with similar rates in serum+Vc or the inhibitor cocktail (**Figure S7b**).

349 This emphasizes that the blocks seen in transformed cells are specific to
350 transformation and are not present in wild-type untransformed cells. Overall, the
351 results indicate that transformation-specific pathways are impairing the ability to
352 reprogram.
353

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

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

360 We explored the epigenetic regulation of the transformed MEFs. As a proxy for
361 overall epigenetic activity, we looked at the expression levels of epigenetic factors
362 involved in activation, repression, and the reading of epigenetic marks, as defined by
363 the Epifactors database (Medvedeva *et al*, 2015). As expected, epigenetic-related
364 factors were uniformly significantly up-regulated in ESCs compared to MEFs (**Figure**
365 **S9a**), reflecting their more complex epigenetic regulation (He *et al*, 2019). For the
366 transformed MEFs, activators were more often significantly up-regulated, compared
367 to untransformed MEFs (**Figure S9a**). Erasers and readers were only up-regulated in
368 several transformed lines, all containing E1A: Hras.E1A and Bcl2.E1A and
369 p53DD.E1A (**Figure S9a**). Nonetheless, whilst the different classes of epigenetic
370 regulators varied across the transformed lines and did not discriminate
371 reprogramming-capable from incapable, the general pattern for the majority of lines
372 was increased expression of epigenetic regulators. Western blot of repressive histone

373 modifications (e.g. H3K27me3, H3K9me3) tended to stay the same in the different
374 lines (**Figure 6a**). However, histone acetylation inversely correlated with
375 reprogramming capability as H3K27ac, H3ac or H4ac was high in reprogramming-
376 incapable lines (Hras.Myc, Hras.E1A and SV40T). The results support epigenetic
377 deregulation as a feature of the transformed lines and suggest increased epigenetic
378 activation through histone acetylation is a feature of transformed reprogramming-
379 incapable MEFs.
380

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

383 To explore changes in the epigenetic state of the MEFs we performed ATAC-seq on
384 three transformed lines: Hdac7SA.Mef2d (reprogramming-capable), Hras.E1A
385 (initially incapable, but can be induced to reprogram with MEK inhibition), and
386 Hras.Myc (reprograming-incapable) and compared them to wildtype MEFs. Heatmaps
387 of open and closed chromatin indicated, surprisingly, that the Hdac7SA.Mef2d lines
388 were distal from all of the other transformed cell lines and had the greatest number of
389 specific peaks (**Figure S9b, Figure 6b**). The loci that were opened or closed in the
390 Hdac7SA.Mef2d lines were significantly correlated with matching changes in gene
391 expression (**Figure S9c**). Intriguingly, the overall ATAC signal was substantially higher
392 in all three transformed cell lines compared to wildtype MEFs (**Figure 6c**), and this
393 effect was despite fewer open loci in the Hras.Myc and Hras.E1A lines (**Figure 6b**).
394 This suggests that the transformed MEFs have increased levels of open chromatin.

395 We next focused on Hras.Myc and Hras.E1A lines, and removed the
396 Hdac7SA.Mef2d from further analysis as it is capable of reprogramming under normal

397 conditions and its ATAC-seq genome-wide pattern was distal from the other
398 transformed cell lines and wildtype MEFs. Hras.Myc and Hras.E1A are interesting as
399 they have relatively similar ATAC-seq patterns (**Figure S9b**), yet Hras.E1A can be
400 converted to reprogramming-capable, whilst Hras.Myc cannot. The two lines have
401 Hras, which does not impede reprogramming independently. Similarly, other lines co-
402 transformed with Myc are also reprogramming-capable (Bcl2.Myc and Hdac7SA.Myc),
403 as E1A transformed lines (Hdac7SA.E1A). Hence, the individual transgenes are not
404 incompatible with reprogramming, only the specific combinations of Hras.E1A and
405 Hras.Myc is important. One key difference between Hras.E1A and Hras.Myc is that
406 only Hras.E1A up-regulated initiation-stage genes and underwent an MET (**Figure 4a,**
407 **d**), suggesting the impairment of reprogramming is already primed in the starting
408 transformed MEFs.

409 To explore the underlying differences, we performed ATAC-seq on Hras.Myc
410 and Hras.E1A lines with and without PD treatment, and compared them to wildtype
411 MEFs. Interestingly, the global upregulation of the ATAC-seq signal at open loci was
412 lost in the PD-treated samples and returned to levels similar to the wild-type (**Figure**
413 **6d**). Western blot of H3K27ac, a mark of open active chromatin, supported this
414 observation as H3K27ac was higher in the transformed MEFs, and treatment with PD
415 for 3 days reduced H3K27ac to wildtype levels (**Figure 6e**). Interestingly the result is
416 transient, as treatment for 2 days followed by washing out reverted H3K27ac levels to
417 the hyperactive state (**Figure 6e**). This suggests that when PD inhibits MEK it leads
418 to the indirect closing of chromatin and reductions in the global levels of H3K27ac.
419 However, this effect was transient, and incomplete as the overall epigenetic landscape

420 did not revert to a wildtype state and 34,913 loci open in wildtype MEFs remained
421 closed in the transformed cell lines (**Figure 6f**).

422 To explore the mechanisms underlying the action of PD, we next divided the
423 loci into two types based upon their state in wildtype MEFs and the effect of PD: (1)
424 those loci that PD opens and reverts to a wildtype state, and (2) those loci that PD
425 closes, reverting to the wildtype state (**Figure 6g**). There was a roughly equal split
426 between the two classes (10,230 peaks versus 12,276 peaks; **Figure 6g**). GO analysis
427 of the peaks that were specifically reverted to wildtype state in the Hras.Myc cells
428 suggested that cell-cell adhesion gene loci were being remodeled (**Figure 6h**). Indeed,
429 chromatin at *Cdh1* and *Epcam* (two epithelial genes) were open in the Hras.E1A and
430 Hras.Myc MEFs, but were closed when the cells were treated with PD and reverted to
431 a wild-type MEF pattern at these genes (**Figure 6i**). Importantly, *Cdh1*, *Epcam,* and
432 other epithelial genes are not expressed in the transformed MEFs (**Figure 6j**),
433 indicating that epigenetic regulation at epithelial gene loci is deleterious for
434 reprogramming of transformed cell lines.

435 We next looked at possible downstream targets of MEK signaling that mediated
436 this indirect effect on chromatin. We used motif analysis to look for changes in
437 transcription factor occupancy at PD-modulated peaks (**Figure 6k**). Interestingly,
438 peaks that were closed in response to PD treatment were specifically enriched with
439 ETS, E2A, and KLF motifs. KLF proteins are major regulators of MET processes and
440 are pluripotent factors, whilst ETS factors have been identified as barriers for wild-type
441 untransformed reprogramming (Li *et al*, 2017a). Loci that PD treatment opens were
442 enriched with REL, HOX, and TFE family TFs (**Figure 6k**). To explore deeper we
443 looked at two small groups of chromatin loci that distinguished reprogramming-

444 capable from -incapable, 10001 (specifically open in WT and Hras.E1A+PD) and
445 01110 (specifically closed in WT and Hras.E1A+PD) (**Figure 6g**). Interestingly, GO
446 analysis of the 01110-group suggested that enhancers near to genes related to cell
447 differentiation were erroneously opened (**Figure S10a**), for example at the
448 developmental-related genes *Isl1* and *Tcf7l2* (**Figure S10b**). This was a purely an
449 effect on chromatin state as *Isl1* and *Tcf7l2* were below the threshold for detection in
450 the RNA-seq data. This pattern extended to all of the genes in the vicinity of the 01110-
451 chromatin state which were not significantly up or down-regulated (**Figure S10c**).
452 Conversely the genes associated with the 10001 group were significantly down-
453 regulated (**Figure S10c**), although GO analysis suggests they are related to basic
454 cellular processes such as cytoskeleton organization (**Figure S10d**). Overall, this
455 suggests the epigenetic activation of developmental genes is deleterious for
456 reprogramming, and MEK inhibition can indirectly modulate the epigenetic state to
457 close chromatin.
458

459 **Discussion**

460 Reprogramming cancerous/transformed cells to pluripotent cells, followed by their
461 differentiation, could form a valuable experimental model to recapitulate the earliest
462 stages of tumorigenesis (Kim & Zaret, 2015; Suknuntha *et al*, 2015). However,
463 reprogramming of primary cancerous tissues is often difficult or impossible for many
464 cancers, and even in possible cases the efficiency is considerably less than wild-type
465 cells. Additionally, there remains doubts about the *bona fide* reprogramming of the
466 cancerous cells (Kumar *et al.*, 2016; Lee *et al.*, 2017; Stricker *et al.*, 2013), and it is
467 not clear if reprogramming of primary human cells is complete. A significant problem

468 is the massive heterogeneity seen in primary cancer cells. To overcome this, we
469 mainly focused on isogenic transformed MEF lines that allow us to carefully control
470 the overall heterogeneity. Although, even in this system there was a surprisingly large
471 variation between the different methods of MEF transformation. Consequently, a clear
472 understanding of the barriers that impair the reprogramming of transformed cells
473 would potentially shed light on this area.

474 There is an intimate but unclear link between reprogramming and
475 transformation (Iglesias *et al.*, 2017; Kim & Zaret, 2015). This link was recently
476 explored in a novel system with multiple inducible genes that could be manipulated to
477 either transform MEFs or perform reprogramming to iPSCs (Huyghe *et al.*, 2022). This
478 elegant exploration showed that reprogramming and transformation followed similar
479 initial paths, but then later diverged. The system they used for transformation was K-
480 ras^{G12D} and *Myc* overexpression (Huyghe *et al.*, 2022), which is similar to the Hras.Myc
481 (*Hras^{G12V}*/*Myc*) line described in this manuscript. Intriguingly, we show that Hras.Myc
482 is not compatible with reprogramming, and those cells fail very early in the
483 reprogramming process. However, several other lines described here can be
484 transformed and remain compatible with reprogramming. Potentially, transformation
485 strategies that remain compatible with reprogramming may recapitulate more of the
486 reprogramming process as they transform, and only diverge from reprogramming at
487 later stages.

488 *Ras* overexpression is beneficial for reprogramming wildtype MEFs, but is a
489 major barrier to reprogramming transformed cells (Ferreiros *et al*, 2019). Indeed, when
490 Hras^{G12V} was overexpressed in p53-null immortalized MEFs reprogramming was
491 blocked (Ferreiros *et al.*, 2019). This is reminiscent of the effect we observed in the

492 Hras.Myc and Hras.E1A lines, which are also resistant to reprogramming. However,
493 we show that inhibition of MEK by PD can only enable Hras.E1A cells to reprogram,
494 and Hras.Myc cells remain reprogramming-incapable. This indicates that $Hras^{G12V}$ is
495 context-dependent when its co-transformation factor is different.

496 Several studies have identified small molecules that assist in reprogramming
497 transformed cells to a more pluripotent-like state. ROCK inhibition and rapamycin can
498 promote the expression of a *NANOG* reporter in a breast cancer cell line (MDA-MB-
499 468) (Yuan *et al*, 2018). However, how completely the *NANOG*-positive cells were
500 reprogrammed is not clear, as they expressed high levels of ectoderm-specific marker
501 genes *SOX1*, *NESTIN,* and *PAX6*. mTOR was also identified as a barrier to the
502 reprogramming of sarcoma cells (Ito *et al*, 2022). In our experiments, mTOR inhibition
503 had no effect, and ROCK inhibition worked only to boost reprogramming in
504 cooperation with MEK inhibition. This again highlights the context-specific nature of
505 the barriers in different transformed cell lines. Overall, MEK inhibition was the
506 dominant factor, which agrees with a previous study that showed that MEK inhibition
507 assisted in reprogramming human transformed cell lines (Kong *et al*, 2020). PD is also
508 a key component of the 2iLIF medium used to maintain mouse ESCs in a naïve
509 groundstate (Silva *et al*, 2008; Ying *et al.*, 2008). It is unclear if PD plays the same role
510 in assisting in reprogramming transformed MEFs and in promoting pluripotency.
511 Indeed, reprogramming with PD is deleterious for reprogramming wild-type MEFs, at
512 least in the early stages of reprogramming (Ferreiros *et al.*, 2019; Shi *et al*, 2008).

513 Ultimately, these studies support a model for the context-specific
514 reprogramming of transformed cell lines. Different transformations lead to diverse
515 blocks at different stages of the reprogramming process. Inhibition of ERK by PD can

516 overcome some of these barriers, primarily in Ras-transformed cells (Ferreiros *et al.*,
517 2019; Kong *et al.*, 2020). However, each transformed cell line can harbor defects that
518 block reprogramming at unique stages. This perhaps reflects the diversity of regulatory
519 features disrupted in cancer and may relate to the 'stemness' of a cancer type or
520 method of transformation (Malta *et al*, 2018). Whilst there are many surprising
521 similarities between tumorigenesis and reprogramming, the nature of the relationship
522 remains unclear (Iglesias *et al.*, 2017). Potentially, insights into reprogramming can
523 inform the process of transformation and provide new models to address
524 tumorigenesis and a clearer understanding of cell type plasticity in cancer
525 development.

526

527 **METHODS SUMMARY**

528 **Animal experimental approval**

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

532

533 **Cell culture and MEF transformation**

534 MEFs were derived from E13.5 embryos from OG2 mice (Zhuang *et al.*, 2018), or a
535 wild-type ICR genetic background. MEF cells were cultured in DMEM high-glucose
536 media containing 10% FBS (GIBCO), 1 × GlutaMAX (GIBCO), and 1x nonessential
537 amino acids (GIBCO), 0.5x Penicillin/Streptomycin (Hyclone). We generated
538 immortalized MEF cell lines by transfecting the following transgenes singly or in
539 combination: P53DD, PMM2-SV40T, PMM2-Hdac7SA, PMM2-Mef2d, Bcl2, PMM2-

540 Myc, Ras, PMM2-E1A. Immortalized MEFs were passaged for at least a month to
541 remove wild-type untransformed MEFs.
542
543 **Somatic cell reprogramming**
544 Wildtype MEFs and transformed MEFs in an OG2 reporter background were
545 reprogrammed as described in (Zhuang *et al.*, 2018). Briefly, OKSM in lentivirus was
546 transfected into 15,000 MEFs in one well of a 12-well plate. One day after transfection
547 the medium was changed to reprogramming medium: DMEM high-glucose media
548 containing 15% FBS (GIBCO), 1× GlutaMAX (GIBCO), 1x non-essential amino acids
549 (GIBCO), 0.5x Penicillin/Streptomycin (Hyclone), 1mM sodium pyruvate (GIBCO),
550 0.1 mM 2-mercaptoethanol (GIBCO), 1000 U/ml leukemia inhibitory factor (LIF)
551 (Millipore) and vitamin C (50 μg/ml, Sigma), as indicated in the figures. The medium
552 was changed daily. Inhibitors were added when the medium was changed to
553 reprogramming medium: PD0325901 (1 μM, Sigma), CHIR99021 (3 μM, Sigma),
554 Y23637 (10 μM, Sigma), BIX-01294 (200 nM, Selleck), and TSA (50 nM, Sigma).
555
556 **Karyotype analysis**
557 When the immortalized MEF cell growth density reached 80 to 90%, colchicine (0.2
558 μg/ml, Selleck), was added to the culture medium to a final concentration of 0.2 μg/ml.
559 The cells were incubated at 37°C for 120 min. After colchicine treatment, cells were
560 washed twice with PBS, and 0.5 ml of 0.25% trypsin was added for digestion. Then, 7
561 ml of 0.075 mol/L KCl solution preheated to 37°C was added, and the cell suspension
562 was blown with a straw and incubated at 37°C in a water bath for 25 mins. Collected
563 cells were fixed with a fixing solution (methanol glacial acetic acid 3:1 preparation) at

564 37°C, 3 min. After fixation, cells were centrifuged at 1200 revolutions per min (rpm) for
565 5 min, after which the supernatant was discarded. 7 ml of fresh fixing solution was
566 added, and the cells were gently beaten with a pipette tip, and fixed in a 37°C water
567 bath for 40 mins. Cells were centrifuged at 1200 rpm for 5 min, after which the
568 supernatant was discarded. Subsequently, 7 ml of fresh fixing solution was added,
569 followed by gentle beating with a filter tip. The cells were fixed in a 37°C water bath
570 for 40 min. After fixation the cells were centrifuged at 1200 rpm for 5 min, and then
571 most of the fixation solution was removed, and the cells were resuspended with part
572 of the fixation solution. Suspended cells were pipetted onto a slide from a distance of
573 about 30 cm. Immediately after dropping the cells on the slide, move the slide into the
574 oven at 75°C and bake for 3h. Next, 0.03 g trypsin powder to 55 ml normal saline was
575 added, followed by a gentle shaking, before adjusting the pH to about 7.2 with 3%
576 Tris-HCl. The preparation was put into trypsin digestion solution for 8-10 seconds,
577 followed by the addition of saline to stop digestion, then add Giemsa staining (C0133,
578 Beyotime Biotech) solution was added for 5-10 minutes, and then pick out the glass
579 slide with tweezers. Both sides were gently rinsed with running water and allowed to
580 dry at room temperature. After the slides dried, they were examined under a
581 microscope to look for a good chromosome split.

582

583 **Teratoma formation**

584 5-10 million mouse PSCs in a slurry of Matrigel and mTeSR (1:1) medium were
585 injected into 8-week-old female nude (BALB/cNj-Foxn1^{nu}/Gpt) immunodeficient mice.
586 Teratoma growth was quantified by measuring approximate elliptical area (mm^2) with
587 calipers measuring the outward width and height after growth for 60 days.

588 Representative tumors were dissected and sectioned and slices were stained with
589 hematoxylin and eosin.
590
591 **Alkaline phosphatase staining**
592 Reprogrammed cells were stained with alkaline phosphatase according to the
593 manufacturer's protocols. In brief, reprogramming cells were fixed in 1% (w/v)
594 formaldehyde, and then cells were stained with BCIP/NBT Alkaline Phosphatase Color
595 Development Kit (C3206, Beyotime Biotech) according to the kit's instructions.
596
597 **Flow cytometry**
598 Reprogramming cells were digested with trypsin and washed with DPBS once, and
599 analyzed or sorted with BD FACS Aria III flow cytometer. To monitor cellular
600 reprogramming status, reprogrammed cells were stained with antibodies: anti-CD44-
601 PE (25-0441-82, 1:500, Thermo Fisher) and anti-ICAM-1-FITC (sc-8439 FITC, 1:200,
602 Santa Cruz Biotechnology). Cells were sorted and analyzed on a BD FACS Aria III
603 instrument.
604
605 **Western blot**
606 Cells were lysed with RIPA buffer (P0013B, Beyotime Biotech). Proteins were
607 resolved in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
608 and transferred onto pre-activated polyvinylidene fluoride (PVDF) membranes
609 (IPVH00010, Millipore, MA, USA). The PVDF Membranes were incubated with anti-
610 H3 (ab1791, 1:3000, Abcam), anti-H3K27ac (ab4279, 1:2000, Abcam), anti-H3ac (06-
611 599,1:2000, Millipore), anti-H4ac (ab46983, 1:2000, Abcam), anti-H3K4me3 (ab8580,

612 1:2000, Abcam), anti-H3K4me (ab176877,1:2000, Abcam), anti-H3K27me3 (07-
613 449,1:2000, Millipore), anti-H3K9me3 (ab8898, 1:2000, Abcam), anti-NANOG (8822S,
614 1:1000, cell signaling), anti-OCT4 (sc-5279, 1:1000, Santa Cruz Biotechnology), anti-
615 P53 (sc-126, 1:1000, Santa Cruz Biotechnology), anti-HDAC7 (ab12174,1:2000,
616 Abcam), anti-MEF2D (ab32845,1:2000, Abcam), anti-ACTIN (ab8227, 1:3000,
617 Abcam), anti-MYC (ab32074,1:2000, Abcam), anti-SOX2 (4900S, 1:1000, cell
618 signaling). Afterward, the membranes were incubated with HRP-conjugated goat
619 antirabbit IgG (ab205718, 1:3000, Abcam) and visualized using an enhanced
620 chemiluminescence BeyoECL method (P0018AS, Beyotime Biotech).
621

622 **Immunofluorescence**

623 Cells were washed with cold PBS and fixed using 4% (w/v) paraformaldehyde in PBS
624 for 10 minutes at room temperature. Then cells were washed with PBS and
625 permeabilized with 0.25% (v/v) Triton X-100 for 10 minutes at room temperature. Fixed
626 cells were then washed with PBS three times and blocked with 3% (w/v) bovine serum
627 albumin for 1 hour at room temperature. Then cells were incubated with primary
628 antibodies: anti-NANOG (8822S, 1:1000, cell signaling), at 4°C overnight in primary
629 antibody. Following overnight incubation, cells were washed with wash buffer (0.1%
630 (v/v) Tween-20 in PBS) and incubated with secondary antibodies Alexa Fluor series
631 (Life Technologies) in wash buffer for 1 hour at room temperature. Cells were washed
632 again and nuclei were stained using DAPI (5 μg/ml, Thermo).
633

634 **RT-qPCR**

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

640

641 **RNA-seq preparation and analysis**

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

645 RNA-seq was analyzed essentially as described in (Hutchins *et al.*, 2017),
646 except scTE/te_counts was used to assign mapped reads to genes and TEs (He *et al*,
647 2021), and the UCSC genome browser repeatmask track and GENCODE vM23
648 annotation were used for TE and gene assignment. Differential expression was
649 determined using DESeq2 (Love *et al*, 2014). A gene was considered differentially
650 expressed had an absolute fold-change of at least 2, and a Bonferroni-Hochberg
651 corrected p-value of 0.01. GSEA was performed using fgsea (Korotkevich *et al*, 2021),
652 and GO was done with goseq (Young *et al*, 2010). Cell type determination was scored
653 with DPre (Steffens *et al.*, 2020). Other analyses were performed using glbase3
654 (Hutchins *et al*, 2014).

655

656 **ATAC-seq preparation and analysis**

657 ATAC-seq library was generated using the Tn5 enzyme from the
658 TruePrep DNA Library Prep Kit V2 (Vazyme, TD501-02) as previously described
659 (Buenrostro *et al*, 2013). Briefly, a total of ~50,000 cells were washed once with 50 µl
660 of cold PBS and resuspended in 50 µl lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM
661 NaCl, 3 mM MgCl2, 0.2% (v/v) IGEPAL CA-630). The nuclei were centrifuged for 10
662 min at 500 × g at 4 °C, followed by the addition of 50 µl transposition reaction mix (25
663 µl TD buffer, 2.5 µl Tn5 transposase (Vazyme), and 22.5 µl nuclease-free water).
664 Samples were PCR-amplified and purified using a MinElute kit (Qiagen). After
665 selecting an appropriate PCR cycle number (See (Buenrostro *et al.*, 2013)) samples
666 were sequenced on an Illumina sequencer. ATAC-seq data was analyzed essentially
667 as described in reference: (Li *et al.*, 2017a). Briefly, reads were aligned to the mm10
668 mouse genome with bowtie2 (Langmead & Salzberg, 2012), peaks were called with
669 MACS2 (Zhang *et al*, 2008), and then the *redefine_peaks* function, which is a
670 generalized reimplementation of the algorithm in (Li *et al.*, 2017a), was used to recover
671 low-scoring peaks by sharing peak information across samples (Ma *et al*, 2021). DNA
672 binding motifs were detected using HOMER (Heinz *et al*, 2010). All other analyses
673 were performed using glbase3 (Hutchins *et al.*, 2014). ATAC-seq data from
674 GSE93029 (Li *et al.*, 2017a) and GSE103980 (Malik *et al*, 2019) were reanalyzed as
675 part of this study.

676

677 **Statistical analysis and biological replication**

678 Differential gene expression was calculated using DESeq2 (v1.36.0). A gene was
679 considered significantly differentially regulated if it had an absolute fold-change of at

least 2 and a Bonferroni-Hochberg corrected p-value (q-value) of <0.01. This criterion was used in **Figures 3a, 4c, Figures S1c, and d**. RNA-seq experiments were performed in at least biological duplicate (different samples on different days, or independent cell lines). Gene ontology analysis was performed using goseq (v1.48.0) and statistics were calculated using goseq's internal statistical model. A gene ontology category was considered significantly enriched if there were at least 50 genes in that GO term and a Bonferroni-Hochberg corrected p-value (q-value) of <0.01. GSEA was performed using fgsea (v1.22.0). Gene sets were considered enriched or depleted if they had an absolute NES (normalized enrichment score) of at least 1.5 and a Bonferroni-Hochberg corrected p-value (q-value) of <0.01. Western blots were repeated at least twice with similar results. **Figure 6a** was repeated three times. FACS analysis (**Figure 1c**) was performed twice. qRT-PCR was performed on at least three biological replicates with three technical replicates each. Transcription factor motif analysis was performed using HOMER. A motif was considered significantly enriched if the uncorrected p-value was <0.00001.

Acknowledgments

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

Author contributions

X.F. designed and performed the majority of the experiments, prepared the manuscript figures and revised the text. Q.Z. designed the project and planned the experiments. 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

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

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

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

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

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

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

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

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

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

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* (2019) Context-Dependent Impact of RAS Oncogene Expression on Cellular Reprogramming to Pluripotency. *Stem Cell Reports* 12: 1099-1112

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Kong Y, Gimple RC, McVicar RN, Hodges AP, Yin J, Liu Y, Zhan W, Snyder EY (2020) "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* 15: 761-775

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

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

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

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

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

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

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

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

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

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

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

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

834 Ma G, Babarinde IA, Zhuang Q, Hutchins AP (2021) Unified Analysis of Multiple ChIP-Seq Datasets.
835 *Methods Mol Biol* 2198: 451-465

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

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

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

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

848 Miyoshi N, Ishii H, Nagai K, Hoshino H, Mimori K, Tanaka F, Nagano H, Sekimoto M, Doki Y, Mori M
849 (2010) Defined factors induce reprogramming of gastrointestinal cancer cells. *Proc Natl Acad Sci U S*
850 *A* 107: 40-45

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

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

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

859 Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei
860 S, Cloutier J *et al* (2012) A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 151:
861 1617-1632

862 Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL
863 (2010) Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation
864 of somatic cell reprogramming. *Cell Stem Cell* 7: 64-77

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

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

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

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

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

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

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

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

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

887 Taapken SM, Nisler BS, Newton MA, Sampsell-Barron TL, Leonhard KA, McIntire EM, Montgomery KD
888 (2011) Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. *Nat*
889 *Biotechnol* 29: 313-314

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

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

894 Utikal J, Polo JM, Stadtfeld M, Maherali N, Kulalert W, Walsh RM, Khalil A, Rheinwald JG, Hochedlinger
895 K (2009) Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature*
896 460: 1145-1148

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

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

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

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

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

Yuan J, Zhang F, You M, Yang Q (2018) Identification of protein kinase inhibitors to reprogram breast cancer cells. *Cell Death Dis* 9: 915

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

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

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

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

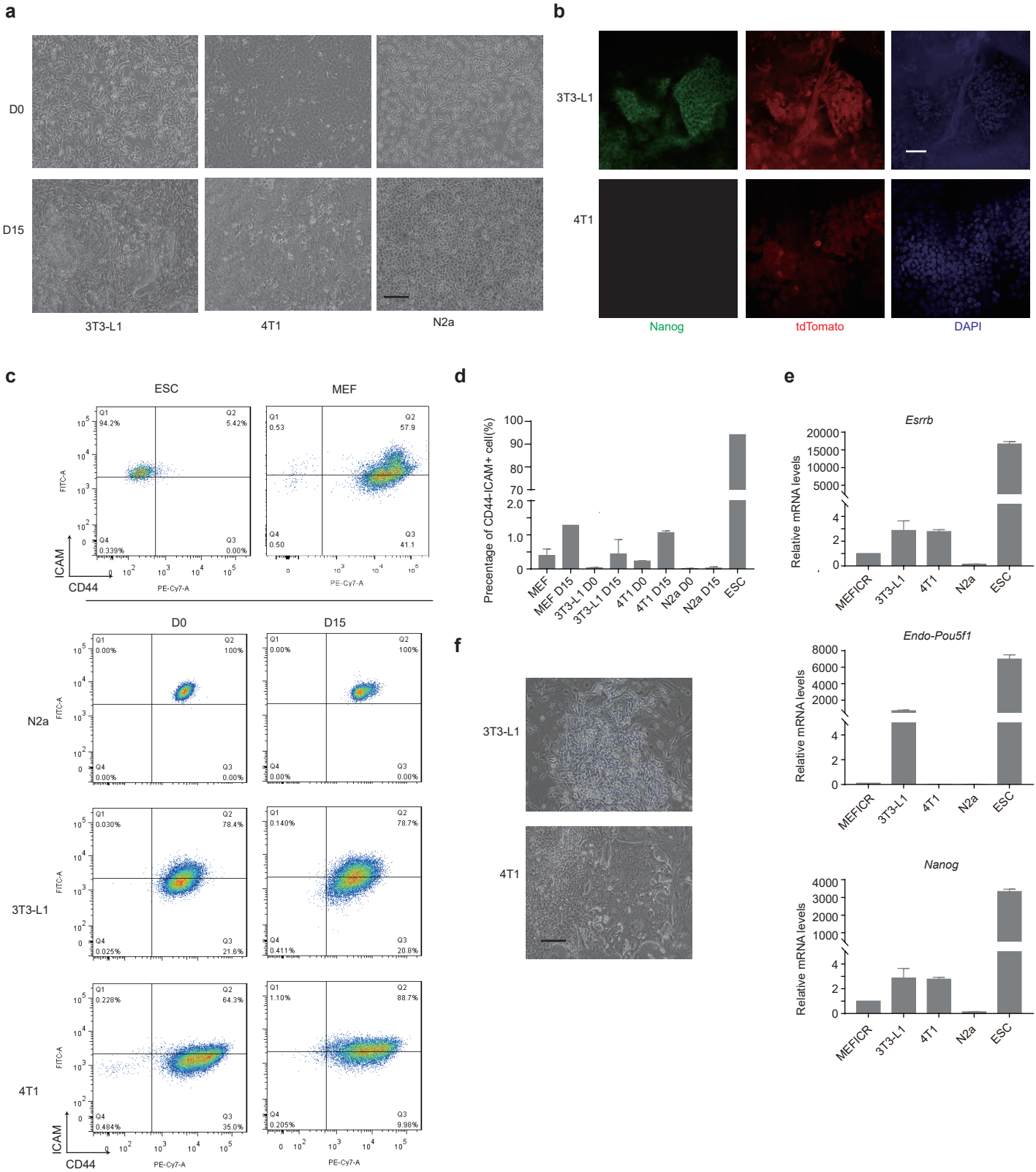
a
D0
D15
3T3-L1
4T1
N2a
b
3T3-L1
4T1
Nanog
tdTomato
DAPI
c
ESC
MEF
D0
D15
N2a
3T3-L1
4T1
ICAM
CD44
FITC-A
PE-Cy7-A
d
Precentage of CD44-ICAM+ cell(%)
MEF
MEF D15
3T3-L1 D0
3T3-L1 D15
4T1 D0
4T1 D15
N2a D0
N2a D15
ESC
e
Esrrb
Endo-Pou5f1
Nanog
Relative mRNA levels
MEFICR
3T3-L1
4T1
N2a
ESC
f
3T3-L1
4T1

Figure 1

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

924 **a** Bright-field images of the reprogramming of 3T3-L1, 4T1, and N2a cell lines in
925 serum+Vc reprogramming conditions. Images are shown from day 0 (D0) or day
926 15 (D15) of the reprogramming experiment. Morphologically iPSC-like cells can
927 typically be observed at D15 and onwards in wild-type reprogramming. Scale
928 bar=100 μm.

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

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

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

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

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

943

944

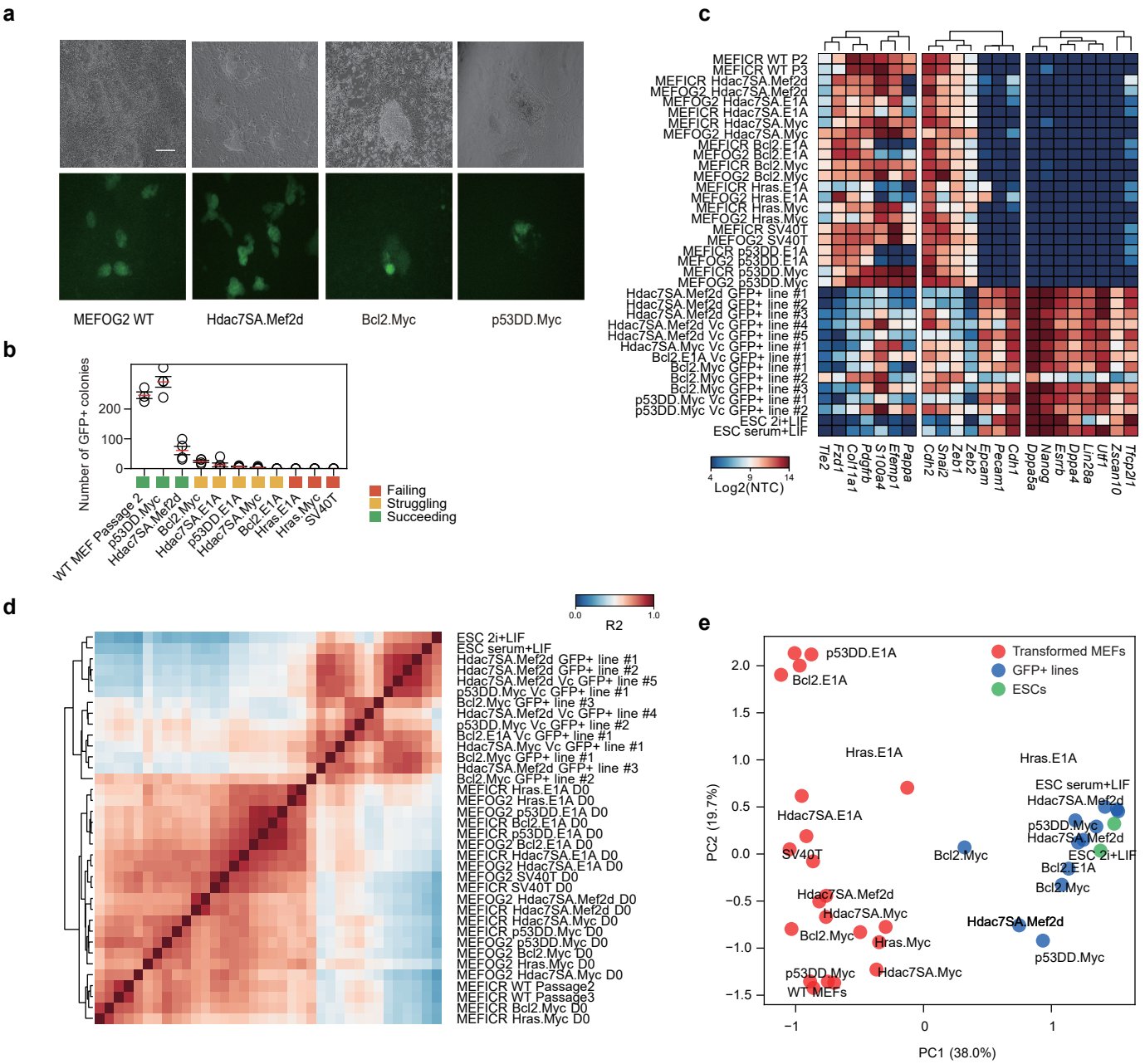
a
MEFOG2 WT
Hdac7SA.Mef2d
Bcl2.Myc
p53DD.Myc
b
Number of GFP+ colonies
200
100
0
WT MEF Passage 2
p53DD.Myc
Hdac7SA.Mef2d
Bcl2.Myc
Hdac7SA.E1A
p53DD.E1A
Hdac7SA.Myc
Bcl2.E1A
Hras.E1A
Hras.Myc
SV40T
Failing
Struggling
Succeeding
c
MEFICR WT P2
MEFICR WT P3
MEFICR Hdac7SA.Mef2d
MEFOG2 Hdac7SA.Mef2d
MEFOG2 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

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

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

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

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

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

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

962

963

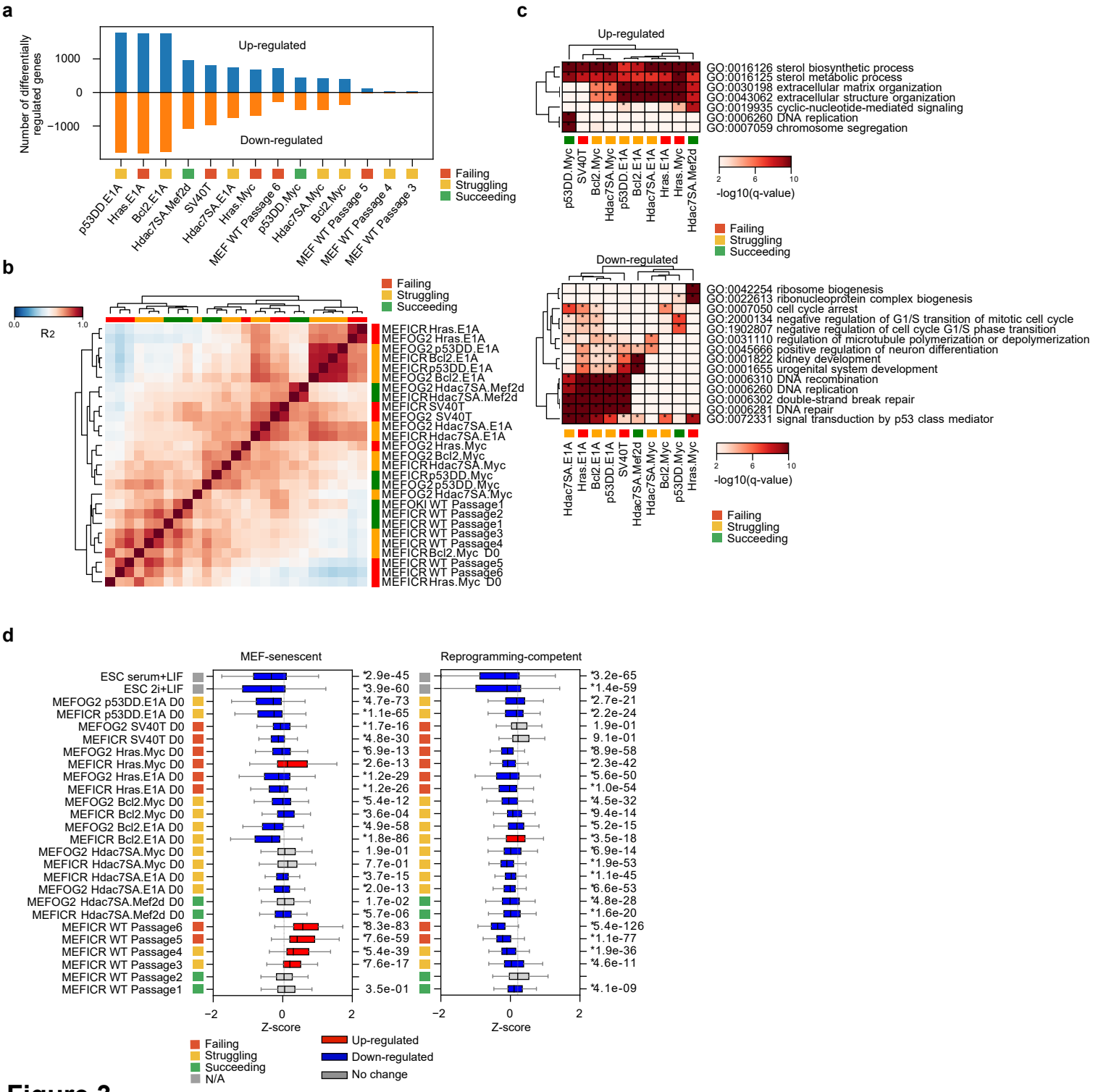
a
Number of differentially regulated genes
1000
0
−1000
Up-regulated
Down-regulated
p53DD.E1A
Hras.E1A
Bcl2.E1A
Hdac7SA.Mef2d
SV40T
Hdac7SA.E1A
Hras.Myc
MEF WT Passage 6
p53DD.Myc
Hdac7SA.Myc
Bcl2.Myc
MEF WT Passage 5
MEF WT Passage 4
MEF WT Passage 3
Failing
Struggling
Succeeding
b
R2
0.0
1.0
Failing
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

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

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

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

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

972 **d** Boxplots for the genes defined as the MEF-senescent signature or the
973 reprogramming-capable signature from **Figure S1c**. Significance is from a two-
974 sided Welch's t-test for pair-wise comparisons versus MEFICR wildtype Passage
975 2 as the reference. A change was considered significant if the p-value was < 0.01.
976 Significant changes are colored red or blue.

977

978

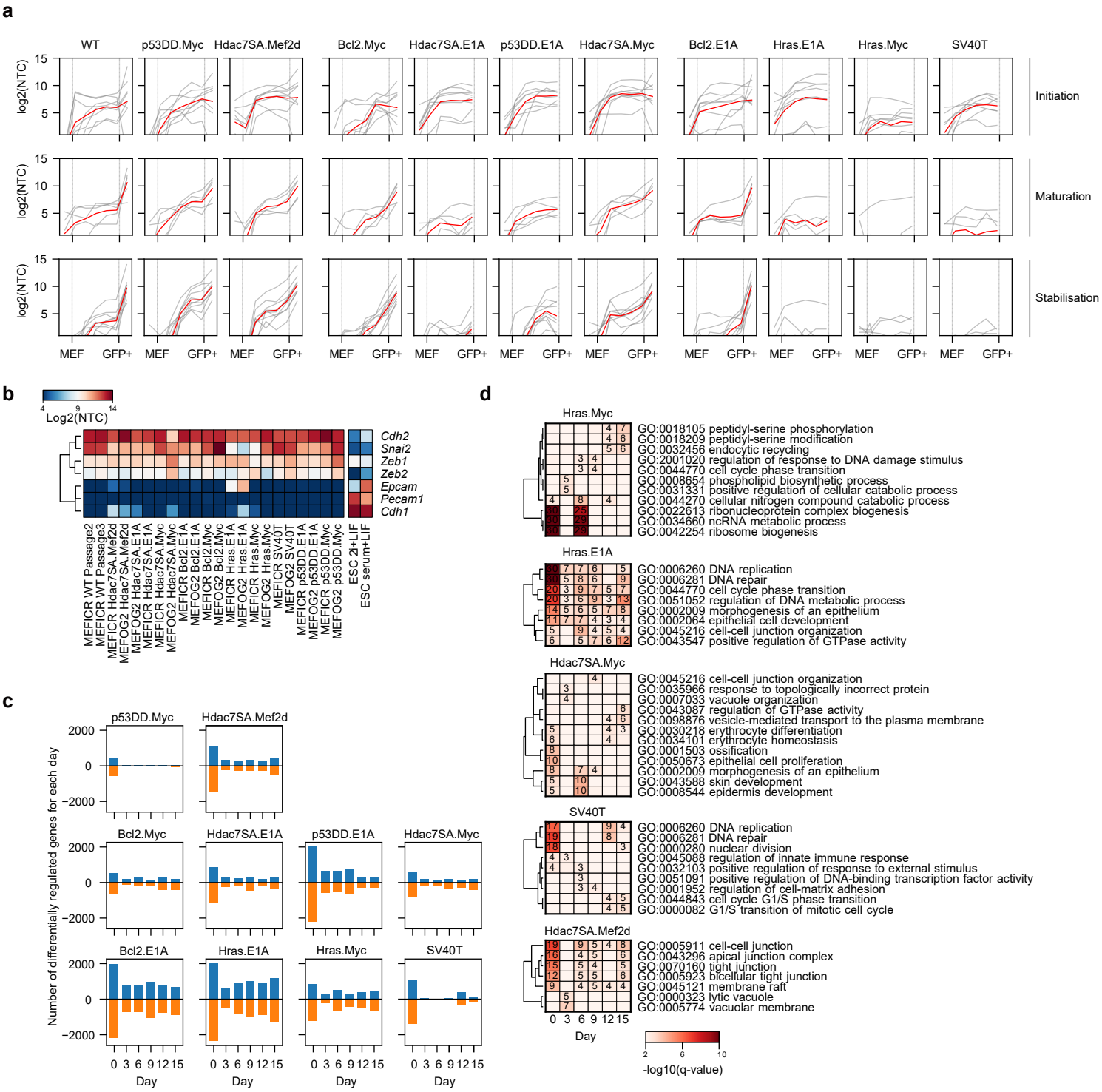
a
WT
p53DD.Myc
Hdac7SA.Mef2d
Bcl2.Myc
Hdac7SA.E1A
p53DD.E1A
Hdac7SA.Myc
Bcl2.E1A
Hras.E1A
Hras.Myc
SV40T
log2(NTC)
Initiation
Maturation
Stabilisation
MEF
GFP+
b
Log2(NTC)
Cdh2
Snai2
Zeb1
Zeb2
Epcam
Pecam1
Cdh1
MEFICR WT Passage2
MEFICR WT Passage3
MEFICR Hdac7SA.Mef2d
MEFOG2 Hdac7SA.Mef2d
MEFOG2 Hdac7SA.E1A
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

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

980 **a** Line charts showing the expression of initiation, maturation, and stabilization genes
981 (Samavarchi-Tehrani *et al.*, 2010), at the indicated time points or in MEFs and
982 GFP+ cells. See **Figure S6** for the expression of the individual genes in each class.
983 The red line is the mean of all genes, the grey lines are the individual genes.

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

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

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

994

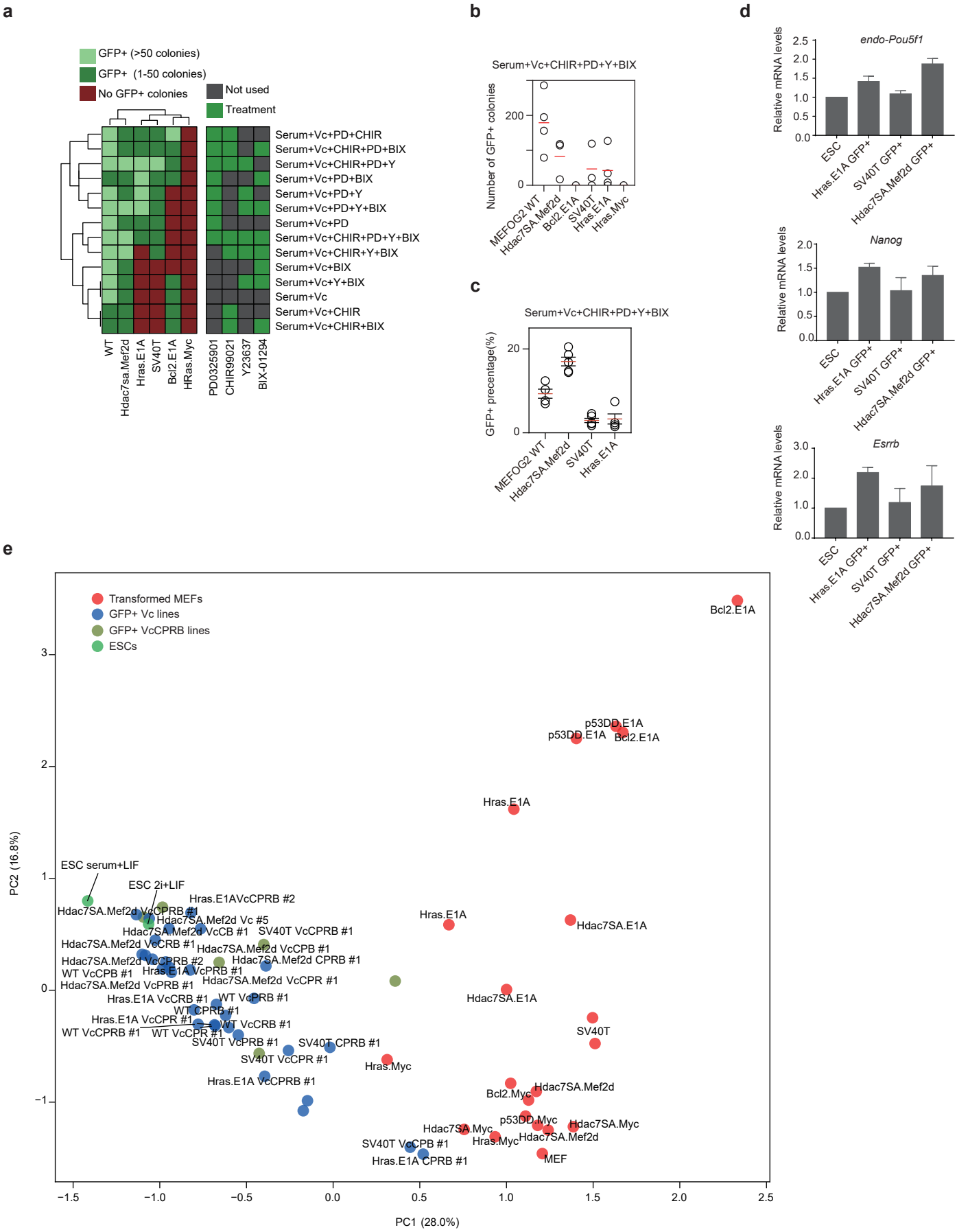
a
GFP+ (>50 colonies)
GFP+ (1-50 colonies)
No GFP+ colonies
Not used
Treatment
Serum+Vc+PD+CHIR
Serum+Vc+CHIR+PD+BIX
Serum+Vc+CHIR+PD+Y
Serum+Vc+PD+BIX
Serum+Vc+PD+Y
Serum+Vc+PD+Y+BIX
Serum+Vc+PD
Serum+Vc+CHIR+PD+Y+BIX
Serum+Vc+CHIR+Y+BIX
Serum+Vc+BIX
Serum+Vc+Y+BIX
Serum+Vc
Serum+Vc+CHIR
Serum+Vc+CHIR+BIX
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

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

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

999 **b** Chart showing the number of GFP+ colonies in the indicated immortal cell lines
1000 with the inhibitor cocktail. Data is from four biological replicates (MEFOGF WT) or
1001 from three biological replicates (all other transformed cell lines).

1002 **c** Percentage of cells sorted by FACS at day 12 of the reprogramming assay in the
1003 indicated transformed cell lines or wildtype MEFs. Data is from four biological
1004 replicates.

1005 **d** RT-qPCR at day 15 of a reprogramming experiment using primers targeting
1006 endogenous Oct4 (endo-*Pou5f1*), *Nanog, and Esrrb*. The mean of three biological
1007 replicates ± s.d. is shown.

1008 **e** PCA of the RNA-seq from the immortalized MEFs, and the reprogrammed iPSCs
1009 and ESCs. The various chemical cocktails used during the reprogramming are
1010 abbreviated for clarity. Key: Vc=Vitamin C; P=PD0325901; C=CHIR99021;
1011 R=Y23637; B=BIX-01294.

1012

1013

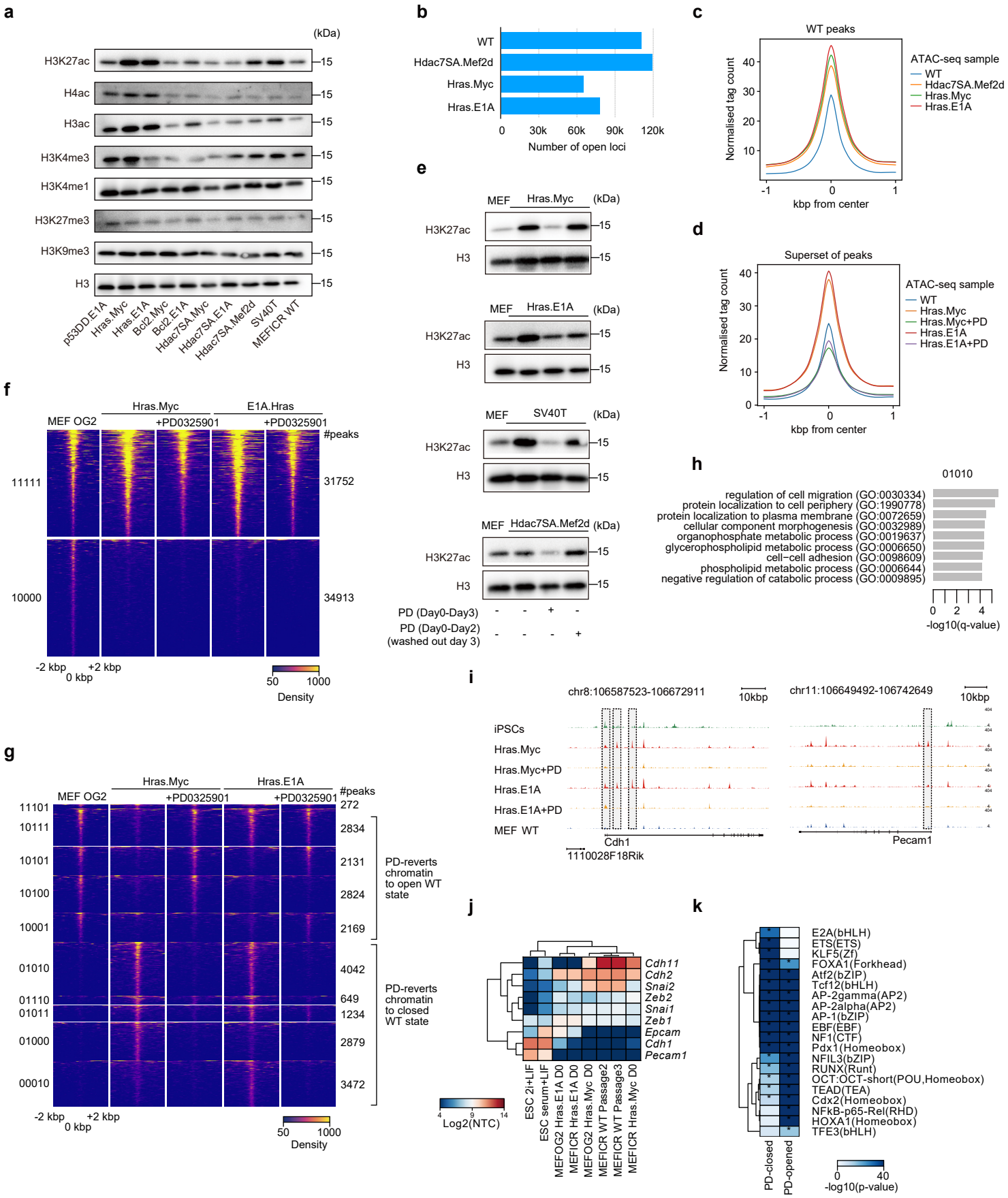
a
(kDa)
H3K27ac
H4ac
H3ac
H3K4me3
H3K4me1
H3K27me3
H3K9me3
H3
15
p53DD E1A
Hras.Myc
Hras.E1A
Bcl2.Myc
Bcl2.E1A
Hdac7SA.Myc
Hdac7SA.E1A
Hdac7SA.Mef2d
SV40T
MEFICR WT
b
WT
Hdac7SA.Mef2d
Hras.Myc
Hras.E1A
0
30k
60k
90k
120k
Number of open loci
c
WT peaks
Normalised tag count
kbp from center
ATAC-seq sample
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

1014 **Figure 6. Inhibition of MEK repairs some epigenetic defects in the transformed**
1015 **MEFs**

1016 **a** Western blot for a selection of epigenetic marks and total histone H3 as a loading
1017 control.

1018 **b** Bar chart showing the number of ATAC-seq loci identified as open/accessible in
1019 the indicated wildtype or transformed MEFs.

1020 **c** Pileups of the ATAC-seq data centered on all wildtype MEF peaks and showing
1021 the flanking 1 kbp (kilo base pairs). Normalized tag counts (reads per million per
1022 bin) are shown for the indicated samples.

1023 **d** Pileups of the ATAC-seq data centered on a non-redundant superset of all peaks
1024 from all of the ATAC-seq samples. The plot is centered on the peak summit and
1025 shows the flanking 1 kbp. Normalized tag counts are shown for the indicated
1026 samples.

1027 **e** Western blot of activatory histone marks upon addition of PD to the MEFs for 3
1028 days, or cells treated with PD for two days and then washed out on the third data
1029 (washed out on day 3).

1030 **f** Heatmap of the ATAC-seq data showing the permanently open (11111), and open
1031 only in the wildtype MEFs (10000). The number of peaks in each class is indicated
1032 on the right-hand side. Read density is shown centered on the peak summit and 2
1033 kbp on each side. The full heatmap containing all groups is in **Figure S9b**.

1034 **g** Heatmap of selected peak groups for the indicated ATAC-seq samples. The full
1035 heatmap containing all groups is in **Figure S9b**.

1036 **h** Gene ontology analysis for nearby genes (a TSS within 10 kbp of an open locus)
1037 from the 01010 group (closed in wildtype MEFs, PD causes these loci to close in
1038 the transformed MEFs) from **Figure 6g**.

1039 **i** Genome pileup view of the ATAC-seq data showing the *Cdh1* and *Epcam* loci in
1040 wildtype MEFs, Hras.E1A, and Hras.Myc MEFs with or without treatment with PD.
1041 Genome assembly is mm10. Regions of open chromatin that are modified in the
1042 PD-treated cells are indicated by vertical grey bars.

1043 **j** Expression heatmap of selected epithelial and mesenchymal genes in the
1044 indicated MEF lines and ESCs.

1045 **k** Motif analysis of selected PD-close and PD-open genes.

1046