Oncogenic mutations disrupt lineage differentiation via epigenetic perturbations

Lingjun Meng[1, 5*], Alice Yu[2], Huanwei Huang[3], Jodie Meng[2], Yongjin Yoo[1], Katie Schaukowitch[1], Justyna A. Janas[1], David A Knowles[4], Sylvia Katina Plevritis[2, 5*].

[1]Institute for Stem Cell Biology and Regenerative Medicine, Department of Pathology, Stanford University, CA 94305, USA

[2]Department of Biomedical Data Science, Radiology, Stanford University, USA

[3]National Institute of Biological Sciences, Beijing 102206, China

[4]New York Genome Center, New York, NY, Department of Computer Science, Columbia University, Department of Systems Biology, Columbia University, Data Science Institute, Columbia University, New York, NY

[5]Senior author

*Correspondence: menglingjundl@gmail.com (L.M.), plevriti@stanford.edu (S.K.P.)

Key word: Oncogenic mutation, DNA methylation, Lineage reprogramming/differentiation, Cancer.

Abstract

Understanding lineage differentiation is an ongoing challenge in the field of stem cell research. We acquired eight gain-of-function (GOF) oncogenic mutations associated with DNA methylation alteration from the TCGA datasets. After introducing each oncogene into wild-type stem cells separately, we found that the mutant stem cells are difficult to reprogram to iPSCs and terminally differentiate to induced neuronal cells, and retain proliferation capacity in the process of differentiation. This findings suggests that oncogenic mutations impair the lineage differentiation. After analyzing the resulting DNA methylation profiles, we identified many novel lineage differentiation genes in mutant cells showing DNA methylation alteration compared with wild-type cells, which is consistent with their methylation states in low-grade

gliomas (LGGs) compared with control brain. Collectively, our results indicate that oncogenic mutations in stem cells are associated with DNA methylation changes for genes related to multipotency maintenance, differentiation resistance, and tumorigenesis, therefore, displaying impairment in lineage commitment.

Introduction

Induced pluripotent stem cell (iPSC) reprogramming and cell differentiation hold great promise for regenerative medicine. However, their potential clinical application is hampered by the low efficiency of somatic cell reprogramming and differentiation. We previously found that cancer cells are extremely difficult to reprogram into iPSCs, except in acute myeloid leukemia (AML)[1]. Reprogramming resets epigenetic states in AML-iPSCs, but leukemic behavior and methylation are reacquired upon hematopoietic differentiation[1]. There appears to be a specific epigenetic profile related to cell behavior during lineage differentiation. A better understanding of oncogenic mutation associated epigenetic profiles holds great promise for enhancing reprogramming and differentiation efficiency. Here, we show that oncogenic mutations interrupt iPSC reprogramming and neural cell differentiation via change of epigenetic states, especially reveals that methylation changes appear on genes play a critical role on multipotency, differentiation, and tumorigenesis.

By analyzing DNA methylation profiles across cancers represented in TCGA (The Cancer Genome Atlas), we identified eight candidate oncogenic mutations that are most associated with genome-wide DNA methylation changes. To investigate the effect of oncogenic mutations in reprogramming and differentiation, we independently introduced each oncogene into wild-type (wt) neural progenitor cells (NPCs) derived from human embryonic stem cells (hESCs) and investigated the epigenetic alteration occurring after the induction of the oncogenes. We found that the induction of gain-of-function mutation of oncogenes effects the change of the epigenetic landscape in mutant NPCs compared to wt NPCs. To understand the role of oncogenes in lineage

54 commitment, we measured the differentiation efficiency of hESCs (or NPCs) to induced
55 neuronal cells (iN), and the reprogramming efficiency of mutant NPCs to human induced
56 pluripotent stem cells (hiPSCs). Through comparing the efficiency of mutant cells with wt cells
57 in the process of lineage reprogramming and differentiation, we found the mutant cells displayed
58 a preference for remaining as stem cells and resisting lineage change. Moreover, the mutant cells
59 retained their proliferation capacity in dysregulated lineage commitment and consequently
60 generated more cells in a differentiation resistance status due to inaccurate lineage commitment.
61 Through comparing the DNA methylation change between wt cells and mutant cells before and
62 after lineage differentiation, we found that many genes that play a role in multipotency
63 maintenance and neural identification had altered methylation, consistent with the same
64 methylation states in low-grade gliomas (LGGs). Our results indicate that the significant role of
65 oncogenes which is associated with the expression of genes that have a function on lineage
66 commitment through epigenetic modification.
67

68 **Results**

69 **Expression of oncogenic mutations blocks differentiation of ES cell into neurons**

70 It was previously shown that IDH1 (isocitrate dehydrogenase 1) mutations occur in the
71 majority of gliomas[2]. IDH1 mutant cells generate R(2)-2-hydroxyglutarate (2HG)[3], which causes
72 GASC1 (lysine-specific demethylase 4C, KDM4C) mediated hypermethylation[4,5]. Building upon
73 these previous studies[2,4], we chose the IDH1-mutation as our first candidate to study the
74 correlation between oncogene and epigenetic alteration. We investigatd the frequency of IDH1
75 mutation within each cancer type in TCGA. We confirmed that IDH1 mutation frequently occurs
76 in low-grade gliomas (LGGs) and glioblastoma (GBM) (Extended Data Fig. 1a). Thus, we
77 interrogated the effect of IDH1-R132* mutations (R132C, R132G, R132H, R132S) on the
78 epigenetic states in GBM and LGG, as measured by DNA methylation. IDH1-R132 mutations
79 are highly associated with DNA methylation changes (Extended Data Fig. 1b, c).

80 To understand whether oncogenes other than IDH1 show similar patterns, we performed a
81 pan-cancer analysis linking both gain-of-function (GoF) and loss-of-function (LoF) mutation
82 changes to epigenetic state, as represented by principal components of DNA methylation. Across
83 The Cancer Genome Atlas, we detected 22 and 17 candidate genes associations of DNA
84 methylation for GoF and LoF mutations respectively (1% false discovery rate) (Extended Data
85 Fig. 1d, e). We picked the top eight gain-of-function genes (IDH1-R132H, TP53-R273C,
86 BRAF-V600E, KRAS-G12C, NRAS-Q61R, HRAS-Q61R, SOX17-S403I, FGFR3-S294C) from
87 the candidate list based on p-value (adjusted pval < 0.05) for the following experimental
88 validation. It indicated that not only the IDH1 mutation reported before, but also a range of
89 oncogenic mutations are associated with epigenetic alteration in pan-cancer DNA methylation
90 analysis.

91 To investigate the function of the eight oncogenes indicated above in cell lineage
92 commitment, we assessed the differentiation efficiency from mutant human embryonic stem cells
93 (hESCs) to induced neuronal (iN) cells. We used a doxycycline (dox)-inducible lentiviral vector
94 system to transduce empty vector plasmid as a control and eight candidate oncogenes into hESC
95 H1 cells, separately. As we showed previously, overexpression of neuronal gene neurogenin-2
96 (Ngn2) efficiently converted hESCs into iN cells with nearly 100% yield and purity in less than
97 two weeks [6] (Fig. 1a). We compared the Ngn2 induced differentiation efficiency between wt and
98 mutant iN cells by staining with neuronal marker Tuj1 (tubulin beta-III). The differentiation
99 efficiency of wt hESCs is in a significant high percentage as nearly 100%. The wt hESC-derived
100 iN cells showed typical neuronal morphology with extension of axons and smaller cell nuclei.
101 The differentiation of iN cells from hESCs forces the cells to completely lose their stem cell
102 identification and proliferation capacity. Remarkably, for all eight tested candidate oncogenes, a
103 high percentage of mutant hESC cells failed to differentiate to iN cells (Fig. 1b). We
104 demonstrated that mutant hESC-derived iN cells have immature iN cell body morphology and
105 short axons (Fig. 1b). Their nuclear size is similar to hESCs, and not as small as wt

106 hESC-derived iN cells (Fig. 1b). We measured neurite length of cells at 12 days of iN
107 differentiation and found markedly shorter neurites in mutant hESC-derived iN cells than in wt
108 hESC-derived iN cells (Fig. 1c). In addition, we see a reduction in mRNA expression level of the
109 neuronal marker Map2 in mutant hESC-derived iN cells compared to wt hESC-derived iN cells
110 by qRT-PCR analysis at day 7 which is the early stage of differentiation (Fig. 1d).

111 Consistent with IDH1-R132H mutant hESCs, the IDH1-R132H mutant NPCs are also hard
112 to differentiate to iN cells compared to IDH1-wt NPCs (overexpress IDH1 without any
113 mutations) (Fig. 1e). From qRT-PCR analysis of neural marker expression at day 7 of
114 differentiation from NPCs, we found a significant decrease in the expression level of Map2,
115 Nestin, Pax6, and NeuroD1 in IDH1-R132H mutation NPC-derived iN cells compared to
116 IDH1-wt NPC-derived iN cells (Fig. 1f, Extended Data Fig. 2a, b, c). Collectively, these results
117 suggest that oncogenic mutations cause difficulty in inducing terminal differentiation from
118 hESCs and NPCs to neuronal cells, and perturb cell fate determination.

119 As known, the terminally differentiated cells, such as the neuronal cells discussed here,
120 cannot proliferate. Interestingly, we noticed that the mutant stem cells not only retain their stem
121 cell state via maintaining the expression of Sox2 and resist neural differentiation, but they also
122 continue to proliferate throughout the differentiation process. To investigate further, we counted
123 the cell number of wt and mutant hESCs with dox induction for 3 days, but did not find an
124 increase in mutant hESCs compared to wt hESCs (Extended Data Fig. 2d). This suggested that
125 there is no proliferation advantage of mutant hESCs. However, we found that the cell number of
126 mutant hESC-derived iN cells was increased after dox for 7 days compared to wt hESC-derived
127 iN cells (Extended Data Fig. 2e). To validate the possibility that oncogenes may enhance cell
128 proliferation, we stained the wt and mutant hESC-derived iN cells with the cell proliferation
129 marker Ki67 (Fig. 1g). For all eight oncogenes, we found that a high percentage of mutant
130 hESC-derived iN cells express Ki67 during the differentiation process, but rarely happen in wt
131 hESC-derived iN cells. To identify the cell type of Ki67+ cells, we co-stained for the neuronal

132 marker Tuj1 and pluripotency marker Sox2 with Ki67. There were three types of cells:
133 Ki67+Sox2+, Ki67+Sox2– and Ki67+Tuj1–. Ki67+Tuj1+ cells were never observed. We
134 counted the Ki67+ cells at 10 days of differentiation. Very few wt hESC-derived iN cells were
135 Ki67+, but more than hundreds times mutant hESC-derived iN cells retained their proliferation
136 capacity and resisted being fully differentiated to neural cells (Fig. 1g). There was a significant
137 reduction in neuronal cell number of mutant hESC-derived iN cells at day 10 (Fig. 1h), even
138 when the mutant hESC-derived iN cells expressed Tuj1, most of them did not grow long axons.
139 There were hundreds times more Sox2+ cells in the eight mutant hESC-derived iN cells than wt
140 hESC-derived iN cells at 10 days of differentiation (Fig. 1i, Extended Data Fig. 2f). In summary,
141 we found that stem cells expressing oncogenes enable themselves to resist differentiation into
142 neural cells. This suggests that many oncogenic mutations may be associated with inhibition of
143 differentiation potential due to epigenetic rewiring, because it has only happened during lineage
144 differentiation with more permissive epigenetic states.
145
146 **Stem cells with oncogenic mutations are difficult to reprogram into iPS cells**
147 As it was previously found that cancer cells are extremely difficult to reprogram into iPSCs[1],
148 we employed human neural progenitor cells (NPCs) as above used, and planned to assess the
149 human iPSC (hiPSC) reprogramming efficiency from NPCs with these eight oncogenes shown in
150 Fig 1, in order to comprehensively understand the lineage commitment blocking. We used NPCs
151 derived from hESCs following the dual smad inhibition protocol [7] (Extended Data Fig. 3). To
152 investigate the effect of each oncogene, we used the dox inducible system, same as discussed
153 above. The NPCs were used to generate hiPSCs by overexpressing the four transcription factors:
154 Oct4, Sox2, c-Myc, and Klf4 (OSKM)[8] (Fig. 2a). Afterward, the reprogramming efficiency of wt
155 and mutant NPC-derived iPSCs were compared by counting the number of iPSC colonies via
156 immunofluorescence staining of an iPSC marker, TRA-1-60 (Fig. 2b, c). Wt and IDH1-wt
157 NPC-derived iPSCs generated many colonies. In contrast, all eight mutant NPCs were difficult to

158 reprogram into iPSCs and generated no iPSC colonies or a few very small colonies (Fig. 2b, c).
159 These results were consistent with our previous study showing that cancer cells were difficult to
160 reprogram to iPSCs (Extended Data Fig. 4), even if there was only one oncogene induced in stem
161 cells with a short-term inducible expression. In summary, stem cells with oncogenes are difficult
162 to reprogram into iPS cells.
163

164 **Oncogenic mutation is associated with epigenetic change in hESCs**

165 To investigate the effect of oncogenic mutations on genome-wide DNA methylation in stem
166 cells, we performed an Infinium MethylationEPIC BeadChip assays. To clarify that the
167 epigenetic change is caused only by one oncogenic mutation, we introduced each oncogene into
168 individual wt hESCs, the same as we previously did, using dox to overexpress the GoF mutation
169 gene for nine days. The principal component analysis (PCA) of DNA methylation profiles
170 demonstrated that the seven mutant hESCs and wt hESCs as control (empty vector) were
171 distinguished by DNA methylated gene loci (Extended Data Fig. 5a, b). 282 CpG sites are
172 differentially methylated in the seven mutants compared with control, 246 out of 286 are
173 associated with specific genes (Extended Data Fig. 5c). TP53I3 and PTPRT were the top two
174 genes differentially methylated in differentially methylated positions (DMPs) distribution.
175 Furthermore, 321 genomic regions exhibited significant differences in differentially methylated
176 regions (DMRs) distribution (p-val < 0.05) (Extended Data Fig. 5d). Based on the distance from
177 CpG islands, there are 401 genomic regions that are differentially methylated in total 4,905
178 regions (p-val < 0.05) (Extended Data Fig. 5e). Geneset enrichment analysis showed that many
179 tumor suppressor genes are methylated within DMPs and DMRs (Extended Data Fig. 5f, g). The
180 genes are enriched for stem-like genesets, such as Wnt signaling pathway (WNT5A, UBC,
181 WWTR1, Sox9), embryonic skeletal system morphogenesis (HOXB1, HOXA7),
182 beta-catenin-TCF complex assembly (HIST1H4B, HIST1H4C), and positive regulation of

183 transcription from RNA polymerase II promoter (ZBTB38). These findings demonstrated that
184 oncogenic mutations are associated with epigenetic alteration in hESCs.
185
186 **Oncogenic mutations block differentiation via epigenetic alteration of stem cell fate genes**
187 As shown above, stem cells with oncogenes were difficult to induce into terminal
188 differentiation from NPCs to iN cells, which may be associated with epigenetic changes (Fig.1).
189 To investigate temporal DNA methylation changes occurring during NPC maintenance in early
190 and later times via genomic DNA methylation measurement by Infinium MethylationEPIC
191 BeadChip assay, we introduced IDH1-R132H, KRAS-G12C, TP53-R273C, BRAF-V600E,
192 NRAS-Q61R, HRAS-Q61R, and SOX17-S403I mutations separately into NPCs at day 0, with
193 dox induction for 2 or 10 days to overexpress the oncogenes. After 10 days of induction, we
194 differentiated the NPCs into iN cells for two more days using the same procedure described in
195 Fig 1 (Fig. 3a), in order to investigate the epigenetic change during differentiation at increase
196 temporal resolution. As a control, we conducted the same analysis on wt-2 days, wt-10 days, and
197 wt-10 days + 2 iN days (Fig. 3a).
198 At later time points we observed an increase in epigenetic variability within the mutant cell
199 lines, notably for IDH1, KRAS, and Sox17 mutations (Fig. 3b). To measure oncogene-specific
200 methylation changes, we identified differentially hyper- (beta-value > 0.8 and logFC > 0.3) and
201 hypo- (beta-value < 0.2, logFC < -0.3) methylated CpG sites. We found that oncogenic mutation
202 is associated with DNA methylation alteration in NPCs and NPC-derived iN cells generally
203 (Extended Data Fig. 6a, b, c). Notably, the DNA methylation change in IDH1, TP53, and KRAS
204 mutation NPCs are more significant, and these mutations are very common in brain tumors and
205 pan cancer. Therefore, to further understand the properties of IDH1, TP53, and KRAS mutation,
206 we identified DMPs at 2 days (Extended Data Fig. 6a, d), 10 days (Extended Data Fig. 6b, e) and
207 10 + 2 iN days (Extended Data Fig. 6c, f). The heatmap of significant DMPs demonstrated that
208 the epigenetic perturbations were enhanced as induction time increased and differentiation

occurred (p-value < 0.01) (Fig. 3c). To identify functional categories of differentially hypermethylated genes, we performed GO enrichment analysis using clusterProfiler, using a cut-off of 1.5-fold 5mC level change to define hypermethylated genes (meandiff > 0.05, padj < 0.01) (Fig. 3d). Genes are hypermethylated in wt NPCs at 2 to 10 days comparisons that are enriched among neural differentiation, regulation of neural precursor cell proliferation, endoderm development, stem cell differentiation, and cell fate commitment. On the contrary, genes are hypermethylated in IDH1, KRAS, and TP53 mutation NPCs at 2 to 10 days comparisons that are involved in the regulation of neuron death, neurotransmitter transport, cell proliferation, regulation of secretory pathway, endocrine system development, regulation of cation transmembrane transport, regulation of cell shape and growth. It was shown that there is a significant difference in gene methylation states between wt and mutants from 2 to 10 days. In addition, the three mutants share a number of hypermethylated genes.

To eliminate the effect that is not related with oncogenes, we did comparative analysis between wt-2days and wt-10days (wt-2/10days), between mutation-2days and mutation-10days (mutation-2/10days), then compared the wt-2/10days and mutation-2/10days. Similarly, we performed the comparison between wt-10days and wt-10+2iNdays (wt-10/10+2iNdays), between mutation-10days and mutation-10+2iNdays (mutation-10/10+2iNdays), then compared the wt-10/10+2iNdays and mutation-10/10+2iNdays (Extended Data Fig. 6g, h, Fig. 3e, f, g, h). Genes that were found to be differentially methylated in wt cells at the relevant comparison periods were removed in the following analysis. Notably, there are a large number of hypomethylated regions in mutant NPCs compared to wt NPCs from 2 to 10 days (Extended Data Fig. 6g, h). However, from 10 to 10 + 2 iN differentiation days, there are a large number of hypermethylated regions in mutant NPC-derived iNs compared to wt NPC-derived iNs (p-val < 0.05, max difference of CpG > 0.2) (Extended Data Fig. 6g, h). The distribution of DNA methylation change showed that most occurred at transcription start site [9] (Extended Data Fig. 6h).

235 To reveal the most distinct methylation changed genes, we delineated overlapping
236 methylated genes between the three mutants at 2 to 10 days and 10 to 10 + 2 iN days
237 comparisons (Fig. 3e). Interestingly, cells have IDH1, KRAS, and TP53 mutation share 25 genes
238 that were hypomethylated at 2 to 10 days comparisons, and share 32 genes were
239 hypermethylated at 10 to 10 + 2 iN days comparisons (Fig. 3e). Simultaneously, each mutant
240 displays different methylation changed genes in a specific manner (Fig. 3e). We next categorized
241 the methylation changed genes based on functionality (Fig. 3f). At 2 to 10 days comparisons, we
242 found that NPCs with IDH1, KRAS, or TP53 mutation shared several of the same
243 hypomethylated genes in the categories of pluripotency, multipotency, differentiation (BAI1,
244 MTL5); neural development (CNPY1, GALR1, GBX2, LHX5); cancer (CACNA2D3, DMTN,
245 FUCA2, IL12RB1, ITGAM, MYOZ3, NR2F1, NR2F1-AS1, UBR3, HIST1H1B) (Fig. 3f). At
246 10 to 10 + 2 iN days comparisons, we found that NPC-derived iN cells have IDH1, KRAS, and
247 TP53 mutation shared several of the same hypermethylated genes in pluripotency, multipotency,
248 differentiation (FOXD3, TAF10); neural development (GABRA1, GPR176, SALL1); cancer
249 (FH, BAZ1A, DCLRE1C, FARSB, HOXC12, ILK, LMBR1, TPP1, UBE2T, ZNF540, ZNF571)
250 (Fig. 3f).

251 In addition to the common methylation change in the three mutants, they also have distinct
252 patterns. Hoxb1 is hypomethylated in IDH1 mutation NPCs, and KLF4 is hypomethylated in
253 KRAS and TP53 mutation NPCs at 2 to 10 days comparisons (Fig. 3f). As reported, when Hoxb1
254 (Homeobox B1) expression is induced in ESC-derived NSCs (neural stem cells), few cells
255 differentiate, but they keep proliferating [10]. KLF4 (Kruppel Like Factor 4) -overexpressing NSCs
256 resisted neuronal differentiation, showing rounded morphology with very short processes[11],
257 which is consistent with our observation. Additionally, at 10 to 10 + 2 iN days comparisons,
258 Foxp1 is hypermethylated in IDH1 and KRAS mutation NPC-derived iN cells (Fig. 3f). Gain or
259 loss of Foxp1 (Forkhead Box P1) function alters the level of progenitor cell divisions and the
260 balance between self-renewal and differentiation[12]. The DNA methylation alterations happened

261 mostly in the functional category related to multipotency and neural identification as well as
262 tumorigenesis. Correspondingly, as previously mentioned, mutant stem cells resist neural
263 differentiation and reprogramming as a consequence. DNA methylation changes in normal NPCs
264 dynamically regulate cell fate and identity. However, aberrant DNA methylation changes in
265 mutant NPCs dysregulate cell fate and identity.

266 To explore the critical effect of differentiation that occurs in mutant stem cells, we plotted
267 the methylation of each DMP for a given region before and after differentiation (Fig. 3g).
268 Interestingly, we found some genes were hypermethylated from 2 to 10 days, but convert to
269 hypomethylated from 10 to 10 + 2 iN days, such as Lin28A in TP53 mutation cells (Fig. 3g). On
270 the contrary, some genes were hypomethylated from 2 to 10 days, but convert to
271 hypermethylated from 10 to 10 + 2 iN days, such as FOXD3 in IDH1 and KRAS mutation cells
272 (Fig. 3g). Lin28A (Lin-28 Homolog A) controls NPC proliferation and neurogenesis in a
273 SOX2-dependent manner[13]. Foxd3 (forkhead box D3) is associated with inactivation of
274 important naive pluripotency genes by its modification of chromatin structures via recruiting
275 histone demethylases and decreasing numerous activating factors[14]. These genes effect stem cell
276 identification and differentiation regulation that were changed on their methylation modification
277 in an aberrant manner, which severely effects the accuracy of lineage commitment.

278 Meanwhile, we found many overlapping gene sites were hypermethylated or
279 hypomethylated in three mutants (Fig. 3f). In particular, BAI1 was hypomethylated from the 2 to
280 10 days, and KMO was hypermethylated from the 10 to 10 + 2 iN days (Fig. 3h). BAI1
281 (brain-specific angiogenesis inhibitor 1) is expressed in normal human brain, mediates p53
282 polyubiquitination[15,16]. KMO (Kynurenine 3-monooxygenase) influences stem cell biology and
283 modulates neurodegeneration[17]. The three mutants share numerous genes with the same
284 methylation site modification, which suggested that there is a specific methylation pattern
285 generally related to oncogenic mutation associated with lineage differentiation disorder.

286

287 **Stem cells with oncogenic mutations share similar epigenetic patterns with brain tumors.**

288 As previously discussed, these oncogenes came from tumor cells via TCGA data analysis
289 (Extended Data Fig. 1d), that are associated with DNA methylation alteration. It is important to
290 evaluate the similarity of epigenetic states between mutant cells and cancer, such as low-grade
291 gliomas (LGGs). To identify the methylation changed genes associated with LGGs, we analyzed
292 the epigenetic profile of LGGs with IDH1-R132H, TP53-R273C, BRAF-V600E, or other
293 mutations separately, from TCGA datasets as measured by DNA methylation, and compared to
294 the non-tumor human whole brain cells as control[18] (see Methods). To evaluate the similarity
295 between mutant cells and LGG cells in DNA methylation patterns broadly, we compared the
296 genomic regions that were differentially methylated between the control brain and LGGs with
297 oncogenes, as well as between wt and mutant cells at 2, 10, and 10 + 2 iN days (Fig. 4a, b,
298 Extended Data Fig. 7a, b). It was shown that many overlapping genes were hypermethylated in
299 mutant cells and LGGs, and gene numbers increased as the increased days of oncogene induction
300 and differentiation initiation (Fig. 4a, b), also many overlapping genes were hypomethylated
301 (Extended Data Fig. 7a, b). The heatmap displays the most significantly differentially
302 hypermethylated genes overlapped in mutant cells and cancer cells (Fig. 4b). Many genes retain
303 the same methylation states from 10 to 10 + 2 iN days (Fig. 4b).

304 We found that many genes were hypermethylated in NPCs and NPC-derived iN cells with
305 oncogenes, that were also hypermethylated in LGGs with the same oncogene (Fig. 4c), such as
306 CASZ1, DACT2, SFRP2, MET, PIWIL1, VGLL4, LATS2, SIX3, UHRF1. As reported, SFRP2
307 is downregulated in glioma patients by promoter hypermethylation, and regulates Wnt/β-catenin
308 activation in glioma cells[19,20]. Inactivation of LATS1/2 kinases causes YAP/TAZ-driven global
309 hypertranscription in neural progenitors that inhibits differentiation[21]. LATS1/2 deficiency
310 reprograms schwann cells (SCs) to a cancerous, progenitor-like phenotype and promotes
311 hyperproliferation[22]. Notably, to fully understand the relation of these hypermethylated genes,

312 we analyzed the biological connection of them. We found that the hypermethylated genes in both
313 mutant cells and tumor cells have a close relation to chromatin-modifying enzymes (Fig. 4d).
314 Moreover, we compared the genomic regions that were differentially methylated between
315 the control brain and LGGs with oncogenes, as well as between wt and mutant NPCs or
316 NPC-derived iN cells at 2 to 10 days or 10 to 10 + 2 iN days comparisons (Extended Data Fig.
317 7c, d). Notably, we found that the genes were hypermethylated in mutant cells from Fig. 3f and
318 were also hypermethylated in LGGs (Extended Data Fig. 7c, d), such as TMEM106A and
319 RARRES1 at 2 to 10 days comparisons (Extended Data Fig. 7e), Foxd3 and HOXC12 at 10 to 10
320 + 2 iN days comparisons (Extended Data Fig. 7f). These genes were hypermethylated in mutant
321 cells and LGGs that have a critical function in cell fate determination via epigenetic modification.
322 Collectively, tumors with various accumulated oncogenes share similar DNA methylation
323 patterns with stem cells have oncogenes after a short period of induced expression.
324 To specifically determine DNA hypermethylation sites consequent to oncogenic mutation
325 associated differentiation impairment, we investigated each methylation locus of
326 hypermethylated genes from Fig. 3g, h, 4c in mutant cells and brains. Elevated methylation
327 occurred at the same loci of genes in mutant cells and LGGs, such as CASZ1 and VGLL4
328 (Extended Data Fig. 7g).
329 Furthermore, we analyzed the biological connection of key genes from Fig. 3f, in order to
330 show the whole picture of the network. Biological process categories were visualized as nodes
331 clustered into functional modules and interconnected based on the indicated interaction
332 confirmed in literature. The red circle labeled genes had methylation alteration as shown in Fig.
333 3f. We categorized the genes by five different colors: the red node indicates genes in
334 pluripotency, multipotency, and differentiation (LIN28A, FOXD3, KLF4, HOXB1, MEIS,
335 NR2F1, BAI1, HOXA5, HOXA9, PHB, FOXP1, OTX2), the green node indicates genes in
336 neural development (GBX2, LHX5, SALL1, PAX2), the blue node indicates genes in cancer
337 (NCOR, NGFR, DDB1, FH, ILK, TPP1, ZNF540, LMO3, KLF7, NFIB), the light green node

338 indicates genes in DNA, RNA binding and translation (RPA1), the purple node indicates genes
339 which function as histone and chromatin enzymes (KDM2A, PRC1/2) (Fig. 4e). The
340 chromatin-modifying enzymes are labeled in yellow circles and shown on top of the
341 visualization. We labeled genes connected with at least three other genes with green circles. The
342 interaction of genes was connected with black lines based on published literature. We applied
343 blue lines to indicate genes connected with chromatin-modifying regulators. The arrow indicates
344 the effect of one gene on another gene. The gene regulatory network demonstrated that the genes
345 play a critical role in multipotency maintenance, lineage differentiation, and cancer formation
346 connect with each other via epigenetic regulation of chromatin-modifying enzymes.

347 To summarize the findings from this comprehensive data analysis, it is not only IDH1
348 mutation that causes methylation change [3,4], but other mutations, such as KRAS, TP53, and
349 SOX17 mutation. The DNA methylation states are important to the maintenance of specific cell
350 types, and also to facilitate differentiation from one cell type to another. Oncogenic mutations
351 occurred in stem cells and progenitor cells which perturb DNA methylation states on lineage
352 commitment genes and their functional performance in the process of differentiation,
353 consequently, emerging a type of cell that maintain multipotency, resist differentiation, and
354 retain proliferation, that is a novel exploration of impairment of differentiation efficiency. We
355 are the first to identify oncogenic mutations that disrupt DNA methylation change on lineage
356 differentiation genes, therefore, mutant stem cells undergoing inaccurate lineage commitment.

357

358 **Discussion**

359 In 2017, Chao et al. generated iPSCs from AML patient samples and found that
360 AML-iPSCs lacked leukemic potential, but when differentiated into hematopoietic cells, they
361 reacquired the ability to give rise to leukemia *in vivo* and reestablished leukemic DNA
362 methylation/gene expression patterns[1]. To investigate the relationship of oncogenic mutation and
363 DNA methylation change during lineage differentiation, we constructed mutant NPCs (or hESCs)

364 for comparing the lineage differentiation efficiency with or without oncogenes. We found stem
365 cells with oncogenes were hard to reprogram and differentiate. During reprogramming and
366 differentiation, a global reset of the epigenome occurs[8]. Also, the epigenetic states are more
367 permissive for gene expression during differentiation. It implies the possibility that these stem
368 cells with oncogenes may initiate differentiation failures via taking advantage of epigenetic reset
369 during lineage differentiation.

370 To investigate the essential DNA methylation changes that occurred in mutant NPCs at the
371 stage of maintenance and differentiation, we performed methylation assay on cells undergoing
372 oncogene induction for 2 days or 10 days, and neural cell differentiation for 2 days. We found
373 that Lin28A was hypermethylated in TP53 mutant cells at 2 to 10 days, but converts to
374 hypomethylated at 10 to 10+2 iN days (Fig. 3g). In the opposite way, some genes were
375 hypomethylated at 2 to 10 days, but convert to hypermethylated at 10 to 10+2 iN days, such as
376 FOXD3 in IDH1 and KRAS mutant cells (Fig. 3g). It had been reported that Lin28 regulates the
377 self-renewal of stem cells through direct interaction with target mRNAs and by disrupting the
378 maturation of certain miRNAs involved in embryonic development[13]. Lin28a/b double knockout
379 mice display neural tube defects and reduced proliferation and precocious differentiation of
380 NPCs[23]. FOXD3 was demonstrated to be required in maintaining pluripotency in mouse
381 embryonic stem cells[14]. Importantly, the DNA methylation states of these genes are converted in
382 mutant NPCs compared to wt NPCs during lineage differentiation. The mechanism of converted
383 methylation states corresponding with differentiation is interesting to explore further.

384 As found previously, these oncogenes came from tumor cells as shown in Extended Data
385 Fig. 1d. Therefore, we evaluated the similarity of epigenetic states between mutant cells and
386 LGGs. Importantly, it was shown that many genes that were hypermethylated in mutant cells
387 also were hypermethylated in LGGs. CASZ1 functions as a neural fate determination gene and
388 tumor suppressor. Piwil1 (piwi like RNA-mediated gene silencing 1) plays an important role in
389 RNA silencing and stem cell self-renewal[24]. Six3 (SIX homeobox 3) is required for development

390 of the rostral forebrain through binding to the Wnt1 promoter region[25]. Uhrf1 (ubiquitin like with
391 PHD and ring finger domains 1) is a widely known hemi-methylated DNA-binding protein,
392 playing a role in DNA methylation through the recruitment of Dnmt1[26]. Uhrf1 regulates cell
393 specification, particularly those associated with neuroectoderm and mesoderm specification[27,28].
394 SFRP2 (secreted frizzled related protein 2) is a Wnt antagonist and inhibits β-catenin[29] and bone
395 morphogenic protein (BMP) signaling pathways, and SOX2 directly binds the SFRP2
396 promoters[30]. SFRP2 improves mesenchymal stem cells engraftment by modulating self-renewal
397 through increasing stem cell survival and by inhibiting differentiation[19]. The inactivation of
398 Hippo pathway kinases LATS1/2 (large tumor suppressor kinase 1/2) causes a YAP/TAZ-driven
399 global increase in transcription activity during brain development, and upregulates many genes
400 involved in cell growth and proliferation, resulting in inhibition of differentiation of neural
401 progenitor cells [21].

402 In summary, our findings reveal an essential function of these lineage differentiation genes
403 as a stabilizer of the epigenome by promoting epigenetic modifications which are necessary for
404 the faithful differentiation and the maintenance of pluripotency and multipotency (Fig. 4e). There
405 is a lineage-specific epigenetic pattern. When stem cells have oncogenes, oncogenic mutation
406 associated epigenetic change effects the expression of pluripotent genes, and impair the accuracy
407 of lineage commitment.

408

409 **Methods**

410 **Key resources table**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-Ki-67	Abcam	Cat# Ab15580; RRID: AB_443209
Mouse anti-Nestin	R&D Systems	Cat# MAB1259; RRID: AB_2251304
Mouse anti-Oct3/4	Santa Cruz Biotechnology	Cat# sc-5279; RRID: AB_628051
Goat anti-Sox1	R&D Systems	Cat# AF3369; RRID: AB_2239879
Goat anti-Sox2	R&D Systems	Cat# AF2018; RRID: AB_355110
Mouse anti-Tuj1	BioLegend	Cat# 801202; RRID: AB_10063408
Rabbit anti-Tuj1	BioLegend	Cat# 802001; RRID: AB_2564645
Mouse anti-Tra-1-60	Millipore	Cat# MAB4360; RRID: AB_2119183
DAPI	Thermo Fisher Scientific	Cat# D1306; RRID: AB_2629482
Anti-Mouse IgG (Alexa Fluor 488-coupled)	ThermoFisher Scientific	Cat# A10037; RRID: AB_2534013
Anti-Rabbit IgG (Alexa Fluor 488-coupled)	ThermoFisher Scientific	Cat# A10042; RRID: AB_2534017
Alexa Fluor 488 donkey anti-goat IgG	Invitrogen	Cat# A-11055; RRID: AB_2534102
Alexa Fluor 488 donkey anti-mouse IgG	Invitrogen	Cat# A-21202; RRID: AB_141607
Alexa Fluor 488 donkey anti-rabbit IgG	Invitrogen	Cat# A-21206; RRID: AB_2535792
Alexa Fluor 555 Donkey Anti-Mouse IgG	Invitrogen	Cat# A-31570; RRID: AB_2536180
Alexa Fluor 555 Donkey Anti-Rabbit IgG	Invitrogen	Cat# A-31572; RRID: AB_162543
Chemicals, Peptides, and Recombinant Proteins		
DMEM media	ThermoFisher Scientific	Cat# 11965092
DMEM/F-12 media	Thermo Fisher	Cat# 1132033
Neurobasal-A Medium	ThermoFisher Scientific	Cat# 10888022
mTeSR1 media	Stemcell Technologies	Cat# 85850
HyClone cosmic calf serum (CCS)	GE Healthcare	Cat# SH30087.04
HyClone fetal bovine serum (FBS)	GE Healthcare	Cat# SH30071.03
Knockout Serum Replacement	ThermoFisher Scientific	Cat# 10828028
Penicillin-Streptomycin	ThermoFisher Scientific	Cat# 15140163
L-Glutamine	ThermoFisher Scientific	Cat# 25030164
Sodium Pyruvate	ThermoFisher Scientific	Cat# 11360070
GlutaMAX Supplememnt	ThermoFisher Scientific	Cat# 35050061
MEM Non-Essential Amino Acids Solution (100X)	ThermoFisher Scientific	Cat# 11140050
2-Mercaptoethanol	Gibco	Cat# 21985023
N-2 supplement (100X)	ThermoFisher Scientific	Cat# 17502048
B-27 supplement (50X)	Thermo Fisher	Cat# 12587010
H-EGF	PeproTech	Cat# 100-47
H-FGF-basic-154	PeproTech	Cat# 10018B
Insulin	Sigma-Aldrich	Cat# I6634

Heparin	Sigma	Cat# H3393
Ascorbic Acid	Sigma	Cat# A4403
HEPES Buffer Solution	Gibco	Cat# 15630049
Matrigel	BD Biosciences	Cat# 356230
Polybrene	Sigma-Aldrich	Cat# H9268
Laminin in Tris-buffered saline	Sigma-Aldrich	Cat# L2020
Mitomycin C	Stress Marq Biosciences	Cat# SIH-246-10MG
Bambanker freezing medium	NIPPON Genetics	Cat# BB03
Trypsin 10x (2.5%)	Sigma	Cat# 59427C-100ML
Accutase	Innovative Cell	Cat# AT104
Doxycycline	Sigma-Aldrich	Cat# D9891
Blasticidin	Sigma-Aldrich	Cat# 203350
Puromycin	Sigma-Aldrich	Cat# P8833
Hygromycin B	Sigma-Aldrich	Cat# H3274
Y-27632	Stemcell Technologies	Cat# 72302
SB431542	Selleckchem	Cat# S1067
LDN193189	Selleckchem	Cat# S2618
CHIR99021	Tocris	Cat# 4423
Recombinant DNA		
pMD2.G	Addgene	Cat# 12259; RRID: Addgene_12259
pRSV-Rev	Addgene	Cat# 12253; RRID: Addgene_12253
pMDLg/pRRE	Addgene	Cat# 12251; RRID: Addgene_12251
FUW-M2rtTA	Addgene	Cat# 20342; RRID: Addgene_20342
Ub-rtTA	Zhang et al., 2013	
TetO-EGFP	Zhang et al., 2013	Addgene #84041
TetO-Ngn2-T2A-PuromycinResistance	Zhang et al., 2013	Addgene #52047
TetO-hygromycin	Generated in this work	N/A
TetO-IDH1-R132H-hygromycin	Generated in this work	N/A
TetO-TP53-R273C-hygromycin	Generated in this work	N/A
TetO-BRAF-V600E-hygromycin	Generated in this work	N/A
TetO-KRAS-G12C-hygromycin	Generated in this work	N/A
TetO-NRAS-Q61R-hygromycin	Generated in this work	N/A
TetO-HRAS-Q61R-hygromycin	Generated in this work	N/A
TetO-SOX17-S403I-hygromycin	Generated in this work	N/A
TetO-FGFR3-S294C-hygromycin	Generated in this work	N/A
Experimental Models: Cell Lines		
Human H1-ESCs	WiCell Research Institute	Cat# WA01

HEK293T cells	ATCC	Cat# CRL-11268
Bacterial and virus strains		
JM109 E. coli	New England Biolabs	Cat# E4107
Critical commercial assays		
RNeasy Mini Kit	Qiagen	Cat# 74106
DNeasy Blood & Tissue Kit	Qiagen	Cat# 69506
Software and Algorithms		
ImageJ	NIH	RRID: SCR_003070
GraphPad Prism		https://www.graphpad.com/; RRID: SCR_002798
Deposited Data		
Infinium MethylationEPIC array	This study	GEO: GSE

Cell culture and media

H1 ES cells were obtained from WiCell Research Resources (WiCell, WI). ES cells were maintained as feeder-free cells in mTeSR™1 medium (Stem Cell Technologies). We used a doxycycline (dox)-inducible lentiviral vector system to transduce empty vector plasmid as a control and oncogenes into the H1 ES cells, separately. Selection was carried out for 7 days in mTeSR™1 medium with 50 μg/ml hygromycin.

Generation of iN cells from human ES cells

H1 human ESCs were dissociated using accutase and plated as single cell in mTeSR™1 medium supplemented with ROCK inhibitor Y-27632 (10 μM, StemCell Technologies, Inc.) on matrigel to a final density of 20,000-30,000 cells per cm^2. On the second day, attached cells were infected with lentivirus for Ngn2-puro and rtTA expression. After 24 hr (day 0), transgene expression was induced by replacing the medium with N2/B27 medium (DMEM/F12, 1% N2 supplement (Life Technologies), 1% B27 supplement (Life Technologies), 10 μg/ml insulin (Life Technologies)), and 2 μg/ml Doxycycline. Selection was carried out for 14 days (day 1 to day 14) in N2/B27 medium with 2 μg/ml puromycin. From day 2, induced neuronal (iN) cells were maintained in N2/B27 medium and changed medium every 2 days. Neuronal cells were defined as cells that stained positive for Tuj1 and had a process at least three times longer than the cell body.

hESCs differentiate to neural progenitor cells

432 hESCs were grown in mTeSR media on matrigel coated dishes until confluent, then
433 enzymatically passaged with accutase, resuspended in Neurobasal-A medium, DMEM/F12, 1%
434 N2, 1% B27 and insulin media, with Y-27632, a specific inhibitor of Rho kinase (ROCK)
435 activity, and cultured in the same size of the dish without matrigel coating. Overnight, colonies
436 will form floating spherical clusters termed "embryoid bodies" (EBs). When EB size grows to
437 200～400 μm, they were gently moved to matrigel coated dishes. The next day, after the EBs
438 attached to the bottom, the media was changed to N2/B27 media supplemented with dual SMAD
439 inhibitors (LDN193189 and SB431542). EBs were fed every second day with N2/B27 media
440 supplemented with 0.1 μM LDN193189 and 10 μM SB431542[7]. Cells are cultured for several
441 days, until neural rosettes begin to appear, characterized as round clusters of neuroepithelial cells
442 with apicobasal polarity. Staining of Sox1 is performed to identify the neural progenitor cells.

443 **Optimized protocol for NPCs differentiation from hESCs**

444 NPCs were differentiated from hESCs using dual SMAD inhibitors[7] (Extended Data Fig. 3).
445 High-density hESCs were identified on day 1, and passaged to the same size of dish to form
446 embryoid bodies in iN cell medium on day 2～3 until the EB size is large as 200～300 μm. EBs
447 were harvested and moved to matrigel coated dishes on day 4. Media was changed to NSC
448 medium with TGF-β and BMP inhibitor to differentiate NPCs from day 5 to day 21. NSC basal
449 medium includes DMEM/F12 and Neurobasal medium at 1:1, 1% Sodium pyruvate, 1% P/S, 1%
450 NEAA, 1 M HEPES, 1% Glutamax, 1% NaHCO3. When used for culturing cells add: 1% B27,
451 20 ng/ml bFGF, 20 ng/ml EGF, 1 μl/ml B27 supplement, 2 μg/ml Heparin, and 0.1% LIF. NPC
452 differentiation medium includes fetal NSC basal medium, 1% N2, 2% B27, 0.1 mM Ascorbic
453 acid, 10 μM SB431542 (TGF-β inhibitor), and 0.1 μM LDN193189 (BMP Inhibitor)[31]. The
454 neural rosette formed around day 6. Cells were passaged and medium changed to NSC medium
455 on day 22. Immunostaining for the pluripotency marker Oct4 and NPC marker Sox1 for NPCs
456 was performed on day 28 (Extended Data Fig. 3). As low density will cause NPC differentiation,
457 the cells were plated at very high density. After four weeks of culturing in NSC medium with

458 dual SMAD inhibitors, all the cells are Sox1 positive NPCs, and there are no Oct4 positive cells
459 present in all field tested by immunofluorescence. However, not all the hESCs successfully
460 differentiate into NPCs in the dual SMAD inhibitors NSC medium. Mesoderm and endoderm
461 cells spontaneously showed in the process. NPCs come from the neural rosette, pick up the
462 neural rosette when passage cells, and maintain NPCs in NSC medium monthly which could
463 potentially help to purify NPCs.

464 **Neural progenitor cells reprogram to iPSCs**

465 The dox-inducible lentiviral transduction of the transcription factors Oct4, Sox2, c-Myc, and
466 Klf4 (OSKM) is an efficient method to reprogram human NPCs to iPSCs[32,33]. We transduced
467 cells with dox-inducible lentiviral vectors carrying the mouse cDNAs encoding the four
468 transcription factors Oct4, Sox2, c-Myc, and Klf4. These cells were simultaneously infected with
469 a constitutively active lentivirus transducing the reverse tetracycline transactivator
470 (FUW-M2rtTA). The infected cells were cultured in the presence of dox, and iPSCs with a
471 human ESC like morphology were detected after approximately 2 weeks in four factors culture.

472 **Virus generation**

473 Lentiviruses were produced as described[33] in HEK293T cells (ATCC, VA) by co-transfection
474 with three helper plasmids (pRSV-REV, pMDLg/pRRE and vesicular stomatitis virus G protein
475 expression vector) (12 μg of lentiviral vector DNA and 6 μg of each of helper plasmid DNA per
476 75 cm^2 culture area) using polyethylenimine (PEI) [34]. Lentiviruses were harvested with the
477 medium 24 and 46 hr after transfection. Viral supernatant was filtered through a 0.45 μm filter,
478 pelleted by centrifugation (21,000 × g for 2 hours), resuspended in DMEM, aliquoted, and
479 snap-frozen in liquid N_2. Only virus preparations with >90% infection efficiency as assessed by
480 EGFP expression or puromycin resistance were used for experiments.

481 **Lentiviral infections**

482 Complementary DNAs for candidate genes were cloned into doxycycline-inducible lentiviral
483 vectors, as described previously[35]. Lentiviral production and hESCs, NPCs infections were

484 performed as described previously[36]. Human ESCs and NPCs were infected with concentrated
485 lentivirus and treated with doxycycline (2 mg/μl) 16～24 hr later.

486 **Immunofluorescence**

487 Cultured iN cells were fixed in 4% paraformaldehyde in PBS for 15 min at room temperature,
488 washed three times with PBS, and incubated in 0.2% Triton X-100 in PBS for 10 min at room
489 temperature. Cells were blocked in PBS containing 10% BSA for 3 hours at 4℃. Primary
490 antibodies were applied overnight. Cells were washed in PBS three times. Secondary antibodies
491 were applied for 3 hours and then washed. Cells were stained with 300 nM DAPI for indication
492 of nuclei, and mounted with mounting medium (containing antifade agent). The coverslips were
493 sealed, and the target antigen was visualized by fluorescence microscopy.

494 **Antibody**

495 The following primary antibodies with indicated dilution in blocking buffer were used: Rabbit
496 anti-Tuj1 (Covance, BioLegend, MRB-435P, 1:1,000), Mouse anti-Tuj1 (Covance, BioLegend,
497 MMS-435P, 1:1,000), Rabbit anti-human Sox1 (Millipore, AB15766, 1:1000), Mouse
498 anti-Tra-1-60 (Millipore, MAB4360, 1:1000), Rabbit anti-Ki67 (Abcam, ab15580, 1:1000), Goat
499 anti-Sox2 (Santa Cruz, Y-17, sc-17320, 1:1000), Mouse anti-Sox2 (R&D, MAB2018, 1:200),
500 Goat anti-Sox2 (R&D, AF2018, 1:200), Mouse anti-Oct-3/4 (Sata Cruz, C-10, sc-5279), Mouse
501 anit-Nestin (DSHB, rat-401, 1:200), Mouse anit-Nestin (R&D, MAB1259, 1:200), Rabbit
502 anti-ASCL1 (Abcam, ab74065, 1:200). After staining with corresponding secondary antibodies
503 in PBS plus 1% BSA, coverslips were washed four times with PBS, each for 10 min, mounted
504 with the mounting medium onto glass slides, and examined under Olympus FluoView FV1000.

505 **qRT-PCR**

506 RNA was isolated through standard methods following manufacturer's instructions (QIAGEN),
507 and reverse transcribed with SuperScript III (Life Technologies). qRT-PCR was performed with
508 FastStart Universal SYBR Green Master Mix (Roche) along with gene specific primers on a
509 7900HT real-time PCR machine (Applied Biosystems). Statistical significance was determined

510 by Students t test based on triplicated experiments.

511 **DNA methylation analysis using Infinium MethylationEPIC array**

512 Genomic DNA was isolated from the indicated hESC, NPC, and NPC-derived iN cell
513 populations through standard methods (QIAGEN) according to the manufacturer's instructions.
514 Samples were processed and run at the Stanford Functional Genomics Facility (SFGF) using the
515 Infinium MethylationEPIC array which interrogates over 850,000 methylation sites at
516 single-nucleotide resolution. All data files that were used in our analysis can be found at NCBI.
517 The GEO accession numbers are GSE268459, GSE268464, GSE268467, GSE268469,
518 GSE268471. The control brain cells came from the whole human brain[18].

519 **Calculating DMPs and DMRs across mutations and timepoints**

520 The ChAMP (Chip analysis methylation pipeline) R package was used to normalize the dataset
521 across all cell lines and time points using the BMIQ (Beta-mixture quantile dilation)
522 normalization method. Each DMP and DMR calculation was done in a pairwise fashion for each
523 mutation and wt cell line independently. We compared each mutant and wt cells at 2 days, 10
524 days, and 10 and 2 days of induction. In addition, 2 days vs. 10 days induction time points were
525 compared together. 10 days vs. 10 days and 2 days of induction were compared together. DMPs
526 were identified using the limma R package. Hypermethylated DMPs were determined as
527 positions that had a beta-value greater than 0.85 and a log fold-change greater than 0.25.
528 Hypomethylated DMPs were determined as positions that had a beta-value less than 0.25 and a
529 log fold-change less than 0.35. At 2 to 10 days comparisons and 10 to 10 + 2 iN days
530 comparisons, the same DMPs for the timeline of comparison calculated in wt cell lines were
531 removed in order to identify mutation-specific DMPs. DMRs were calculated using the
532 Bumphunter R package and annotated with the ilm10b4.hg19 genome. Both hypomethylated and
533 hypermethylated DMRs were defined as regions with FDR < 0.1, number of cpgs < 2, and
534 absolute maximum difference of CpG expression > 0.2. UCSC genome browser (GRCh37/hg19)
535 and Integrative Genomics Viewer (IGV) illustration of the methylation regions were shown in

536 Fig. 3, 4 and Extended Data Fig. 5, 6, 7. The GO analysis was performed with ClusterProfiler.
537 The DNA methylation data included in this publication have been deposited in NCBI's Gene
538 Expression Omnibus repository (GSE). https://www.ncbi.nlm.nih.gov/geo/info/geo_illu.html

539 **Statistical Tests**

540 Microsoft Excel was used for data analysis. Cumulative plots were generated from parameters
541 collected from individual cells under similar experimental conditions. Average data are presented
542 as bar graphs indicating means ± SD. Statistical comparisons between bar graphs were made
543 using the unpaired, two-tailed, Student t test (NS: $p > 0.05$, *$p < 0.01$, all versus control).

544

545 **Acknowledgements**

546 We thank the Stanford Functional Genomics Facility (SFGF) for providing excellent sequencing
547 by using the Infinium MethylationEPIC array. We also would like to thank Daniel Haag, Qianyi
548 Lee, Gernot Neumayer, Yohei Shibuya, Wendy Mei Fong, Nan Yang, Samuele Giuseppe Marro,
549 Soham Chanda, Cheen Euong Ang, Bo Zhou, Bahareh Haddad Derafshi, Madhuri Vangipuram
550 for critical discussions, sharing reagents, and providing technical assistance. Funding: This study
551 was supported by grants from the NIH (R01) and Emerson Collective.

552

553 **Author Contributions**

554 L.M. conceived and designed the study; L.M., A.Y., H.H., D.K., J.M., and Y.Y. planned and
555 performed the experiments; A.Y., H.H., D.K., and J.M. performed DNA methylation data
556 analyses; S.K.P. provided critical advice on experimental designs and data interpretations; K.S.
557 and J.A.J provided experimental material; L.M. wrote the manuscript. L.M., K.S., and A.Y.
558 edited the manuscript and the figures.

559

560 **Declaration of Interests**

561 The authors declare no conflicts of interest in the presented study.

562 **FIGURE LEGENDS**

563 **Fig. 1. Expression of oncogenic mutations blocks differentiation of ES cell into neurons**

564 **a,** Experimental scheme for generating induced neuronal (iN) cells differentiated from human
565 embryonic stem cells (hESCs) or neural progenitor cells (NPCs) by overexpression of the
566 neuronal gene, Ngn2.

567 **b,** Brightfield and fluorescent images of iN cells stained for the neural cell marker Tuj1 (red),
568 pluripotency marker Sox2 (red), cell proliferation marker Ki67 (green) from wt hESC-derived iN
569 cells and mutant hESC-derived iN cells after 2 weeks of differentiation. Data are from at least
570 three independent experiments. Scale bars, 50 μm.

571 **c,** Neurite length in cells were measured at 12 days of iN differentiation (*$p < 0.01$).

572 **d,** Quantitative RT-PCR analysis of neural cell marker Map2 expression level at 7 days of hESC
573 differentiation. Two-tailed t-test, Mean ± SD (n=3, *$p < 0.01$).

574 **e,** Brightfield and fluorescent images of NPC-derived iN cells stained for the neural cells marker
575 Tuj1 (green) from IDH1-wt NPC-derived iN cells and IDH1-R132H mutation NPC-derived iN
576 cells after 2 weeks of differentiation.

577 **f,** Quantitative RT-PCR analysis of neural cell marker Map2 expression at 7 days of NPC
578 differentiation. Values represent mean ± SD (n=3, *$p < 0.01$).

579 **g,** Cell number counting of Ki67 positive wt hESC-derived iN cells and mutant hESC-derived iN
580 cells at 10 days of differentiation (*$p < 0.01$).

581 **h,** Cell number counting of Tuj1 positive wt hESC-derived iN cells and mutant hESC-derived iN
582 cells at 10 days of differentiation (*$p < 0.01$).

583 **i,** Cell number counting of Sox2 positive wt hESC-derived iN cells and mutant hESC-derived iN
584 cells at 10 days of differentiation. Values represent the number of cells showing positive in each
585 situation ± SD (n=3, *$p < 0.01$).

586

587 **Fig. 2. Stem cells with oncogenic mutations are difficult to reprogram into iPS cells**

588 **a,** Experimental scheme for generating human induced pluripotent stem cells (hiPSCs) from
589 human NPCs by overexpression of Oct4, Sox2, Klf4, and c-Myc (OSKM) four transcription
590 factors.
591 **b,** The fluorescent images of NPC-derived iPSCs stained for the pluripotency marker TRA-1-60
592 (red) from wt NPC-derived iPSCs and mutant NPC-derived iPSCs. Nuclear staining by DAPI
593 (blue). Scale bars, 25 μm.
594 **c,** The number of TRA-1-60 positive iPSCs colonies generated from the reprogramming of wt
595 and mutant NPCs (n=3, *$p < 0.05$).
596
597 **Fig. 3. Oncogenic mutations block differentiation via epigenetic alteration of stem cell fate**
598 **genes**
599 **a,** A schematic timeline representation detailing experimental design.
600 **b,** Principal component analysis (PCA) of DNA methylation profiles of seven oncogenes and wt
601 NPCs are distinguished by DNA methylation states, from induced genes express 2 days, 10 days
602 and 10 days plus 2 days induce to iN cells. PCA plot of the different methylation samples
603 color-coded in time of induction (n=2).
604 **c,** Heatmap analysis of significant DMPs with the blue-red color scale representing the hypo- to
605 hypermethylation trend. Hierarchical clustering of rows and columns represents beta-value of
606 each group, respectively (p-value < 0.01).
607 **d,** Gene ontology (GO) term enrichment analysis in wt, IDH1, KRAS, and TP53 mutation NPCs
608 at 2 to 10 days comparisons (p-value < 0.01). GO terms show gene enrichment of wt and mutants
609 are different, but three mutants share similarity.
610 **e,** Venn diagram representing number of genes found within DMRs in IDH1, KRAS, and TP53
611 mutation cells compared with wt cells at 2 to 10 days and 10 to 10 + 2 iN days comparisons.
612 **f,** Heatmap showing genes found within DMRs in IDH1, KRAS, and TP53 mutation cells
613 compared with wt cells at 2 to 10 days and 10 to 10 + 2 iN days comparisons (p-value < 0.01).

614 Color indicates maximum CpG difference within the DMR that the gene was found within.
615 Geneset classification was identified using the hallmark genesets in GSEA. The gene
616 classification is based on their major function and hallmark pathway signaling.
617 **g,** Methylation beta-values across DMR where genes that flipped from hypo- to
618 hypermethylation across time points were found. LIN28A in TP53 mutation cells, and FOXD3 in
619 KRAS and IDH1 mutation cells. Asterisks represent DNA methylation marks at gene loci.
620 **h,** Methylation beta-values across DMR where genes were found to be changed across
621 mutational states (BAI1 and KMO) of IDH1, TP53, and KRAS mutation cells. Asterisks
622 represent DNA methylation marks at gene loci.
623
624 **Fig. 4. Stem cells with oncogenes share similar epigenetic patterns with brain tumors**
625 **a,** Venn diagram representing number of hypermethylated genes found in DMRs come from the
626 control brain VS low-grade gliomas (LGGs) with IDH1 or TP53 mutation, and wt cells VS IDH1
627 or TP53 mutation cells at 2, 10, and 10 + 2 iN days (p-value < 0.01).
628 **b,** Heatmap of differentially hypermethylated genes found within the overlap node from Fig. 4a
629 (p-value < 0.01). The colour scale indicates methylation level, from high (yellow) to low (blue).
630 **c,** The box plot demonstrates the mean DNA methylation level and variation in methylation
631 levels of genes from Fig. 4b in LGGs with IDH1-R132H, TP53-R273C, BRAF-V600E, other
632 mutations and control brain cells as measured by DNA methylation from TCGA dataset (NS: p >
633 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001). The horizontal line in
634 box-plots shows median values, and the lower and upper edges of the boxes represents the 25th
635 and 75th percentiles, respectively, and upper and lower lines outside the boxes represent
636 minimum and maximum values (error bars). Each black dot represents mean DNA methylation
637 level of one affected case in the corresponding cohort. The methylation data of control brain
638 came from whole human brain, and more details are in the method.

d, Visualization of network of hypermethylated genes. The red circles mark genes from Fig. 4b, c. The yellow circles mark chromatin regulation enzymes are on the top. The green circles mark genes which play a critical role in pluripotency that connect with genes that are hypermethylated in LGGs and connect with genes which function as chromatin regulators.

e, Visualization of the gene regulatory network. Biological Process categories were visualized as nodes clustered into functional modules and interconnected based on the indicated interaction confirmed in literature. There are five categories cycled as different colors: the red node indicates genes of pluripotency, multipotency, and differentiation (LIN28A, FOXD3, KLF4, HOXB1, MEIS, NR2F1, BAI1, HOXA5, HOXA9, PHB, FOXP1, OTX2); the green node indicates genes of neural generation and development (GBX2, LHX5, SALL1, PAX2); the blue node indicates genes of cancer (NCOR, NGFR, DDB1, FH, ILK, TPP1, ZNF540, LMO3, KLF7, NFIB); the light green node indicates genes that function in DNA, RNA binding and translation (RPA1); the purple node indicates genes that function in histone and chromatin regulation (KDM2A, PRC1/2). The red circles mark genes that are found to have methylation changes in Fig. 3f. The yellow circles mark chromatin regulation enzymes are on the top. The green circles mark genes that connect with at least three other genes. The black lines represent the interaction of two genes which has been confirmed in literature. The blue lines represent the genes connected with chromatin-modifying regulators. The arrows indicate an effect of one gene on another gene.

SUPPLEMENTARY MATERIAL

Extended Data Fig. 1. Oncogenic mutations are associated with epigenetic change from the The Cancer Genome Atlas dataset

a, Depicting the frequency of IDH1 mutation within each cancer type analysed.

b, c, The effect of IDH1-R132* mutations on epigenetic state in low-grade glioma (LGG) and glioblastoma (GBM) as measured by DNA methylation.

665 **d, e,** The 22 and 17 candidate genes in gain-of-function and loss-of-function mutations
666 respectively (1% FDR, false discovery rate) involved in regulating the epigenetic landscape in
667 cancer, by performing a cancer analysis linking both gain-of-function and loss-of-function
668 mutations to changes in the epigenetic state by measure DNA methylation.
669
670 **Extended Data Fig. 2. Stem cells with oncogenic mutations retain their potential for**
671 **proliferation in differentiation**
672 **a,** Cell number counting of wt hESCs and mutant hESCs after 3 days of dox-induction (n=3, NS:
673 $p > 0.05$).
674 **b,** Cell number counting of wt hESC-derived iN cells and mutant hESC-derived iN cells at 7
675 days of differentiation (n=3, *$p < 0.01$).
676 **c,** The ratio of the Sox2 positive hESC-derived iN cell number and Tuj1 positive hESC-derived
677 iN cell number at 10 days of differentiation (*$p < 0.01$).
678
679 **Extended Data Fig. 3. Optimize differentiation protocol for generation of NPCs from**
680 **hESCs**
681 **a,** Experimental scheme for generating human NPCs differentiated from hESCs use dual SMAD
682 inhibitors.
683 **b,** Brightfield of hESCs at day 1, embryoid bodies (EBs) at day 3, neural rosette at day 6, photo
684 NPCs at day 28, and immunostaining pluripotency marker Oct4 (red) (yellow arrow) and NPC
685 marker Sox1 (green) for NPCs at day 28. Nuclear staining by DAPI (blue). Scale bars, 25 μm.
686
687 **Extended Data Fig. 4. Cancer cells are difficult to reprogramming and differentiation**
688 **a,** The morphology of U251 cells. U251 cell line is an IDH1 mutation glioblastoma cell line.
689 Scale bars, 25 μm.

690 **b, c,** Brightfield and fluorescent images of immunostaining TRA-1-60 (red) of U251 cells at 14
691 days reprogramming with OSKM for colony identification.
692 **d, e,** Brightfield and fluorescent images of immunostaining U251 cells at 14 days differentiation
693 by Ngn2 overexpression for the neural cells marker Tuj1 (green).
694 **f,** The morphology of K562 cells. K562 cell line is an IDH1 mutant leukemia cell line.
695 **g, h,** Brightfield and fluorescent images of K562 cells immunostaining for TRA-1-60 (green) at
696 14 days reprogramming with OSKM. No TRA-1-60 positive staining colony.
697 **i, j,** Brightfield and fluorescent images of staining Tuj1 (red) and DAPI (blue) on K562 cells at
698 14 days differentiation by Ngn2 overexpression.
699 **k,** The morphology of HCT116 cells. HCT116 cell line is an IDH1 mutant colon cancer cell line.
700 **l, m,** Brightfield and fluorescent images of staining TRA-1-60 (green) and DAPI (blue) on
701 HCT116 cells at 14 days reprogramming with OSKM. HCT116 cells cluster on feeders and grow
702 more as colony-like clumps as shown in the photo. All cells are TRA-1-60 positive.
703 **n, o,** Brightfield and fluorescent images of staining Tuj1 (red) and DAPI (blue) on HCT116 cells
704 at 14 days differentiation by Ngn2 overexpression.
705 **p, q, r, s,** Immunostaining U251 cells with the pluripotency marker Sox2 (red), neural
706 stem/progenitor cell markers Nestin (green) and ASCL1 (red), and neural cell marker Tuj1
707 (green).
708
709 **Extended Data Fig. 5. Oncogenic mutations are associated with epigenetic change in hESC**
710 **cells**
711 **a,** Principal component analysis (PCA) of DNA methylation profiles using Infinium
712 MethylationEPIC BeadChip assay demonstrates that seven oncogene mutants and one wt hESC
713 cells are distinguished by DNA methylation states. Each point represents wt as control or a gene
714 modified cell sample. The proportions of variance explained by PC1 and PC2 are indicated.
715 **b,** Heatmap plot shows the potential involvement of the most significant DNA methylation

716 changes in mutant hESCs compared with wt hESCs by Infinium MethylationEPIC BeadChip
717 assay.
718 **c,** Genomic location of differentially methylated sites (DMS), for CpG sites methylated in wt and
719 mutant hESCs. Odds ratio and 95% confidence intervals are displayed in wt and mutant hESCs,
720 comparing their localization in different genomic locations as provided by Illumina (TSS1500,
721 TSS200, 5′UTR, 1stExon, Body, Exon boundaries [ExonBnd], and 3′UTR). Odds ratios were
722 computed against the general distribution of the 684,838 CpGs of our dataset. Indicated 282 out
723 of 684,838 CpG sites are differentially methylated in seven mutant hESCs compared with wt
724 hESCs, 246 out of 286 are associated certain gene.
725 **d,** The distribution of DMRs between wt hESCs and mutant hESCs. 321 genomic regions are
726 differentially methylated ($P < 0.05$) (in total 4,905 regions from UCSC refgene).
727 **e,** Enrichment of DMRs according to genic regions (left) and distance from CpG island (right)
728 (with a CpG shore, CpG shelve and open sea located 1-2 kbp, 2-4 kbp and >4 kbp away from the
729 nearest CpG island, respectively) between wt and mutant hESCs. 401 genomic regions are
730 differentially methylated ($P < 0.05$) (in total 4,905 regions).
731 **f, g,** Gene Ontology (GO) categories of genes differentially methylated between wt and mutant
732 hESCs (top categories ranked according to p-value). GO enrichment analysis by DAVID using
733 DMPs associated 246 genes, and DMRs associated 125 genes.
734
735 **Extended Data Fig. 6. Oncogenic mutations are associated with epigenetic alteration in**
736 **NPCs and iN cells**
737 **a, b, c,** Number of hypermethylated (beta-value > 0.8 and logFC > 0.3) and hypomethylated
738 (beta-value < 0.2, logFC < -0.3) DMPs in mutant cells compared to wt cells at 2 days, 10 days,
739 and 10 days plus 2 days of neural induction.
740 **d, e, f,** Bar plot shows the percentage of genomic location each DMP was found in mutant cells
741 compared to wt cells at 2 days, 10 days, and 10 days plus 2 days of neural induction (p-value <

742 0.01).

743 **g,** Number of DMPs and DMRs at 2 to 10 days and 10 to 10 + 2 iN days comparisons for mutant
744 cells (IDH1, KRAS, and TP53). DMPs found in the same comparison for wt cells were removed
745 from the total DMP count.

746 **h,** Number of DMS within the DMRs in (G) across the IDH1, KRAS, and TP53 mutation cells.
747 Black or grey color bar plots represent DMRs found in hyper- or hypomethylated regions.

748

749 **Extended Data Fig. 7. Stem cells with oncogenic mutations share similar epigenetic**
750 **patterns with brain tumors**

751 **a,** Venn diagram representing the number of hypomethylated genes found in DMRs that come
752 from the control brain VS low-grade gliomas (LGGs) with IDH1 or TP53 mutation, and wt cells
753 VS IDH1 or TP53 mutation cells at 2, 10, and 10 + 2 iN days (p-value < 0.01).

754 **b,** Heatmap of differentially hypomethylated genes found within the overlap node from extended
755 data Fig. 7a at 2, 10, and 10 + 2 iN days (p-value < 0.01). The color scale indicates methylation
756 level from high (yellow) to low (blue).

757 **c,** Venn diagram representing the number of genes found in DMRs that come from the control
758 brain VS LGGs with IDH1 (TP53) mutation, and wt cells VS IDH1 (TP53) mutation cells, at 2
759 to 10 days comparisons and 10 to 10 + 2 iN days comparisons (p-value < 0.01).

760 **d,** Heatmap of differentially methylated genes found within the overlap node from extended data
761 Fig. 7c at 2 to 10 days comparisons and 10 to 10 + 2 iN days comparisons (p-value < 0.01). The
762 color scale indicates methylation level from high (yellow) to low (blue).

763 **e, f,** From comparing the epigenetic states in LGGs with IDH1-R132H, TP53-R273C,
764 BRAF-V600E, other mutations and control brain cells as measured by DNA methylation from
765 the TCGA dataset, the box plot demonstrates the distribution of DNA methylation of
766 hypermethylated genes in cells at 2 to 10 days comparisons and 10 to 10 + 2 iN days
767 comparisons. (NS: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001).

768 **g,** Distribution of DNA methylation marks at CASZ1 and VGLL4 gene loci. The arrows indicate
769 differentially methylation sites.
770
771

772 **REFERENCES**

773

774 1 Chao, M. P. *et al.* Human AML-iPSCs reacquire leukemic properties after differentiation and model clonal
775 variation of disease. *Cell Stem Cell* **20**, 329-344 e327 (2017).
776 2 Turcan, S. *et al.* IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*
777 **483**, 479-483 (2012).
778 3 Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744
779 (2009).
780 4 Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation.
781 *Nature* **483**, 474-478 (2012).
782 5 Cloos, P. A. *et al.* The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone
783 H3. *Nature* **442**, 307-311 (2006).
784 6 Zhang, Y. *et al.* Rapid single-step induction of functional neurons from human pluripotent stem cells.
785 *Neuron* **78**, 785-798 (2013).
786 7 Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of
787 SMAD signaling. *Nat Biotechnol* **27**, 275-280 (2009).
788 8 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult
789 fibroblast cultures by defined factors. *Cell* **126** (2006).
790 9 Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup
791 of glioma. *Cancer Cell* **17**, 510-522 (2010).
792 10 Gouti, M. & Gavalas, A. Hoxb1 controls cell fate specification and proliferative capacity of neural stem
793 and progenitor cells. *Stem Cells* **26**, 1985-1997 (2008).
794 11 Qin, S., Liu, M., Niu, W. & Zhang, C. L. Dysregulation of kruppel-like factor 4 during brain development
795 leads to hydrocephalus in mice. *Proc Natl Acad Sci U S A* **108**, 21117-21121 (2011).
796 12 Li, X. *et al.* Foxp1 regulates cortical radial migration and neuronal morphogenesis in developing cerebral
797 cortex. *PLoS One* **10**, e0127671 (2015).
798 13 Cimadamore, F., Amador-Arjona, A., Chen, C., Huang, C. T. & Terskikh, A. V. SOX2-LIN28/let-7
799 pathway regulates proliferation and neurogenesis in neural precursors. *Proc Natl Acad Sci U S A* **110**,
800 E3017-3026 (2013).
801 14 Krishnakumar, R. *et al.* FOXD3 regulates pluripotent stem cell potential by simultaneously initiating and
802 repressing enhancer activity. *Cell Stem Cell* **18**, 104-117 (2016).
803 15 Zhu, D., Hunter, S. B., Vertino, P. M. & Van Meir, E. G. Overexpression of MBD2 in glioblastoma
804 maintains epigenetic silencing and inhibits the antiangiogenic function of the tumor suppressor gene BAI1.
805 *Cancer Res* **71**, 5859-5870 (2011).
806 16 Zhu, D. *et al.* BAI1 suppresses medulloblastoma formation by protecting p53 from Mdm2-mediated
807 degradation. *Cancer Cell* **33**, 1004-1016 e1005 (2018).
808 17 Jones, S. P., Guillemin, G. J. & Brew, B. J. The kynurenine pathway in stem cell biology. *Int J Tryptophan*
809 *Res* **6**, 57-66 (2013).

810 18 Braun, P. R. *et al.* Genome-wide DNA methylation comparison between live human brain and peripheral
811 tissues within individuals. *Transl Psychiatry* **9**, 47 (2019).
812 19 Alfaro, M. P. *et al.* sFRP2 suppression of bone morphogenic protein (BMP) and Wnt signaling mediates
813 mesenchymal stem cell (MSC) self-renewal promoting engraftment and myocardial repair. *J Biol Chem*
814 **285**, 35645-35653 (2010).
815 20 Wu, Q., Yin, X., Zhao, W., Xu, W. & Chen, L. Downregulation of SFRP2 facilitates cancer stemness and
816 radioresistance of glioma cells via activating Wnt/beta-catenin signaling. *PLoS One* **16**, e0260864 (2021).
817 21 Lavado, A. *et al.* The Hippo pathway prevents YAP/TAZ-driven hypertranscription and controls neural
818 progenitor number. *Dev Cell* **47**, 576-591 e578 (2018).
819 22 Wu, L. M. N. *et al.* Programming of schwann cells by Lats1/2-TAZ/YAP signaling drives malignant
820 peripheral nerve sheath tumorigenesis. *Cancer Cell* **33**, 292-308 e297 (2018).
821 23 Herrlinger, S. *et al.* Lin28-mediated temporal promotion of protein synthesis is crucial for neural progenitor
822 cell maintenance and brain development in mice. *Development* **146** (2019).
823 24 Le Thomas, A. *et al.* Piwi induces piRNA-guided transcriptional silencing and establishment of a
824 repressive chromatin state. *Genes Dev* **27**, 390-399 (2013).
825 25 Lagutin, O. V. *et al.* Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for
826 vertebrate forebrain development. *Genes Dev* **17**, 368-379 (2003).
827 26 Berkyurek, A. C. *et al.* The DNA methyltransferase Dnmt1 directly interacts with the SET and RING
828 finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic
829 center to hemi-methylated DNA. *J Biol Chem* **289**, 379-386 (2014).
830 27 Kim, K. Y. *et al.* Uhrf1 regulates active transcriptional marks at bivalent domains in pluripotent stem cells
831 through Setd1a. *Nat Commun* **9**, 2583 (2018).
832 28 Ramesh, V. *et al.* Loss of Uhrf1 in neural stem cells leads to activation of retroviral elements and delayed
833 neurodegeneration. *Genes Dev* **30**, 2199-2212 (2016).
834 29 Kaur, A. *et al.* sFRP2 in the aged microenvironment drives melanoma metastasis and therapy resistance.
835 *Nature* **532**, 250-254 (2016).
836 30 Kasashima, H. *et al.* Stromal SOX2 upregulation promotes tumorigenesis through the generation of a
837 SFRP1/2-expressing cancer-associated fibroblast population. *Dev Cell* **56**, 95-110 e110 (2021).
838 31 Topol, A., Tran, N. N. & Brennand, K. J. A guide to generating and using hiPSC derived NPCs for the
839 study of neurological diseases. *J Vis Exp*, e52495 (2015).
840 32 Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**,
841 1035-1041 (2010).
842 33 Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse
843 somatic cells. *Cell Stem Cell* **2**, 151-159 (2008).
844 34 Pang, Z. P., Cao, P., Xu, W. & Sudhof, T. C. Calmodulin controls synaptic strength via presynaptic
845 activation of calmodulin kinase II. *J Neurosci* **30**, 4132-4142 (2010).
846 35 Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during
847 fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**, 230-240 (2008).
848 36 Chen, C. & Okayama, H. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol Cell*

849 *Biol* **7**, 2745-2752 (1987).

850

Figure 1

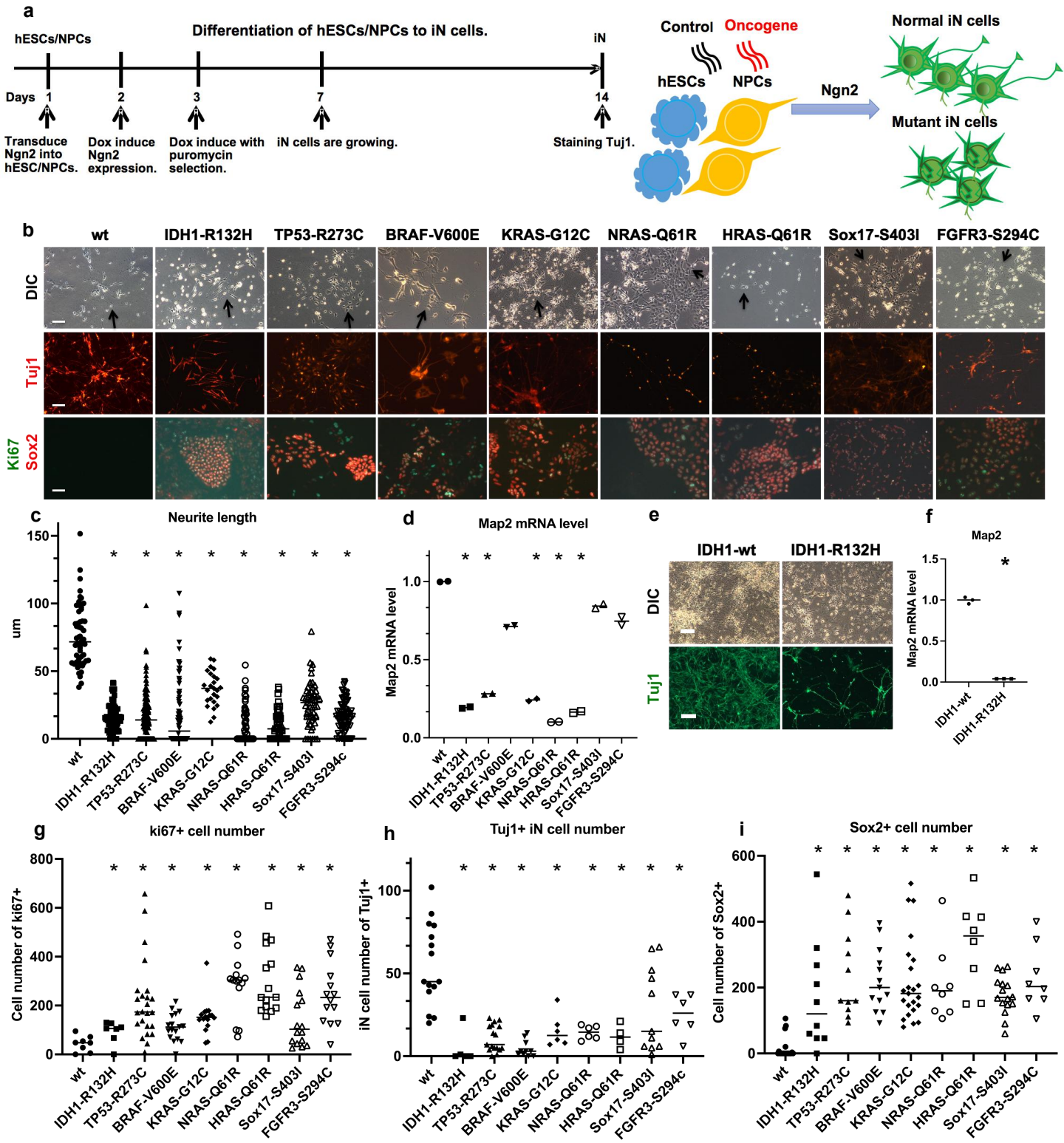
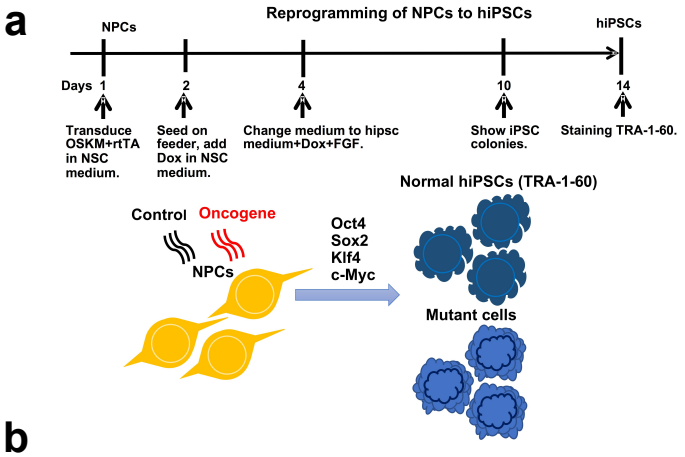
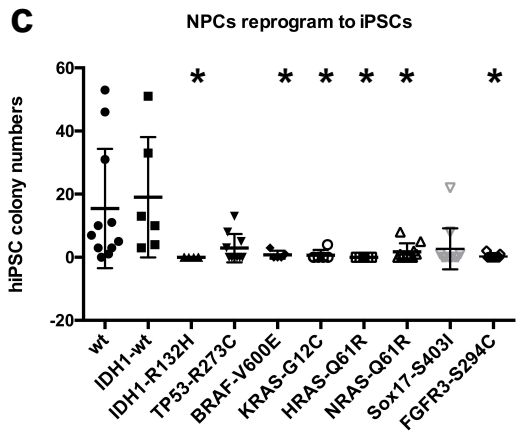

a
Differentiation of hESCs/NPCs to iN cells.
hESCs/NPCs
iN
Days 1 2 3 7 14
Transduce Ngn2 into hESC/NPCs.
Dox induce Ngn2 expression.
Dox induce with puromycin selection.
iN cells are growing.
Staining Tuj1.
Control
Oncogene
hESCs
NPCs
Ngn2
Normal iN cells
Mutant iN cells
b
wt
IDH1-R132H
TP53-R273C
BRAF-V600E
KRAS-G12C
NRAS-Q61R
HRAS-Q61R
Sox17-S403I
FGFR3-S294C
DIC
Tuj1
Ki67
Sox2
c
Neurite length
um
d
Map2 mRNA level
e
IDH1-wt
IDH1-R132H
DIC
Tuj1
f
Map2
Map2 mRNA level
g
ki67+ cell number
Cell number of ki67+
h
Tuj1+ iN cell number
iN cell number of Tuj1+
i
Sox2+ cell number
Cell number of Sox2+

Figure 2


a
Reprogramming of NPCs to hiPSCs
NPCs
hiPSCs
Days 1
2
4
10
14
Transduce OSKM+rtTA in NSC medium.
Seed on feeder, add Dox in NSC medium.
Change medium to hipsc medium+Dox+FGF.
Show iPSC colonies.
Staining TRA-1-60.
Normal hiPSCs (TRA-1-60)
Control
Oncogene
NPCs
Oct4
Sox2
Klf4
c-Myc
Mutant cells


c
NPCs reprogram to iPSCs
hiPSC colony numbers
60
40
20
0
-20
wt
IDH1-wt
IDH1-R132H
TP53-R273C
BRAF-V600E
KRAS-G12C
HRAS-Q61R
NRAS-Q61R
Sox17-S403I
FGFR3-S294C

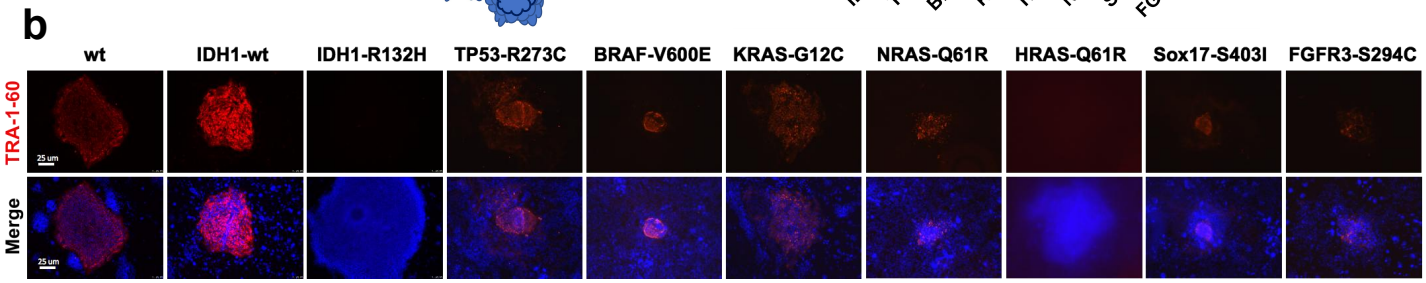

b
wt
IDH1-wt
IDH1-R132H
TP53-R273C
BRAF-V600E
KRAS-G12C
NRAS-Q61R
HRAS-Q61R
Sox17-S403I
FGFR3-S294C
TRA-1-60
Merge
25 um
25 um

Figure 3

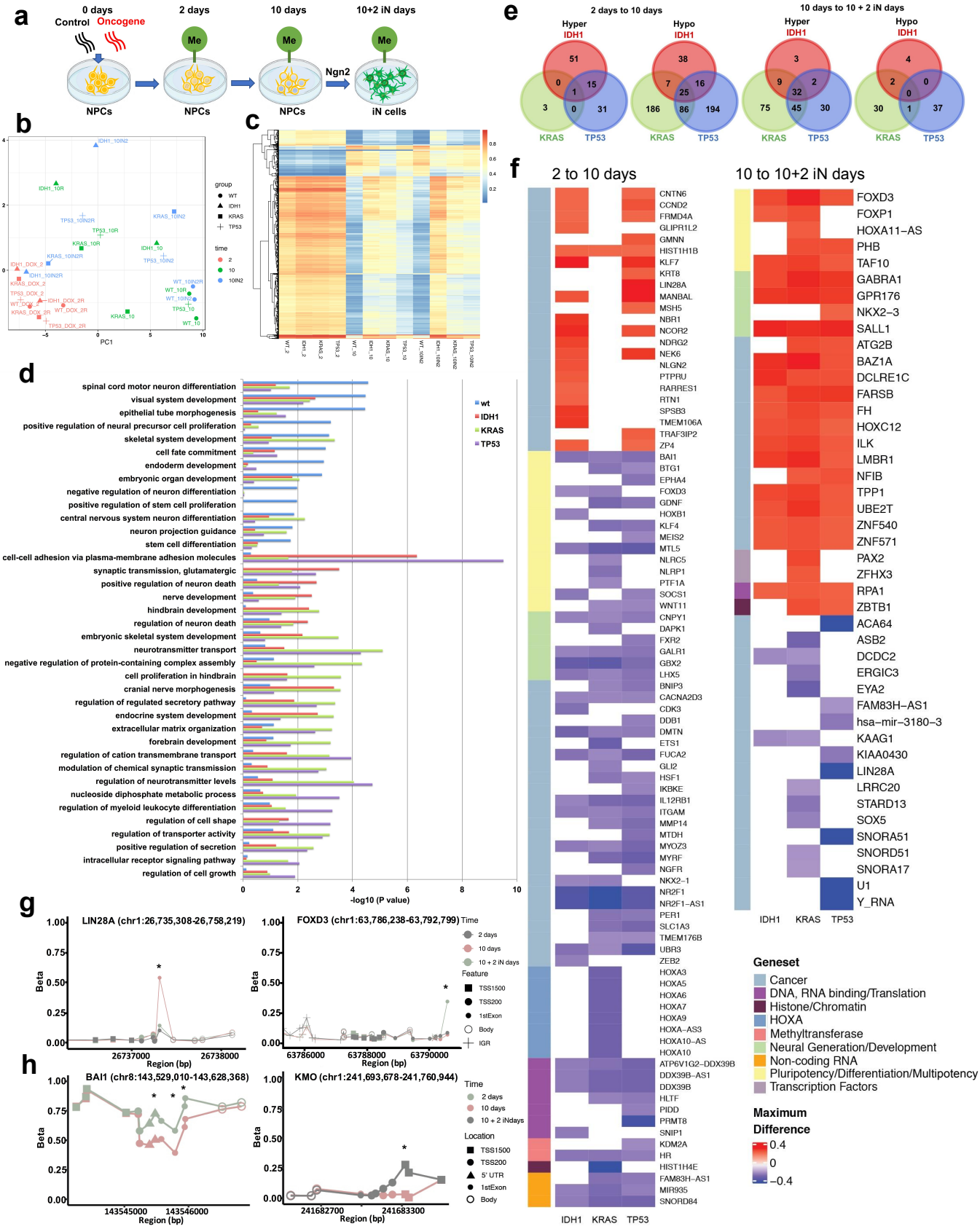
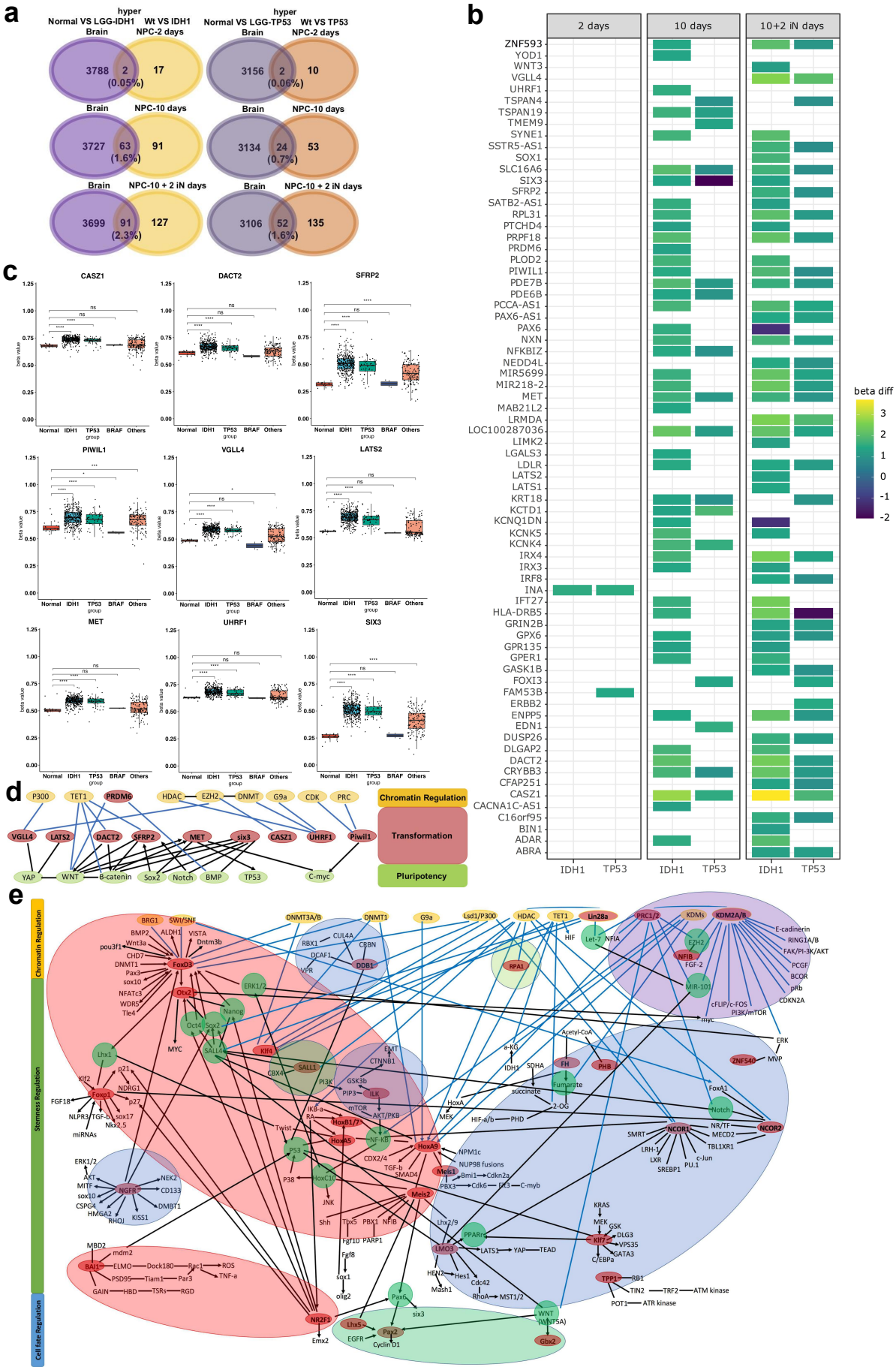
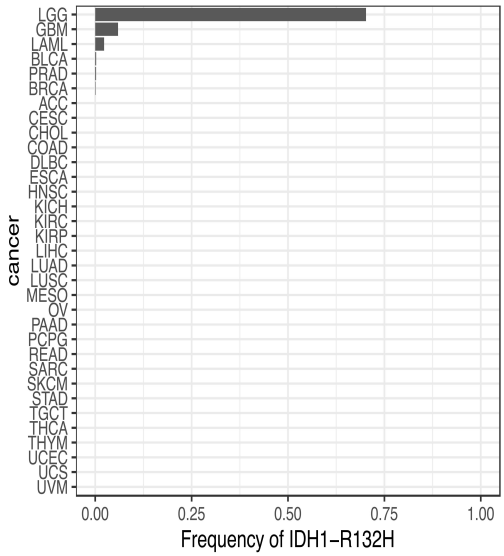

a
0 days
Control
Oncogene
2 days
10 days
10+2 iN days
Me
Ngn2
NPCs
iN cells
b
PC1
PC2
group
WT
IDH1
KRAS
TP53
time
2
10
10IN2
c
d
spinal cord motor neuron differentiation
visual system development
epithelial tube morphogenesis
positive regulation of neural precursor cell proliferation
skeletal system development
cell fate commitment
endoderm development
embryonic organ development
negative regulation of neuron differentiation
positive regulation of stem cell proliferation
central nervous system neuron differentiation
neuron projection guidance
stem cell differentiation
cell-cell adhesion via plasma-membrane adhesion molecules
synaptic transmission, glutamatergic
positive regulation of neuron death
nerve development
hindbrain development
regulation of neuron death
embryonic skeletal system development
neurotransmitter transport
negative regulation of protein-containing complex assembly
cell proliferation in hindbrain
cranial nerve morphogenesis
regulation of regulated secretory pathway
endocrine system development
extracellular matrix organization
forebrain development
regulation of cation transmembrane transport
modulation of chemical synaptic transmission
regulation of neurotransmitter levels
nucleoside diphosphate metabolic process
regulation of myeloid leukocyte differentiation
regulation of cell shape
regulation of transporter activity
positive regulation of secretion
intracellular receptor signaling pathway
regulation of cell growth
wt
IDH1
KRAS
TP53
-log10 (P value)
e
2 days to 10 days
10 days to 10 + 2 iN days
Hyper
Hypo
IDH1
KRAS
TP53
f
2 to 10 days
10 to 10+2 iN days
IDH1
KRAS
TP53
Geneset
Cancer
DNA, RNA binding/Translation
Histone/Chromatin
HOXA
Methyltransferase
Neural Generation/Development
Non-coding RNA
Pluripotency/Differentiation/Multipotency
Transcription Factors
Maximum Difference
0.4
0
−0.4
g
LIN28A (chr1:26,735,308-26,758,219)
FOXD3 (chr1:63,786,238-63,792,799)
Beta
Region (bp)
Time
2 days
10 days
10 + 2 IN days
Feature
TSS1500
TSS200
1stExon
Body
IGR
h
BAI1 (chr8:143,529,010-143,628,368)
KMO (chr1:241,693,678-241,760,944)
Time
2 days
10 days
10 + 2 iNdays
Location
TSS1500
TSS200
5' UTR
1stExon
Body

Figure 4

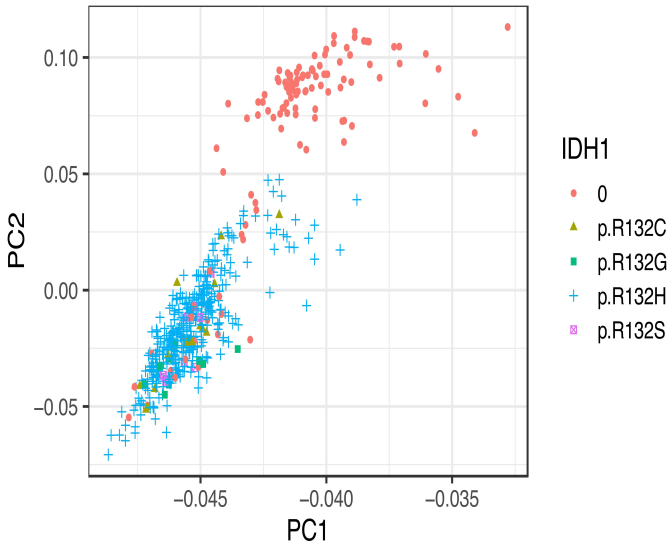

a
hyper
Normal VS LGG-IDH1
Wt VS IDH1
Brain
NPC-2 days
3788
2
(0.05%)
17
NPC-10 days
3727
63
(1.6%)
91
NPC-10 + 2 iN days
3699
91
(2.3%)
127
hyper
Normal VS LGG-TP53
Wt VS TP53
3156
2
(0.06%)
10
3134
24
(0.7%)
53
3106
52
(1.6%)
135
b
2 days
10 days
10+2 iN days
IDH1
TP53
beta diff
c
CASZ1
DACT2
SFRP2
PIWIL1
VGLL4
LATS2
MET
UHRF1
SIX3
beta value
group
Normal
IDH1
TP53
BRAF
Others
d
P300
TET1
PRDM6
HDAC
EZH2
DNMT
G9a
CDK
PRC
Chromatin Regulation
VGLL4
LATS2
DACT2
SFRP2
MET
six3
CASZ1
UHRF1
Piwil1
Transformation
YAP
WNT
β-catenin
Sox2
Notch
BMP
TP53
C-myc
Pluripotency
e
Chromatin Regulation
Stemness Regulation
Cell fate Regulation
BRG1
SWI/SNF
DNMT3A/B
DNMT1
G9a
Lsd1/P300
HDAC
TET1
Lin28a
PRC1/2
KDMs
KDM2A/B
FoxD3
Otx2
Nanog
Oct4
Sox2
SALL4
Klf4
Lhx1
Foxp1
ERK1/2
Meis1
Meis2
HoxA5
HoxB1/2
HoxC10
NF-κB
P53
NGFR
NR2F1
Pax6
Pax2
Lhx5
Gbx2
WNT
(WNT5A)
LMO3
PPARγ
Klf7
TPP1
RB1
ZNF540
MVP
FoxA1
Notch
NCOR1
NCOR2
FH
PHB
Fumarate
succinate
2-OG
Acetyl-CoA
s-KG
IDH1
SDHA
HIF-a/b
PHD
HoxA
MEK
mTOR
AKT/PKB
ILK
PI3K
GSK3b
EMT
CTNNB1
Twist
IK6-a
RA
CDX2/4
TGF-b
SMAD4
NPM1c
NUP98 fusions
Bmi1
Cdkn2a
PBX3
Cdk6
Fli3
C-myc
JNK
P38
Shh
Tbx5
PBX1
NFIB
Fgf10
PARP1
Fgf8
sox1
olig2
Emx2
EGFR
Cyclin D1
six3
Hes1
Hes2
Mash1
Cdc42
RhoA
MST1/2
LATS1
YAP
TEAD
KRAS
MEK
GSK
DLG3
VPS35
GATA3
C/EBPa
TIN2
TRF2
ATM kinase
POT1
ATR kinase
ROS
TNF-a
ELMO
Dock180
Rac1
PSD95
Tiam1
Par3
GAIN
HBD
TSRs
RGD
MBD2
mdm2
BAI1
ERK
MVP
c-Jun
PU.1
SREBP1
LXR
LRH-1
SMRT
NR/TF
MECD2
TBL1XR1
E-cadinerin
RING1A/B
FAK/PI-3K/AKT
PCGF
BCOR
pRb
CDKN2A
cFLIP/c-FOS
PI3K/mTOR
myc
FGF-2
Let-7
NFIA
EZH2
NFIB
MIR-101
FGF18
NLPR3
TGF-b
sox17
Nkx2.5
miRNAs
p21
NDRG1
p27
ERK1/2
AKT
MITF
sox10
CSPG4
HMGA2
RHOJ
KISS1
NEK2
CD133
DMBT1
MYC
ALDH1
VISTA
Dntm3b
BMP2
Wnt3a
pou3f1
CHD7
DNMT1
Pax3
sox10
NFATc3
WDR5
Tle4
CUL4A
RBX1
DCAF1
VPR
CBBN
DDB1
RPA1
HIF
Let-7
NFIA
CTNNB1
GBX4
SALL1
PI3K
PIP3
GSK3b
ILK
mTOR
AKT/PKB
IK6-a
RA
HoxB1/2
HoxA5
NF-κB
HoxA9
P53
HoxC10
Meis1
Meis2
Bmi1
Cdkn2a
PBX3
Cdk6
Fli3
C-myc
Lhx2/9
PPARγ
LMO3
Klf7
TPP1
RB1
Pax6
Pax2
Lhx5
Gbx2
WNT
(WNT5A)
NR2F1
Emx2
EGFR
Cyclin D1
six3
Hes1
Hes2
Mash1
Cdc42
RhoA
MST1/2
LATS1
YAP
TEAD
KRAS
MEK
GSK
DLG3
VPS35
GATA3
C/EBPa
TIN2
TRF2
ATM kinase
POT1
ATR kinase
ROS
TNF-a
ELMO
Dock180
Rac1
PSD95
Tiam1
Par3
GAIN
HBD
TSRs
RGD
MBD2
mdm2
BAI1
ERK
MVP
c-Jun
PU.1
SREBP1
LXR
LRH-1
SMRT
NR/TF
MECD2
TBL1XR1
E-cadinerin
RING1A/B
FAK/PI-3K/AKT
PCGF
BCOR
pRb
CDKN2A
cFLIP/c-FOS
PI3K/mTOR
myc
FGF-2
Let-7
NFIA
EZH2
NFIB
MIR-101
FGF18
NLPR3
TGF-b
sox17
Nkx2.5
miRNAs
p21
NDRG1
p27
ERK1/2
AKT
MITF
sox10
CSPG4
HMGA2
RHOJ
KISS1
NEK2
CD133
DMBT1
MYC
ALDH1
VISTA
Dntm3b
BMP2
Wnt3a
pou3f1
CHD7
DNMT1
Pax3
sox10
NFATc3
WDR5
Tle4
CUL4A
RBX1
DCAF1
VPR
CBBN
DDB1
RPA1
HIF
Let-7
NFIA
CTNNB1
GBX4
SALL1
PI3K
PIP3
GSK3b
ILK
mTOR
AKT/PKB
IK6-a
RA
HoxB1/2
HoxA5
NF-κB
HoxA9
P53
HoxC10
Meis1
Meis2
Bmi1
Cdkn2a
PBX3
Cdk6
Fli3
C-myc
Lhx2/9
PPARγ
LMO3
Klf7
TPP1
RB1
Pax6
Pax2
Lhx5
Gbx2
WNT
(WNT5A)
NR2F1
Emx2
EGFR
Cyclin D1
six3
Hes1
Hes2
Mash1
Cdc42
RhoA
MST1/2
LATS1
YAP
TEAD
KRAS
MEK
GSK
DLG3
VPS35
GATA3
C/EBPa
TIN2
TRF2
ATM kinase
POT1
ATR kinase
ROS
TNF-a
ELMO
Dock180
Rac1
PSD95
Tiam1
Par3
GAIN
HBD
TSRs
RGD
MBD2
mdm2
BAI1
ERK
MVP
c-Jun
PU.1
SREBP1
LXR
LRH-1
SMRT
NR/TF
MECD2
TBL1XR1
E-cadinerin
RING1A/B
FAK/PI-3K/AKT
PCGF
BCOR
pRb
CDKN2A
cFLIP/c-FOS
PI3K/mTOR
myc
FGF-2
Let-7
NFIA
EZH2
NFIB
MIR-101

Supplemental Figure 1

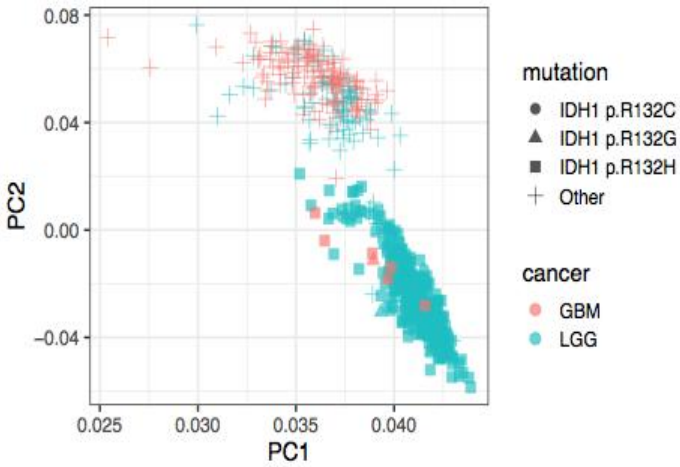
a



b IDH1 mutations in LGG



c IDH1-R132* mutations in GBM and LGG



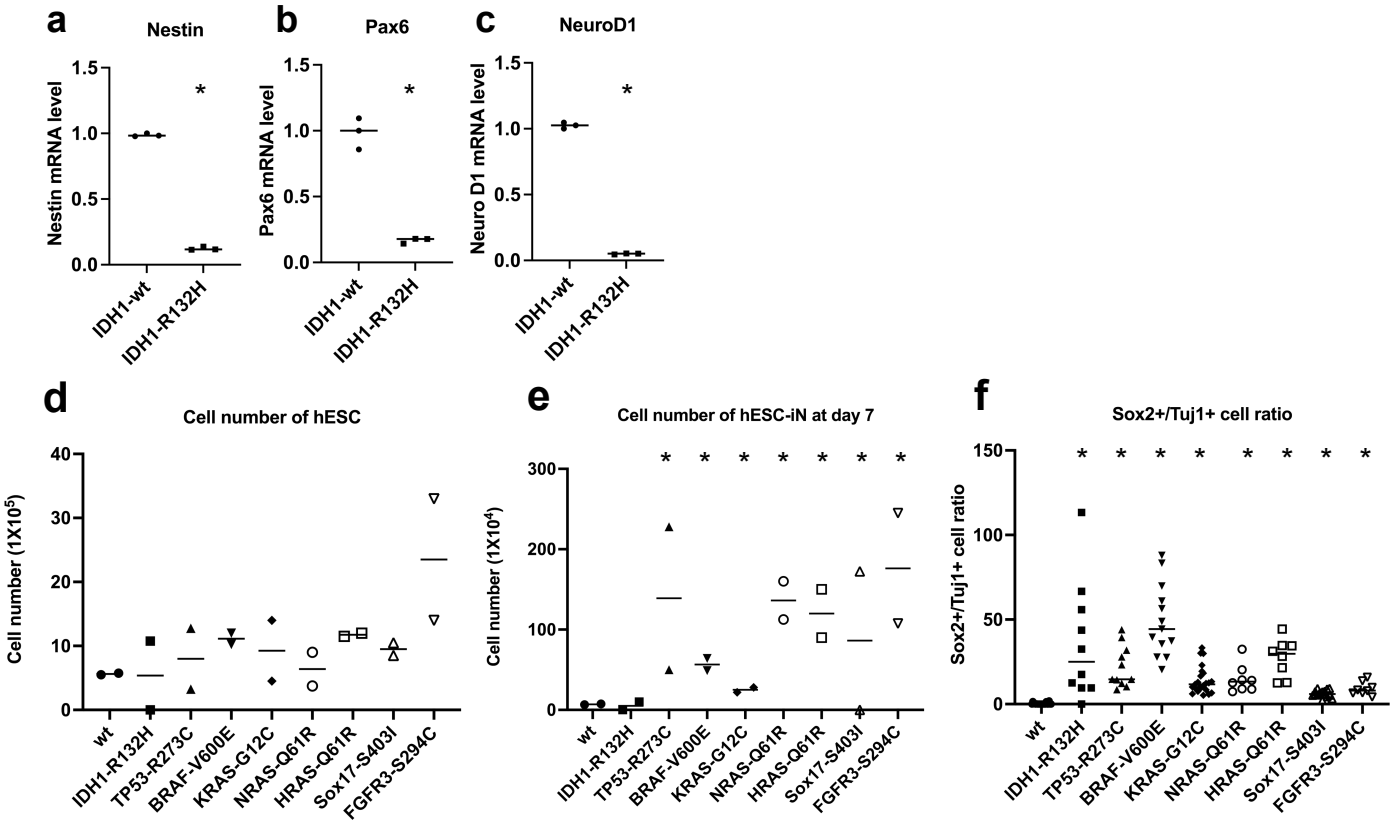
d gain-of-function mutations

Cancer	Gene	Mutation	p_adjust
THCA	BRAF	V600E	1.96E-44
LGG	IDH1	R132H	4.76E-25
THYM	GTF2I	L424H	8.13E-15
COAD	BRAF	V600E	7.88E-14
SKCM	BRAF	V600E	1.11E-10
THCA	NRAS	Q61R	5.05E-10
UCEC	PPP2R1A	P179R	1.57E-07
BLCA	FGFR3	S249C	2.29E-06
THCA	HRAS	Q61R	1.34E-05
LGG	TP53	R273C	1.88E-03
LGG	IDH1	R132G	3.14E-03
LGG	IDH1	R132C	6.37E-03
UCEC	CTGLF11P	H132R	8.40E-03
PCPG	HRAS	Q61R	1.34E-02
LUAD	BRAF	V600E	1.38E-02
BLCA	FGFR3	Y373C	2.74E-02
THCA	NRAS	Q61K	3.05E-02
SKCM	IDH1	R132C	3.32E-02
UCEC	SOX17	S403I	3.35E-02
LGG	IDH2	R172K	4.48E-02
LGG	TP53	R175H	4.52E-02
LUAD	KRAS	G12C	4.90E-02

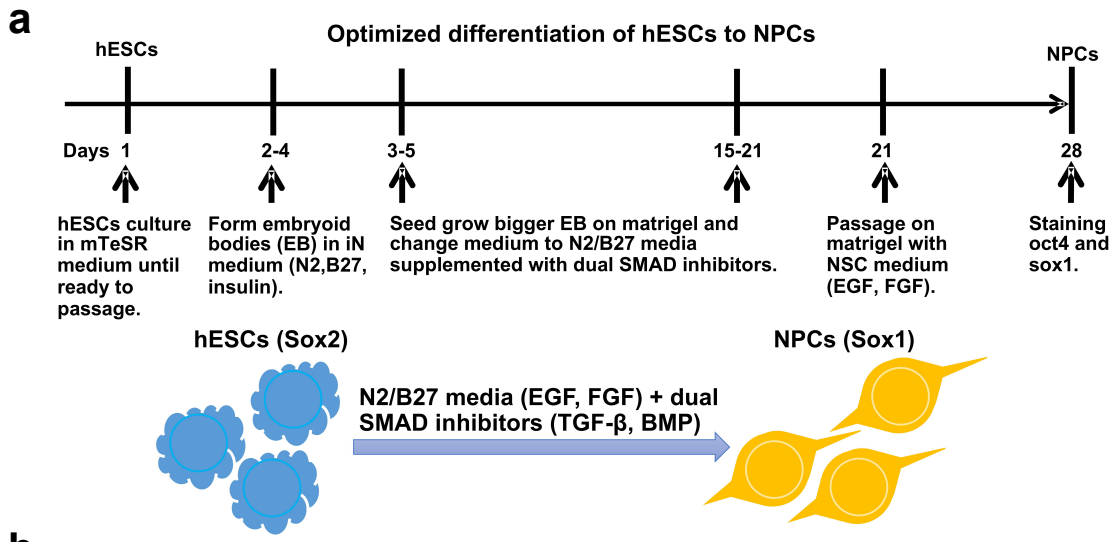
e loss-of-function mutations

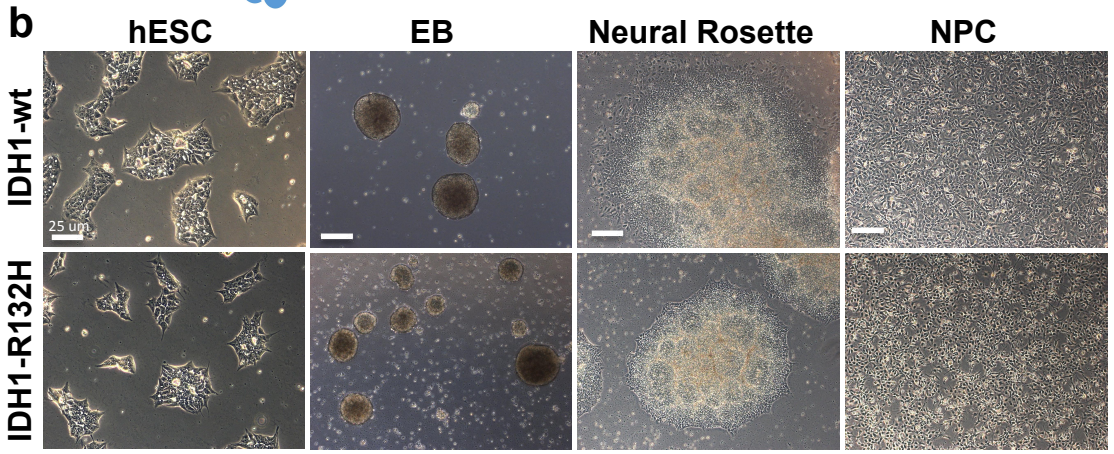
Gene	Cancer	PC	p_adjust
PTEN	UCEC	2	4.22E-09
ATRX	LGG	4	2.10E-05
TP53	LGG	4	1.47E-03
KMT2D	BLCA	6	1.47E-03
STAG2	BLCA	4	1.47E-03
AXIN1	LIHC	9	3.41E-03
RB1	BLCA	8	4.86E-03
NSD1	UCEC	4	4.96E-03
BMPR2	COAD	2	4.96E-03
ABCA13	UCEC	4	4.96E-03
ARID1A	UCEC	2	4.96E-03
SMARCA4	LUAD	4	7.86E-03
DYNC2H1	UCEC	4	8.17E-03
NSD1	UCEC	7	8.46E-03
TTN	STAD	2	8.46E-03
APC	COAD	3	8.46E-03
SMARCA4	LUAD	6	9.73E-03

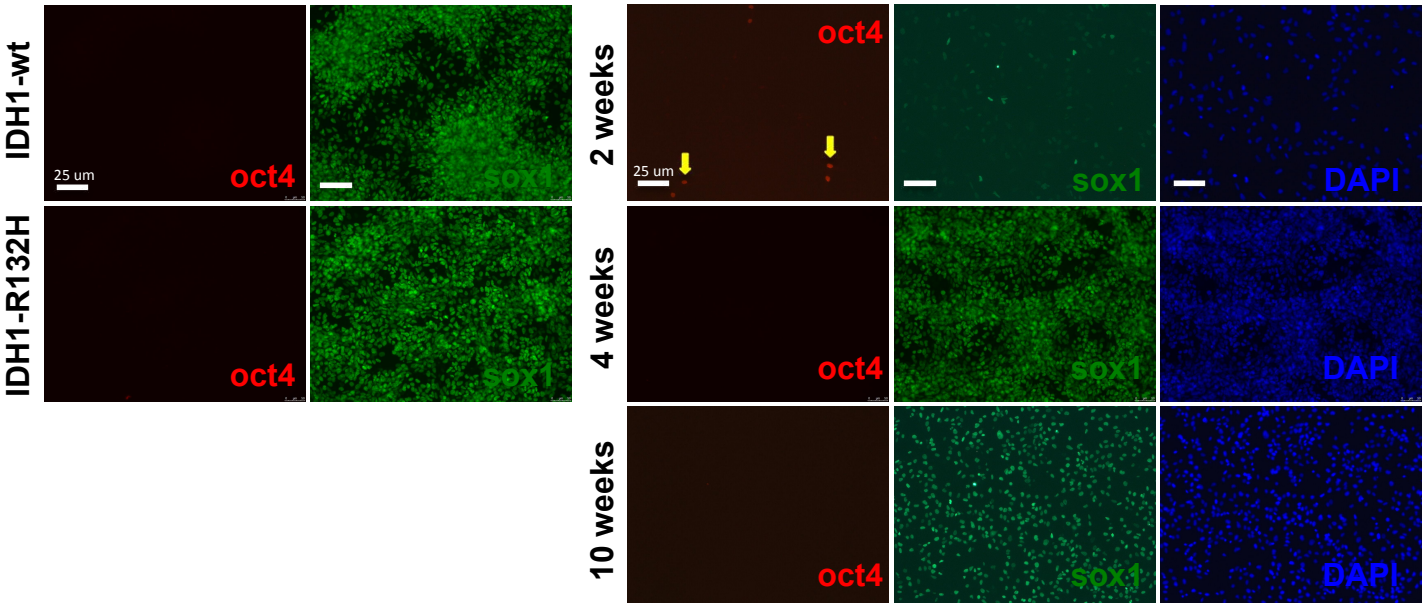
Supplemental Figure 2



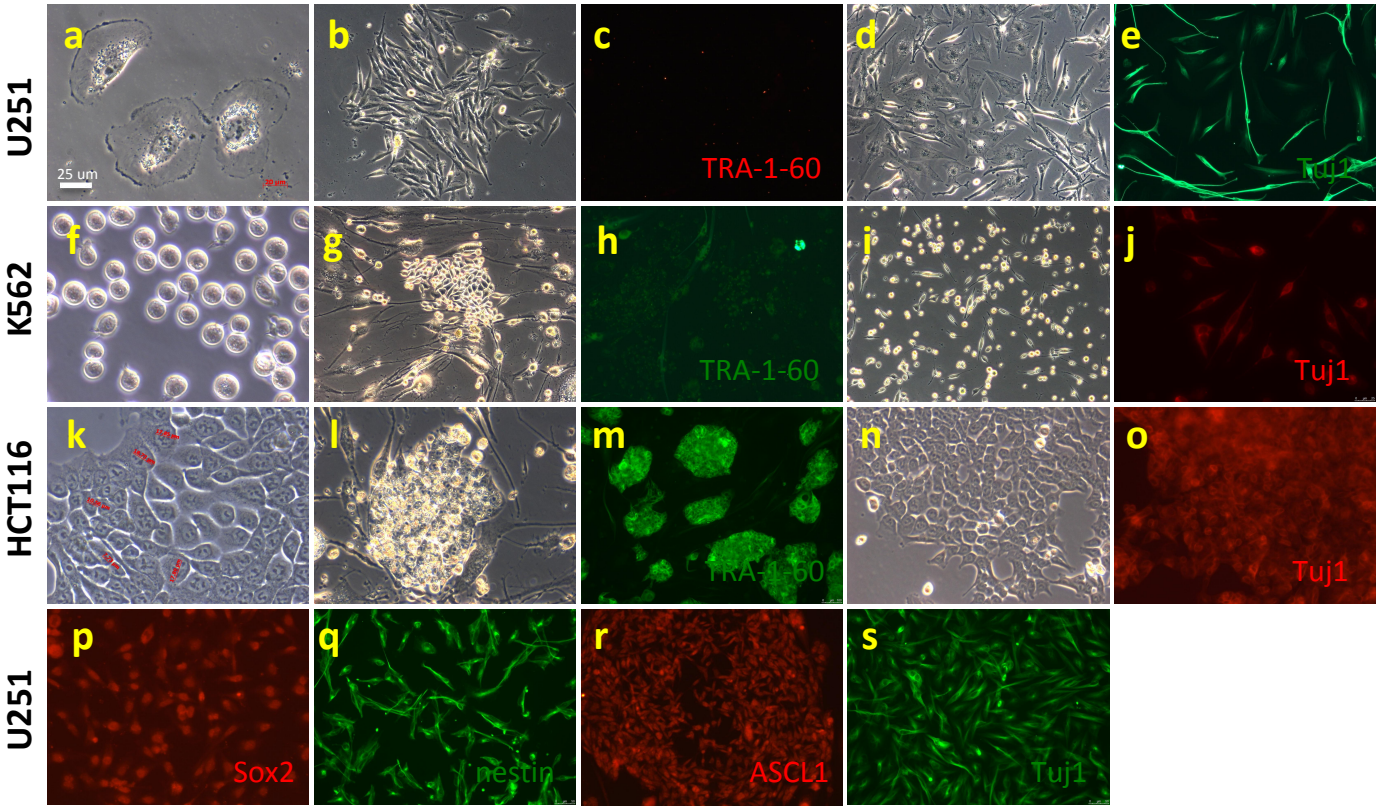
Supplemental Figure 3



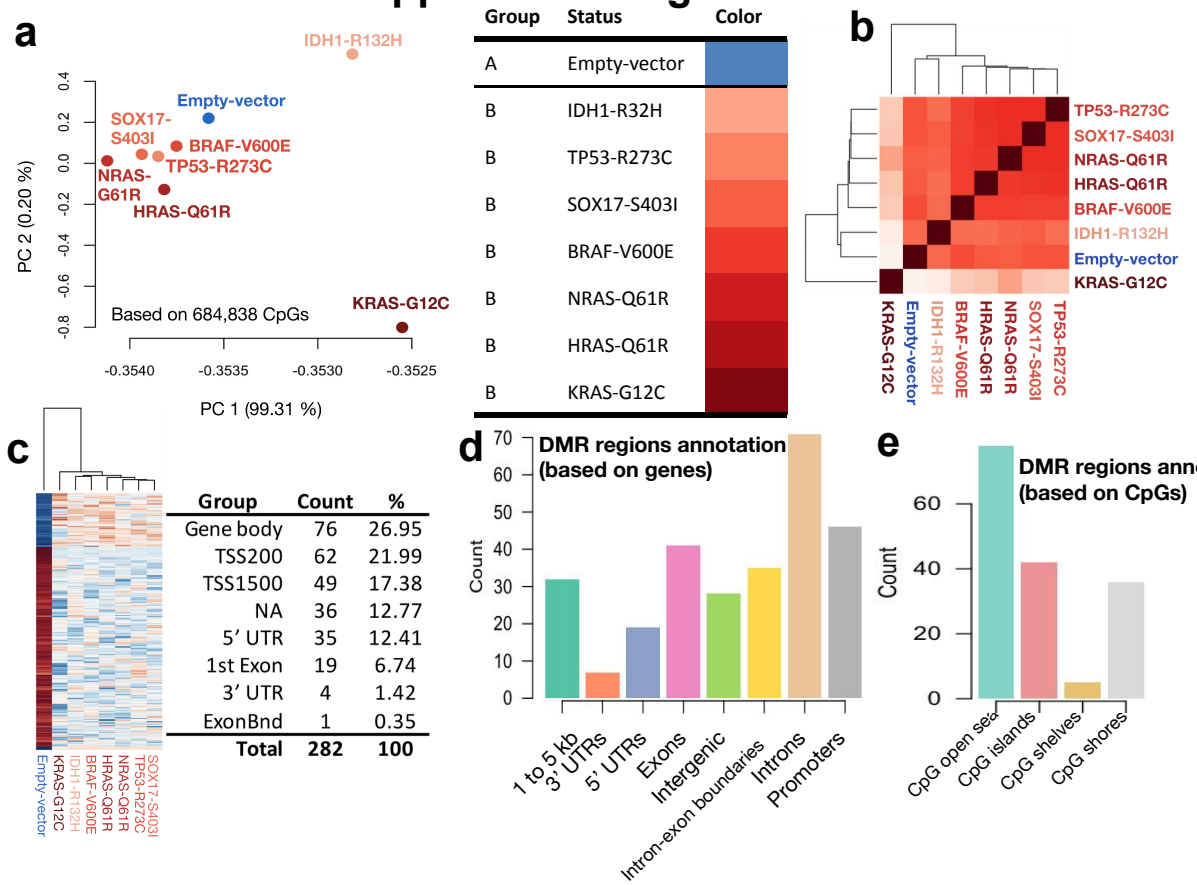




Supplemental Figure 4



Supplemental Figure 5



f Differentially methylated positions (DMPs): enrichment analysis

Category	Term	Count	*P*	Fold Enrichment	adj. *P*
KEGG_PATHWAY	Transcriptional misregulation in cancer	7	0.03	3.07	0.98
KEGG_PATHWAY	Pathways in cancer	11	0.04	2.05	1.00
KEGG_PATHWAY	Regulation of actin cytoskeleton	7	0.06	2.44	1.00

Category	Term	Count	*P*	Fold Enrichment	Bonferroni
GO:MF	transcription factor activity, sequence-specific DNA binding	22	0.00398	1.96	0.81
GO:MF	transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	9	0.00641	3.27	0.93
GO:MF	sequence-specific DNA binding	14	0.0076	2.32	0.96
GO:MF	zinc ion binding	24	0.0091	1.76	0.98
GO:MF	protein binding	118	0.01824	1.15	1.00
GO:MF	DNA binding	29	0.03095	1.48	1.00
GO:MF	metal ion binding	34	0.0358	1.41	1.00
GO:MF	ligase activity	8	0.03814	2.54	1.00
GO:MF	chemoattractant activity	3	0.03893	9.52	1.00
GO:MF	ATP binding	26	0.04065	1.49	1.00
GO:MF	histone acetyltransferase binding	3	0.04161	9.18	1.00

g Differentially methylated regions (DMRs): enrichment analysis

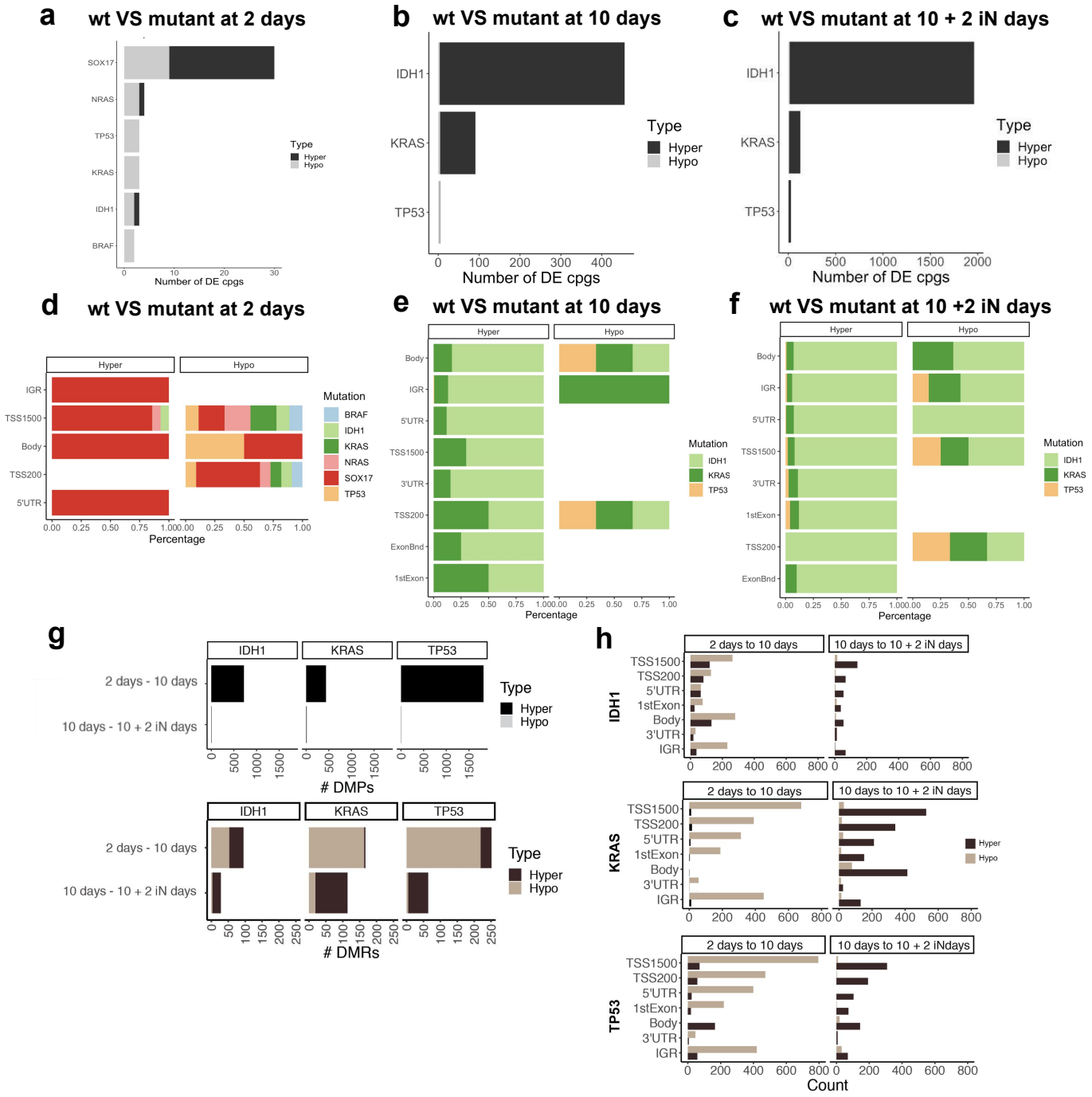
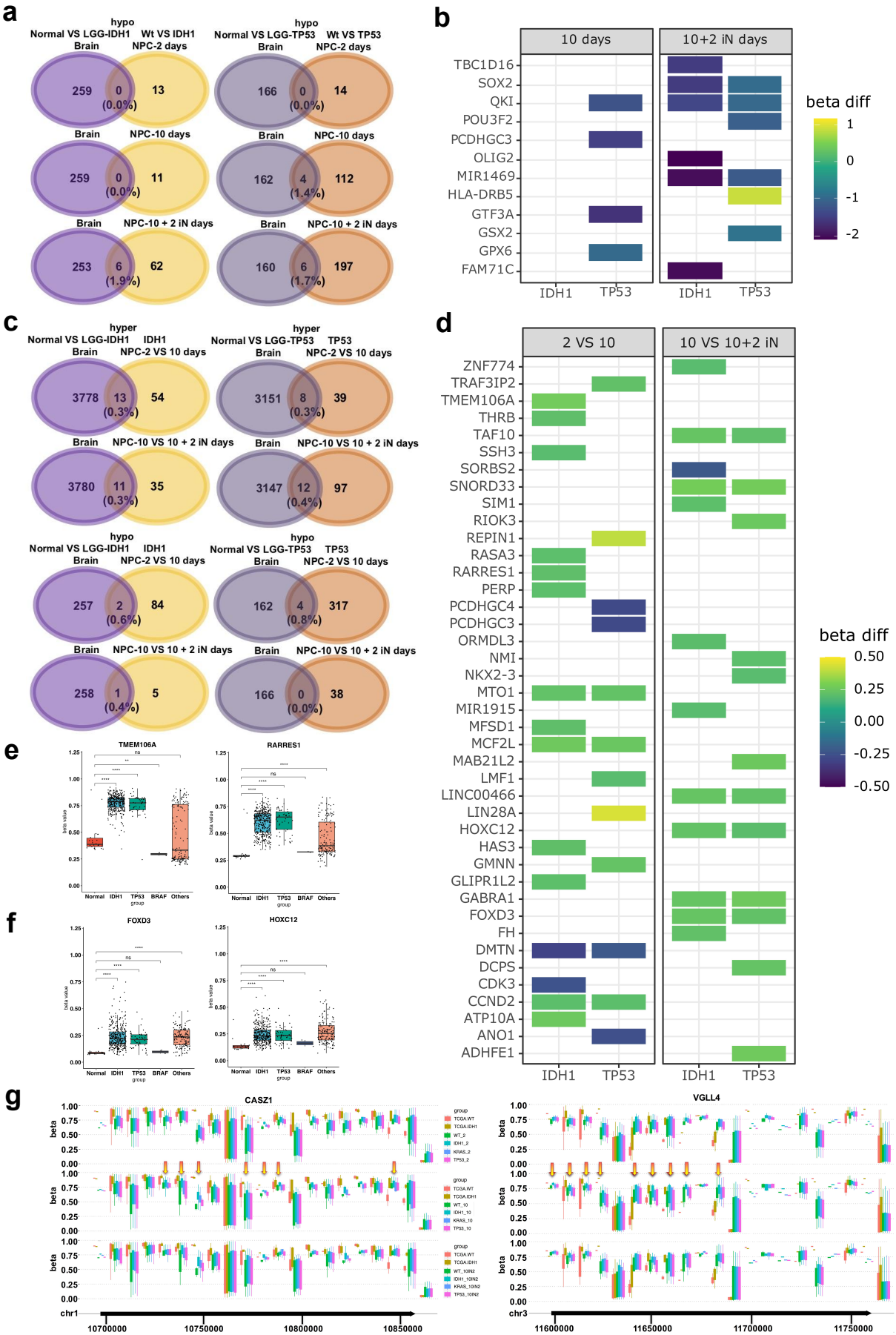
Category	Term	Count	*P*	Fold Enrichment	adj. *P*
GO:BP	negative regulation of canonical Wnt signaling pathway	7	0.00	7.67	0.45
GO:BP	male gonad development	4	0.02	7.60	21.30
GO:BP	intestinal epithelial structure maintenance	2	0.02	119.09	22.75

Category	Term	Count	*P*	Fold Enrichment	adj. *P*
GO:CC	cell junction	12	0.00	4.91	0.04
GO:CC	synapse	6	0.00	6.23	3.18
GO:CC	nuclear membrane	6	0.01	4.92	8.37

Category	Term	Count	*P*	Fold Enrichment	adj. *P*
GO:MF	phosphorylase kinase regulator activity	2	0.02	91.74	23.97
GO:MF	protein domain specific binding	5	0.03	4.41	28.45
GO:MF	palmitoyl-(protein) hydrolase activity	2	0.03	61.16	33.71

Category	Term	Count	*P*	Fold Enrichment	FDR
KEGG_PATHWAY	Long-term potentiation	4	0.01	10.17	7.36

Supplement Figure 6



Supplement Figure 7