Oct4 redox sensitivity potentiates reprogramming and differentiation

Zuolian Shen[1,2,¶], Yifan Wu[1,2,¶], Asit Manna[1,2], Kimberly J. Evason[1,2], Mahesh B. Chandrasekharan[2,3] and Dean Tantin[1,2,*]

[1]Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT 84112, USA

[2]Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT 84112, USA

[3]Department of Radiation Oncology, University of Utah School of Medicine, Salt Lake City, UT 84112, USA

[*]Correspondence: dean.tantin@path.utah.edu.

[¶]Co-first authors

Running head: Oct4 oxidative regulation in reprogramming

Words in abstract: 208

Abstract

The transcription factor Oct4/Pou5f1 is a key component of the regulatory circuitry governing pluripotency. Oct4 is also widely used to generate induced pluripotent stem cells (iPSCs) from somatic cells. These observations provide compelling rationale to understand Oct4's functions. Here we used domain swapping and mutagenesis to compare Oct4's reprogramming activity with the paralog Oct1/Pou2f1, identifying a redox-sensitive DNA binding domain cysteine residue (Cys48) as a key determinant of both reprogramming and differentiation. In combination with the Oct4 N-terminus, mutating Oct1's serine at this position to cysteine is sufficient to confer strong reprogramming activity. Conversely, Oct4^{C48S} strongly reduces reprogramming potential. Oct4^{C48S} renders the protein insensitive to oxidative inhibition of DNA binding. The same mutation prevents oxidation-mediated protein ubiquitylation. An engineered *Pou5f1*^{C48S} point mutation has little effect on undifferentiated cells embryonic stem cells (ESCs), but upon retinoic acid (RA)-mediated differentiation causes retention of Oct4 expression, decreased proliferation, and increased apoptosis. Transcriptional profiling reveals that the retention of pluripotency also manifests at the RNA level. *Pou5f1*^{C48S} ESCs also contribute poorly to adult somatic tissues and form less differentiated teratomas. Collectively, the data support a model in which Oct4 redox sensing is a positive reprogramming determinant during one or more reprogramming steps, and mediates Oct4 degradation during differentiation.

Keywords Oct4 (Pou5f1); Oct1 (Pou2f1); induced pluripotent stem cells (iPSCs); oxidative stress; ubiquitylation

41 **Introduction**

42 The early mammalian embryo contains undifferentiated, pluripotent cells capable of generating all
43 embryonic tissue lineages. Cultured embryonic stem cells (ESCs) are derived from the early embryo and
44 have similar capabilities[1]. The transcription factor Oct4/Pou5f1 governs the establishment of pluripotent,
45 undifferentiated inner cell mass and epiblast cells, as well as primordial germ cells, germline progenitor
46 cells and ESCs[2]. Both ESC differentiation in culture and blastocyst implantation in animals are
47 accompanied by Oct4 loss[3]. Together with transcription factors such as Nanog and Sox2, Oct4 forms a
48 network of pluripotency master transcriptional regulators that bind to their own and each other's genes to
49 reinforce transcription[4].

50 Oct4 is also widely used to generate induced pluripotent stem cells (iPSCs) from somatic cells[5, 6].
51 Reprogramming methods that omit Oct4 nevertheless rely on activation of the endogenous *Pou5f1* (*Oct4*)
52 gene[7, 8]. These findings pinpoint Oct4 as a central reprogramming component and place priority on
53 understanding its functions. Reprogramming strategies that increase efficiency[9, 10, 11, 12, 13], or improve the
54 quality of iPSC clones[14, 15, 16, 17] reveal difficult trade-offs: efficiently generating iPSC that are of non-
55 optimal quality, or inefficiently generating iPSCs with improved developmental potential.

56 Pluripotent cells co-express Oct4 together with the widely expressed prototypic POU transcription
57 factor Oct1/Pou2f1[18, 19, 20, 21, 22]. The two proteins share similar DNA binding domains and recognize
58 similar 8-bp 5'ATGCAAAT3' consensus octamer sequences and sequence variants[23]. The two proteins
59 also associate with many of the same target genes[22, 24]. Nevertheless, Oct1 as well as other mammalian
60 Oct4 paralogs do not efficiently reprogram[5, 25]. Oct1 in particular is poor at reprogramming relative to
61 most other Oct transcription factors[26]. In contrast, other reprogramming factors such as Sox2, Klf4 and c-
62 Myc can be replaced by paralogs[27, 28]. Structure/function studies comparing Oct4 with its paralogs have
63 identified Oct4 transcriptional determinants[29, 30, 31]. An example is three residues present in the Oct4 DNA
64 binding domain but absent in Oct6/Pou3f1 needed to form Sox2 heterodimers. Introducing these residues,
65 plus the Oct4 N- and C-termini, results in efficient reprogramming[30]. However, several Oct4 paralogs

66 including Oct1 contain these residues but are nevertheless reprogramming-incompetent, indicating the
67 presence of additional reprogramming determinants present in Oct4 (and likely Oct6) but absent in Oct1.
68 Recent studies show that omission of Oct4 from reprogramming cocktails necessitates using target
69 cells of high mitotic potential and higher-than-typical levels of the other factors, and is still inefficient,
70 but results in qualitatively superior iPSC clones with enhanced developmental potential[17, 32]. The
71 counterproductive activities of Oct4 are caused by transient promotion gene expression programs that
72 deviate cells from reprogramming trajectories[17]. A logical extension of these findings is that transient
73 down-regulation of counterproductive Oct4 activities may be an important step in standard
74 reprogramming.
75 Here, we identify a new central Oct4 reprogramming determinant – a single DNA binding domain
76 cysteine residue (Cys48) – that accounts for much of its differential reprogramming potential vis-à-vis
77 Oct1. A serine residue is present in Oct1 at this position. In the presence of the Oct4 N-terminus, changing
78 the Oct1 serine to cysteine at this position is sufficient to confer levels of reprogramming activity
79 comparable to wild-type Oct4. Conversely, mutating this cysteine to serine in the context of Oct4 reduces
80 reprogramming efficiency by ~60%. Mechanistically, we find that this cysteine sensitizes DNA binding
81 activity to oxidative stress, and mediates Oct4 protein ubiquitylation via a redox regulation mechanism.
82 We generated CRISPR-edited Oct4^{C48S} embryonic stem cell (ESC) lines, which are phenotypically normal
83 until differentiated. Upon differentiation, edited ESC lines abnormally retain Oct4 expression, proliferate
84 less and undergo apoptosis at higher rates compared to unedited parental controls. These cells also form
85 less differentiated teratomas and contribute poorly to the development of viable adult mice. Cumulatively,
86 our findings identify a redox-sensitive Oct4 residue that restrains protein function, but nevertheless
87 functions positively in reprogramming.

88 **Results**

89 **Oct4 vs. Oct1 differential reprogramming potential localizes to the DNA binding domain**

90 The POU-specific and POU-homeodomain (POU_S and POU_H) DNA binding subdomains of Oct1 and
91 Oct4 are similar, with 71% and 59% identity in mouse, respectively. In contrast, the N- and C-termini,
92 and the linker between the subdomains, are highly divergent (Fig. 1a and not shown). The human proteins
93 are structured similarly. Because the DNA binding domains are similar, the amino acid numbering can be
94 standardized by referring to the first amino acid of the POU_S DNA binding domain as residue 1 (Fig. 1a).
95 To identify Oct4 determinants important for reprogramming, we generated Oct4/Oct1 chimeric constructs
96 containing the distinct domains (Fig. 1b). Because introducing different domains and epitopes can
97 confound the analyses with protein-specific antibodies, we included C-terminal FLAG-tags on Oct4, Oct1
98 and all chimeric proteins to track protein expression. Unlike an N-terminal FLAG tag which was
99 deleterious (Supplementary Fig. 1), engineering a C-terminal tag results in reprogramming efficiency
100 similar to untagged Oct4 (Fig. 1c). Prior work comparing Oct4 and another reprogramming-incompetent
101 POU transcription factor, Oct6, also identified a lysine residue within the POU_S DNA binding subdomain
102 (position 40 of the Oct4 DNA binding domain, Fig. 1a) as important for reprogramming[31]. We therefore
103 mutated N40 of the Oct1 DNA binding domain to lysine in some constructs (Fig. 1d-f).

104 To measure reprogramming efficiency, we transduced primary mouse embryonic fibroblasts
105 (MEFs) expressing GFP under the control of the endogenous *Pou5f1* gene[33] with mSTEMCCA
106 lentiviruses express the four Yamanaka factors[34] (Fig. 1b). Substituting Oct1 for Oct4 eliminated all
107 reprogramming potential in this assay (Fig. 1d,e). The other reprogramming factors (Sox2, Klf4 and c-
108 Myc) can mediate reprogramming without Oct4[17, 32], however reprogramming is inefficient. The results
109 are therefore consistent with the original reprogramming findings showing that Oct4 is essential for
110 conventional reprogramming[5], and set a useful baseline for making domain-swaps and mutations to assess
111 reprogramming potency.

112 We then replaced the Oct1 N- and C-termini with that of Oct4, and mutated asparagine 40 of the
113 Oct1 DNA binding domain to the lysine present in Oct4. Prior work[31] showed that the Oct4 lysine at this
114 position (Fig. 1a) is important for Oct4 reprogramming. Of these constructs, only fusion of the Oct4 N-
115 terminus generated detectable colony-forming potential, at ~5% the level of WT Oct4 (Fig. 1d,e). These
116 findings are consistent with work establishing the importance of the Oct4 N-terminus[35]. Similar data were
117 obtained at multiple timepoints (not shown), indicating that this result does not simply reflect kinetic
118 differences in reprogramming. Anti-FLAG immunoblotting showed equivalent expression (Fig. 1f). Viral
119 titers and expression of Klf4 from the mSTEMCCA vector were also similar (not shown). Because a
120 construct with both the Oct4 N- and C-termini displayed only ~5% efficiency (Fig. 1d,e), we conclude
121 that determinants within the Oct4 DNA-binding domain confer the bulk of reprogramming potential.
122
123 **Oct4 vs. Oct1 differential reprogramming potential concentrates in a DNA binding domain cysteine**
124 The POU DNA binding domain comprises two subdomains connected by a peptide linker. The two
125 subdomains make separate sequence-specific contacts with DNA[23]. These domains are termed POU_S and
126 POU_H (Fig. 1a,b). We mutagenized DNA binding domain amino acid residues that differ between Oct1
127 and Oct4. We focused on possible Oct4 post-translationally modifiable residues: Lys, Ser, Thr and Cys.
128 This analysis identified 14 possible amino acid substitutions, not including the aforementioned Lys40
129 (Fig. 1a and not shown). Work from others has shown that Oct6 is capable of reprogramming when
130 modified to enable Sox2 dimerization[36]. Filtering the data using conservation between Oct4 and Oct6
131 resulted in just six residues. Replacing one of these, an Oct1 serine at position 48 with the Oct4 cysteine
132 residue found at this position, and using the Oct4 N-terminus, conferred the majority of the differential
133 Oct4 activity, increasing activity >7-fold over the "4N" construct containing the Oct4 N-terminus alone
134 (Fig. 2a,b). We term this the "1-4All" construct. iPSC colonies formed under these conditions showed a
135 pluripotent morphology and expressed GFP from the endogenous *Pou5f1* locus at levels comparable to
136 cells reprogrammed with Oct4 (Supplementary Fig. 2). The remaining activity relative to Oct1 appeared

137 to be concentrated in the linker domain, as a construct containing the Oct4 N-terminus, Cys48 and the
138 Oct4 linker (4N-S48C-4L) had 100% activity (Fig. 2a,b). The Oct4 linker has been shown to regulate
139 reprogramming via selective interactions with different cofactors[29, 37]. The differences were maintained
140 over time, with the 1-4All construct consistently generating ~60% activity over multiple days, indicating
141 that the differences were qualitative and not simply a function of different colony formation rates
142 (Supplementary Fig. 3).

143 Replacing the Oct4 Cys with Ser at this position reduced activity by 60% (Fig. 2a,b), indicating
144 that more than half of Oct4's differential reprogramming activity vis-à-vis Oct1 is lost by changing this
145 residue. The converse experiment, replacing Oct1 S48 with cysteine, resulted in no reprogramming
146 activity (Fig. 2a,b, Oct1-S48C), indicating that Cys48 is important but not sufficient, and highlighting its
147 combined importance with the Oct4 N-terminus. The chimeric proteins were equivalently expressed (Fig.
148 2c). These results indicate that in the context of the Oct4 N-terminus, a cysteine residue at position 48
149 confers robust reprogramming activity. In the co-crystal structure of the Oct1 DNA binding domain bound
150 to DNA[38], Ser48 hydrogen bonds with the DNA backbone (Fig. 2d). In a recent Oct4:Sox2:nucleosome
151 structure[39], the Cys48 sulfhydryl also allows for hydrogen bonding. Predictably, mutation of Cys48 to
152 glycine greatly diminishes Oct4 DNA binding in vitro[40]. Oct6, which can reprogram when changes are
153 made to POU_H allowing for Sox2 dimerization[30], contains a cysteine at this position (Fig. 2e). These
154 structural studies along with our findings suggest that Cys48 has minimal effects on DNA binding vis-à-
155 vis a serine (but not glycine) at this position, but nevertheless plays a role in reprogramming.

156

157 **Oct4 Cys48 is redox-sensitive and mediates degradation in response to oxidative stress**

158 Oct4 is known to be sensitive to oxidative stress[40, 41, 42] in a manner reversed by thioredoxin[42], implicating
159 cysteine thiols in redox sensitivity. We reproduced this result (Supplementary Fig. 4), which with the
160 results above suggested that Cys48 may potentiate reprogramming through a mechanism involving redox
161 sensitivity. We measured wild-type and C48S mutant Oct4 DNA binding using electrophoretic mobility

162 shift assays (EMSA) with Cy5-labeled DNA containing a canonical 5'ATGCAAAT octamer element[24].
163 To measure redox sensitivity, we used methods similar to Guo et al.[42] in which Oct4 is pre-treated with
164 the oxidizing agent diamide prior to addition of the DNA probe.

165 We first used lysates from 293T cells transfected with WT or C48S mutant Oct4. 293T cells do
166 not express Oct4, but do express endogenous Oct1 as an internal control (Fig. 3a, lanes 2-3). Oct4 bound
167 to octamer DNA specifically (Fig. 3a, lanes 7-8) and was hypersensitive to diamide treatment (lanes 9-
168 11). $Oct4^{C48S}$ bound DNA equivalently (lane 13) and also conferred resistance to oxidation, in particular
169 at higher concentrations (lanes 14-16). In contrast, little sensitivity was observed with Oct1. Quantification
170 from three independent replicate experiments is shown in Fig. 3b. The recombinant Oct4 contained a C-
171 terminal FLAG tag, allowing for purification (Supplementary Fig. 5). Similar relative resistance to
172 oxidative inhibition of DNA binding was observed using purified C48S recombinant Oct4 protein (Fig.
173 3c). Quantification from three independent replicates is shown in Fig. 3d. These results indicate that Oct4
174 cysteine 48 sensitizes Oct4 to oxidative inhibition of DNA binding in vitro.

175 Oct4 oxidation has been linked to its degradation[40]. In studying recombinant Oct4 expressed in
176 293T cells, we noted in over-exposed immunoblots the presence of multiple slow-migrating high
177 molecular weight forms of wild-type Oct4. These high-molecular weight bands were strongly attenuated
178 with C48S mutant protein (Fig. 3e, lanes 2-3). The bands were also increased by hydrogen peroxide
179 treatment using wild-type but not mutant protein (lanes 4-5), indicating that they are increased by oxidative
180 stress in a manner that requires Cys48. They were also specifically increased using the proteasome
181 inhibitor MG-132, strongly suggesting that the high-molecular weight forms are ubiquitylated species
182 (lanes 6-7). Interestingly, the combination of both H_2O_2 and MG-132 increased the intensity of the banding
183 pattern over that observed using either treatment individually (lanes 8-9). These treatments alone and in
184 combination did not affect short-term viability of the cells (Supplementary Fig. 6).

185 To identify the nature of the ubiquitylated Oct4 species, we purified Oct4 expressed from 293T
186 cells using a C-terminal streptavidin tag, and subjected the isolated proteins to immunoblotting using

187 antibodies specific to different ubiquitin linked chains. Analysis using antibodies to pan linked Ub chains
188 revealed strong diminution with Oct4^{C48S} as expected (Fig. 3f, FK2). Strong diminution was observed
189 with K48 linkage-specific polyubiquitin antibodies (Fig. 3f), while K63 linkage-specific antibodies
190 showed more modest differences compared to total Oct4 high-molecular weight bands (Supplementary
191 Fig. 7). Similar results were observed using endogenous Oct4 in wild-type ESCs, where high-molecular
192 weight Oct4 bands could be detected (Fig. 3g, lane 2) that were augmented by H_2O_2 and MG-132 (lanes
193 4 and 6). We conclude that Cys48 promotes Oct4 ubiquitylation and degradation in response to oxidative
194 stress, largely via catalysis of K48-linked polyubiquitin chains.
195
196 **ESCs with endogenous Oct4^{C48S} mutation differentiate abnormally**
197 The ability of Oct4 Cys48 to promote both pluripotency and redox sensitivity suggested that sensitivity to
198 oxidative stress mediated by Cys48 helps potentiate one or more reprogramming steps. To test if Oct4
199 oxidative stress sensitivity also regulates forward developmental steps, we studied the role of Cys48 in
200 ESC differentiation. Using CRISPR, we generated ESCs with a single-base missense mutation that
201 converts Cys48 to Ser (see Methods). Two silent point mutations were also engineered in the gRNA that
202 create an *Xba*I site to monitor targeting and facilitate genotyping. Two low-passage parent ESC lines were
203 chosen for targeting, one on a pure C57BL/6 background and one on a mixed BL6/129 background.
204 Multiple independent clones were generated for each line. Homozygous mutation was confirmed by
205 HRMA (not shown) and resequencing (Fig. 4a). Tested clones were confirmed to be karyotypically normal
206 and wild-type for *Tp53* (not shown). Undifferentiated derived ESC clones were phenotypically
207 indistinguishable to the parent clones in terms of Oct4, Sox2 and Nanog expression and morphology (Fig.
208 4b,c), as well as long-term maintenance of pluripotency factor expression and proliferative capacity (not
209 shown).
210 To test if C48S mutation protects endogenous Oct4 against oxidative stress, we studied Oct4
211 ubiquitination in parent and mutant cells with and without H_2O_2 treatment. Oct4 was immunoprecipitated

212 from ESC lysates, and the resulting material was immunoblotted using pan-Ub (FK2) and Oct4 antibodies.
213 Ubiquitylation of endogenous Oct4 in ESCs was clearly visualized and was increased with H_2O_2 exposure
214 (Fig. 4d, lanes 1-4). Ubiquitylation was attenuated in mutant cells (lanes 5-6), and further was not affected
215 by oxidative stress (lanes 7-8). Oct4 immunoblotting confirmed that the protein was immunoprecipitated,
216 with decreased parent Oct4 recovered with oxidative stress (presumably due to degradation) and more
217 Oct4 recovered in mutant cells.

218 Parent and derived mutant cells were then differentiated in vitro using retinoic acid (RA) to assess
219 Oct4 expression, morphology and cell counts. Oct4 protein expression was abnormally retained early
220 during differentiation, mostly at day 2 (Fig. 4e, lanes 3 and 7). Quantification from multiple experiments
221 revealed that the effect was transient, with a >2.8-fold increase of Oct4 at day 3 that was lost by day 4
222 (Fig. 4f). These results were recapitulated using a second clone from the same parental ESC line
223 (Supplementary Fig. 8a,b), as well as a C48S clone from a different 129×C57BL/6 ESC line, G4-N, which
224 differentiates more slowly and showed even greater differences (Supplementary Fig. 8c).

225 Because Oct4 is a transcription regulator whose expression is retained during differentiation by
226 C48S mutation, we performed transcriptional profiling of parent and cells at RA differentiation day 2, as
227 well as undifferentiated controls. Three replicates per condition were used, which showed tight
228 concordance (Supplementary Fig. 9a). Between 20 and 31 million sequence reads were generated for each
229 sample, ~80% of which aligned perfectly to the *mm39* reference genome. We identified 421 genes with
230 significantly elevated gene expression in differentiating mutant relative to parent cells (p<0.001, >2.5-fold
231 increased, Fig. 4h, Supplementary Table 1). Intersecting these genes with the set of genes normally
232 decreased upon differentiation of parent cells (1536 genes) showed that 40% of the 421 genes with retained
233 expression (169 genes) are normally down-regulated with differentiation. These included *Oct4* itself
234 (*Pou5f1*), as well as *Sox2*, *Nanog* and *Klf4* (Fig. 4h). Comparing these genes with established Oct4 ChIP-
235 seq [43] identified 114 genes (67%) as direct Oct4 targets. The complete set of 169 genes were enriched for
236 gene ontology (GO) terms associated with pluripotency such as "response to leukemia inhibitory factor"

237 and "stem cell population maintenance" (Fig. 4g). Similar results were obtained using a hierarchical
238 clustering approach. Because clustering all differentially expressed genes did not reveal obvious trends
239 (not shown), we utilized just the 100 top differentially expressed genes upon differentiation for both
240 genotypes (150 total genes in the union of the two sets). Although the differentiating parental and mutant
241 replicates clustered together, there was a cohort of pluripotency associated genes (including *Pou5f1*) that
242 shared less down-regulation in mutant cells (Supplementary Fig. 9b). A second cohort of differentiation-
243 specific genes showed poorer expression in mutant cells upon differentiation. These included *Krt18*,
244 *Mest*/*Peg1* and *Dag1* (Supplementary Fig. 9b). Comparing the complete set of 150 genes to Oct4 ChIP-
245 seq [43] 117 genes (78%) as direct Oct4 targets. These results indicate that Oct4^{C48S} mutant ESCs maintains
246 aspects of pluripotency gene expression during differentiation.

247 The duration of retained Oct4 expression during differentiation can affect lineage specification
248 outcomes[44, 45, 46]. To determine if Oct4^{C48S} mutation affects differentiation, we treated mutant and parental
249 control cells for 12 days with RA. Longer RA-mediated differentiation revealed a significant depletion of
250 cells later in the differentiation timecourse (Fig. 5a,b). The decreased cell numbers observed during
251 differentiation of C48S-edited cells with RA could be due to decreased proliferation, increased death or
252 some combination of the two. To investigate the nature of the effect on differentiation more closely, we
253 studied proliferation using cell trace violet and apoptosis using TUNEL staining. Adherent, viable
254 differentiating C48S cells showed decreased proliferation at days 3 and 4 (Fig. 5c) and increased apoptosis
255 at day 2 (Fig. 5d). These results indicate that C48S-edited ESCs differentiate abnormally.

256

257 **Oct4^{C48S} mutant ESCs form less differentiated teratomas and contribute poorly to embryonic**
258 **development**

259 We used these same ESC lines to form teratomas in immunocompromised mice. Parent and derived
260 Oct4^{C48S} mutant clones were injected into contralateral flanks of NOD.*Rag1*$^{-/-}$*Il2rg*$^{-/-}$ (NRG) female mice,
261 which were monitored for two weeks before tumor excision and weighing. Tumors were fixed, H&E

262 stained, and pathologically scored in a blinded fashion. Parent tumors contained areas of maturing
263 structures such as squamous tissue, respiratory epithelium (cilia and/or goblet cells), maturing neural
264 components, gastric foveolar epithelium and fat. In contrast, tumors formed with Oct4^{C48S} mutant ESCs
265 did not contain maturing components, being entirely comprised of immature neuroepithelial elements,
266 immature cartilage, embryonal carcinoma and yolk sac. with greater homogeneity and fewer differentiated
267 structures (Fig. 6a). Representative images are shown in Fig. 6b. Additional images are shown in
268 Supplementary Fig. 10. At later time points, larger mutant teratomas did begin to accumulate mature
269 structures, but they remained fewer than parent controls (not shown).

270 One essential hallmark of pluripotency is the ability to contribute to cells and tissues of the
271 embryo[47]. We conducted chimerism assays to assess the ability of CRISPR-edited ESCs to contribute to
272 development. We used nine lines of mutant ESCs, derived from three independent wild-type parent lines
273 with different strain backgrounds to minimize background and clonal effects. Derived clones had
274 karyotypes similar to parent clones, with the proportion of diploid cells ranging between 71 and 93%.
275 Compared to the parent cells, which contributed to offspring at a rate of 10-20% (Fig. 6c, chimerism
276 estimated to range from 2% to 99%), none of the derived Oct4^{C48S} mutant ESCs were able to contribute
277 to adult mice (Fig. 6c). The results indicates that Oct4^{C48S} ESCs do not contribute to embryonic
278 development.

279 **Discussion**

280 Oct4 is a central mediator of pluripotency and is widely used to generate iPSCs from somatic cells[25, 27, 48,
281 49]. Together with other factors it sustains pluripotency by activating "core" targets such as *Pou5f1* (which
282 encodes Oct4), *Sox2* and *Nanog*[4]. Oct4 also maintains silent targets, including critical developmental
283 regulators, in a readily inducible "poised" state[50, 51]. Yet the mechanisms by which Oct4 is itself regulated,
284 particularly at a post-transcriptional level, are poorly defined. Here, we used Oct4/Oct1 domain-swapping
285 to pinpoint Oct4 redox sensitivity as an important determinant in reprogramming and differentiation. Oct1
286 itself appears to have no role in pluripotency establishment or maintenance, but upon differentiation and
287 Oct4 loss becomes important for guiding developmental-mediated induction of targets shared with Oct4[22].

288 Pluripotent cells co-express Oct4 together with the widely expressed prototypic Oct/POU factor
289 Oct1/Pou2f1[21]. Despite having similar DNA binding domains and recognizing the same 8-bp consensus
290 octamer sequence (5'ATGCAAAT3')[23, 52], Oct1 and other Oct4 paralogs do not function in
291 reprogramming[5]. Structure/function studies comparing Oct4 with these paralogs, in particular Oct6, have
292 identified Oct4 transcriptional determinants[29, 30, 31, 35]. An example is three residues (K7,T22,151S) present
293 in the Oct4 DNA binding domain (Fig. 1a) but absent in Oct6 needed to form Sox2 heterodimers.
294 Introducing these residues plus the Oct4 N- and C-termini into Oct6 results in efficient reprogramming[30].
295 However, several Oct4 paralogs including Oct1 contain these residues and readily heterodimerize with
296 Sox2 but cannot reprogram, indicating that additional reprogramming determinants may be present in
297 Oct4 (and Oct6) but absent in Oct1. We identify an additional central Oct4 reprogramming determinant –
298 a single DNA binding domain cysteine residue that confers redox sensitivity of Oct4.

299 We found that Oct4 C48S mutation maintains DNA binding and confers resistance of oxidation-
300 induced DNA binding inhibition in vitro, while reducing reprogramming efficiency by more than half.
301 Cys48 is located within helix 3 (the DNA recognition helix) of the POU-specific domain (Fig. 1a). This
302 residue makes hydrogen bond contacts with the DNA phosphate backbone, predicting that any oxidation
303 event will block DNA binding. In the context of the Oct4 N-terminus, a single S48C mutation in the Oct1

304 DNA binding domain is sufficient to confer reprogramming activity. This result suggests that any unique
305 Oct4 cofactors required for reprogramming vis-à-vis Oct1 are likely to be recruited through the N-
306 terminus. The ability of the Oct4 N-terminus to mediate liquid-liquid phase transitions[35] may also play a
307 role. Mutation of the endogenous Oct4 Cys48 to serine does not appear to affect the undifferentiated state,
308 but results in numerous phenotypes upon RA-mediated differentiation, including abnormal retention of
309 Oct4 expression, decreased proliferation and increased apoptosis. Prominent roles for Oct4 in both cell
310 survival and proliferation during post-implantation embryonic development have been documented[46].

311 The redox sensitivity of Cys thiols is known to be enhanced by immediately adjacent basic amino
312 acid residues[53], and Cys48 is followed by a conserved Arg (Fig. 1a). In Oct1, this residue is a serine,
313 strengthening hydrogen bonding due to the increased electronegativity of the oxygen in the hydroxyl
314 relative to the sulfur in the sulfhydryl. Redox-regulated transcription factors include Nrf2, Yap1, p53 and
315 HIF-1[54, 55, 56]. A subset of these are regulated by Cys oxidation, including disulfide bond formation[55, 56,
316 57]. Oct4 is similarly known to be sensitive to oxidative stress[40, 41, 42] in a manner reversed by thioredoxin[42]
317 (Supplementary Fig. 4). Relative to other Oct proteins, mouse Oct4 has unusually high cysteine content
318 in the DNA binding domain, with two cysteines in each subdomain (Fig. 1a), including cysteine 142
319 (Cys50 of the POU-homeodomain), which also lies in the recognition helix. Cys50 is the distinguishing
320 feature of POU-homeodomains relative to other members of the homeodomain family[23, 52]. All other
321 homeodomains, from bacteria to flies to mammals, have bulkier residues such as Q or K at this position
322 that make base contacts[23]. Yet in Oct4, mutating Cys142 to Gly has no effect on DNA binding[40]. In Oct1,
323 C61A and C142A mutation does not interfere with DNA binding, but rather improves crystallization
324 properties[58]. These findings suggest that variable Cys oxidation interferes with uniform crystal packing.

325 Prior studies have also identified a link between Oct4 oxidation and degradation[40]. In addition to
326 mediating relative sensitivity to oxidative stress, we found that Cys48 increases protein ubiquitylation, in
327 particular following exposure to oxidative stress or MG-132. Oct4 is known to be ubiquitylated via K63-
328 linked chains by the Stub1 E3 Ub ligase to reduce protein half-life[59], however we found that Oct4 marked

329 with K48-specific linkages were more strongly affected with mutation of cysteine 48 to serine, suggesting
330 a possibly related by distinct pathway.

331 Transcription factors with complex (both positive and negative) roles in reprogramming include
332 Oct4 and Smad3[17, 32, 60]. Although endogenous Oct4 must be activated and expressed as a key component
333 of the core pluripotency network, recent studies show that removal of Oct4 from reprogramming cocktails
334 is highly inefficient but results in qualitatively superior iPSC clones that more closely resemble a
335 pluripotent "ground state"[17, 32]. As a consequence, the few iPSC colonies derived in the absence of Oct4
336 have enhanced developmental potential. For example, iPSCs reprogrammed without Oct4 are better able
337 to generate viable mice in a single generation using tetraploid complementation. The counter-productive
338 activity of Oct4 is likely due to transient promotion of off-target, ectoderm-specific gene expression[17].
339 The studies indicate that transient down-regulation of counterproductive Oct4 activities can be an
340 important step for efficient reprogramming. Our work identifying a role of Oct4 redox sensitivity in
341 reprogramming may explain this process and may provide a route to generating superior iPSC
342 reprogramming outcomes with high efficiency.

343 **Methods**

344

345 **Cell culture**

346 ESCs/iPSCs were cultured as described previously[61, 62] with "2iV" supplementation: the MEK inhibitor
347 PD03259010 (1 μM, LC laboratories) and the glycogen synthase kinase-3β (GSK3β) inhibitor
348 CHIR99021 (3 μM, LC Laboratories). Culture medium was also supplemented with 50 μg/mL ascorbic
349 acid (Sigma). H_2O_2 (Sigma) was supplied at 1 mM for 2 hr and MG-132 (Sigma) was supplied at 10 μM
350 for 2 hr. Cell trace violet proliferation assays were conducted using a kit (Cell Trace Violet Cell
351 Proliferation Kit for Flow Cytometry, ThermoFisher) according to the manufacturer instructions. TUNEL
352 assays were conducted using a kit (Apo-Direct, Becton-Dickinson) according to the manufacturer
353 instructions.

354

355 **FLAG-tagged and chimeric Oct4/Oct1 constructs for iPSC generation**

356 The "mSTEMCCA" 4-in-1 lentiviral vector expressing Oct4, Klf4, Sox2 and c-Myc was used for these
357 studies[63]. To identify determinants of reprogramming, the Oct4 cDNA cassette was replaced with c-
358 terminally FLAG-tagged full-length mouse Oct4, mouse Oct1 or chimeric constructs. FLAG-tagged Oct4
359 was generated by overlap PCR using overlapping primers containing the FLAG peptide sequence:
360 DYKDDDDK (Forward: 5'-GATTACAAGGATGACGACGATAAGGGAAGTGGCGTGAAACAGA;
361 Reverse: 5'-CTTATCGTCGTCATCCTTGTAATCGTTTGAATGCATGGGAGAG). The first fragment
362 was amplified using a forward primer containing a *Not*I restriction enzyme site (5'-
363 AATGAAAAAAGCGGCCGCCATGGCTGGACACCT) and overlapping reverse primer containing the
364 FLAG sequence. A second fragment was amplified using the reverse primer containing an *Nde*I restriction
365 site (5'-GGGAATTCCATATGTGTGGCCATATTATCATCGTGT) and overlapping forward primer.
366 Mixed first fragment and second fragment as template and amplified the whole fragment by using forward
367 and reverse primers containing restriction enzymes (*NotI* and *Nde*I). After purification and digestion with

368 *Not*I and *Nde*I , the purified fragment was ligated with the mSTEMCCA vector digested by *Not*I and *Nde*I.
369 A similar strategy was applied to generate other tagged constructs. For point mutation, a site-directed
370 mutagenesis kit (New England Biolabs Q5 Site-Directed Mutagenesis kit) was used following
371 manufacturer instructions. When making fusions between the different linker domains and POU_H, we took
372 care to maintain the "RKRKR" sequence at the N-terminus of Oct4 POU_H, which has been shown to be
373 important for reprogramming[30]. Chimeric constructs between the Oct1 linker and Oct4 POU_H contained
374 the amino acid sequence N-AEGLN/ARKRKR as shown in Fig. 1a.
375
376 **iPSC generation**
377 293T cells were transfected by polycation polyethylenimine (PEI) with mSTEMCCA encoding FLAG-
378 tagged Oct4, Oct1 and different domain-swaps and mutants as well as the packaging plasmids
379 (pMDLg/pRRE and pRSV-Rev) and envelope plasmid(VSV-G). Lentiviral supernatants were filtered and
380 used to transduce passage 3 primary MEFs in which EGFP is expressed from the endogenous *Pou5f1*
381 locus (Oct4-GFP MEFs, Jackson Labs strain #008214). Cells were cultured in ESC medium (above).
382 Unless stated otherwise, green iPSC colony emergence was quantified after 12 days using an Olympus
383 IX-71 epifluorescence microscope.
384
385 **Immunoblotting**
386 Antibodies for immunoblotting were as follows: Oct4, Santa Cruz sc-5279; Oct1, Bethyl A301-
387 716A+A301-171A; Nanog, GeneTex GTX100863; Sox2, GeneTex GTX101507; GAPDH, EMD
388 Millipore MAB374; α-Tubulin, Santa Cruz sc-5286; β-actin, Santa Cruz sc-47778.
389
390 **Lentiviral Oct1 complementation**
391 The *Oct1* (*Pou2f1*) cDNA and IRES (internal ribosomal entry site) elements were amplified and cloned
392 together by overlap PCR. In the first PCR, primers to the 5' end of *Oct1* containing a *Not*I restriction site

393 and to the 3' end of *Oct1* that contained a 5' extension of IRES-complementary DNA were used. The
394 sequences were: Oct1-NotI-For: 5'-AATGAAAAAAGCGGCCGCCATGAATAATCCATCAGAAAC-
395 3'; Oct1-Rev-IRES: 5'-TTAGGGGGGGGGGAGGGATCTTCACTGTGCCTTGGAG-3'. In the second
396 PCR, an IRES sequence was amplified using primers to the 5' end of the IRES containing a 5' extension
397 of DNA complementary of the *Oct1* 3' end, and primers to the 3' end of the IRES containing an *Nde*I
398 restriction site. The sequences were: IRES-overlap-FOR: 5'
399 AGATCCCTCCCCCCCCCCTAACGTTACTGGCCGAA-3'; IRES-Rev-NdeI: 5'-
400 GGGAATTCCATATGTGTGGCCATATTATCATCGTGT-3'. The third PCR used as a template the
401 PCR products from the first two rounds, along with the Oct1-NotI-For and IRES-Rev-NdeI primers. This
402 process generated a DNA fragment containing an *Oct1* cDNA fused to an IRES at the 3' end, along with
403 a *Not*I site at the 5' terminus and an *Nde*I site at the 3' terminus. The fragment was cloned into the
404 optimized, self-inactivating, nonreplicative pHAGE lentiviral vector using the *Not*I and *Nde*I restriction
405 sites. To insert a *Puro* cassette after the IRES, the cDNA was amplified using primers containing 5' *Nde*I
406 and 3' *Cla*I restriction sites. The sequences were Puro-NdeI-For: 5'-
407 GGAATTCCATATGATGACCGAGTACAAGCCCACGGT-3'; Puro-ClaI-Rev: 5'
408 GGTTTATCGATTCAGGCACCGGGCTTGC-3'. Because the IRES apparently attenuated expression of
409 the *Puro* resistance cassette in this vector, puromycin selection was performed at 0.75 μg/mL. To generate
410 an empty vector control, the vector was cut with *Nde*I and *Not*I, filled in with Klenow fragment, and re-
411 ligated.
412

413 **Electrophoretic mobility shift assay**

414 EMSA was performed using published methods[64, 65] Briefly, reactions were assembled on ice in 0.6×
415 buffer D (12 mM HEPES pH 7.9, 60 mM KCl, 0.06 mM EDTA, 12% glycerol, all buffer reagents from
416 Sigma) supplemented with 0.1 mM dithiothreitol (DTT, Sigma), 0.5 mM phenylmethylsulphonyl fluoride
417 (PMSF, Sigma), 50 μg/mL poly-dI•dC (Sigma), 1 mg/mL bovine serum albumin (BSA, Sigma) and added

418 DNA probe and lysate or recombinant protein. Diamide (Sigma) was added at final concentrations ranging
419 from 0 to 300 μM on ice during reaction assembly. 5'Cy5-labeled double-stranded DNA probes were as
420 follows: wild-type octamer, 5'-TGTCGAATGCAAATCACT-3'; mutant octamer, 5'-
421 TGTCGAATGCAAGCCACT-3'. Reactions were incubated for 30 min at room temperature prior to gel
422 loading. For experiments involving reversal of oxidative inhibition, thioredoxin (Sigma) was supplied at
423 a final concentration of 0 to 60 μM after 15 min incubation with diamide, and both thioredoxin-treated
424 and control samples were incubated for a further 15 min at 37°C prior to gel loading. Reactions were
425 resolved be electrophoresis through 0.5× TBE (45 mM Tris-Cl pH 8.3, 45 mM boric acid, 1 mM EDTA,
426 all reagents from Sigma), 4% polyacrylamide gels containing 1% glycerol. Gels were run for 90 min at
427 room temperature. DNA was visualized using a Molecular Dynamics Typhoon system.
428
429 **Recombinant Oct4 protein**
430 Wild-type and C48S mutant Oct4 constructs were generated bearing dual C-terminal FLAG and Twin-
431 Strep tags. A g-block (IDT) fragment coding for the Twin-Strep and FLAG epitope tags was inserted
432 between the *Nco*I and *Kpn*I sites in the transient expression vector pACE-MAM2 (Geneva Biotech). The
433 resultant construct was then digested with *Nco*I and *Pvu*II to insert the g-block fragment coding for either
434 wild-type or C48S mutant murine Oct4. After Sanger sequencing to confirm veracity, the plasmids were
435 transfected into Expi293F cells (ThermoFisher Scientific) using ExpiFectamine 293 Tranfection Kit
436 (ThermoFisher Scientific) following the manufacturer instructions. After 2 days, transfected cells were
437 harvested by centrifugation at 4°C for 5 min at 500×*g*, washed with ice-cold PBS, and centrifuged again
438 at 4°C for 5 min at 500×*g*. Cells were resuspended in Buffer A (20 mM HEPES pH 8.0, 1.5 mM $MgCl_2$,
439 10 mM KCl, 0.25% NP-40, 0.5 mM DTT, 2x protease inhibitor cocktail/Roche), incubated for 10 min on
440 ice, and homogenized using a Dounce homogenizer (Wheaton). The homogenate was centrifuged at 4°C
441 for 5 min at 3,800×*g*). Pelleted nuclei were resuspended in Buffer C (20 mM HEPES pH 8.0, 25% glycerol,

442 1.5 mM $MgCl_2$, 420 mM KCl, 0.25% NP-40, 0.2 mM EDTA, 0.5 mM DTT, 2x protease inhibitor cocktail)
443 and homogenized again using a Dounce homogenizer (Wheaton). Nuclei were extracted for 30 min at 4°C
444 with end-to-end rotation, then centrifuged at 4°C for 30 min at 20,000×*g*. The nuclear extract was then
445 subjected to immunoprecipitation with anti-FLAG (M2)-agarose beads (Sigma). Bead-bound proteins
446 were eluted with 200 ng/μL 3×FLAG peptide (Sigma) by incubating at 4°C for 30 min. The protein was
447 pure and monodisperse as verified using Coomassie blue staining and size exclusion chromatography.
448

449 **Generation of Oct4^{C48S} mutant mouse ESCs**

450 Mutant ESCs with a single-base missense mutation that converts Cys48 to Ser were generated using a
451 templated CRISPR approach. Briefly, a gRNA was designed to target Cas9 and cutting near the target
452 nucleotide, 5' GGCCTCGAAGCGA<u>C</u>AGATGG 3', where the underlined C indicates the target
453 nucleotide. The sgRNA was ordered as a modified synthetic sgRNA from Synthego and was complexed
454 with Cas9 protein (IDT). to form a ribonucleoprotein (RNP) complex that was electroporated with
455 program CG-104 into into 6×10^4 mouse ESCs using a LONZA 4D-Nucleofector X-unit. The RNP was
456 co-electroporated with a corresponding single-stranded oligodeoxynucleotide (ssODN) with 5' and 3'
457 phosphorothioate modifications, 5'-
458 ctcactctgtttgatcggcctttcagGAAAGGTGTTCAGCCAGACCA**CCATCT<u>C</u>T<u>A</u>G<u>A</u>TTCGAGGCC**TTGC
459 AGCTCAGCCTTAAGAACATGTGTAAGCTG-3' where lower-case denotes intronic sequence. The
460 ssODN contained the target mutation (underlined), silent mutations in the seed region (double underlined)
461 to block CRISPR re-cutting after incorporation of the donor and to generate a unique *Xba*I restriction
462 enzyme site for screening purposes (the sgRNA site is in bold). Incorporation of the ssODN sequence was
463 identified via PCR and restriction enzyme digestion, and confirmed by sanger sequencing of the 709 bp
464 region flanking the mutation. Mutant clones were generated from a pure C57BL/6 ESC line, a 129/BL6
465 hybrid line expressing constitutive GFP and a reporter line used to de-differentiate ESCs into a 2C-like
466 state. Clones were tested by karyotype and any clones with <70% normal diploid content were excluded.

467

468 **RNA-seq**

469 ESCs were first depleted by feeder cells by incubation on plates coated with gelatin. Parental and derived
470 Oct4^{C48S} ESC lines were differentiated in media lacking LIF and containing 1.5mg/L RA for 2 days.
471 Undifferentiated and RA-differentiated cells were collected, and RNA was prepared using a kit (RNeasy,
472 Qiagen). RNA poly-A selected for mRNA using NEBNext rRNA Depletion Kit v2 (Human, Mouse, Rat)
473 (E7400). Bulk RNA was processed and sequenced as reported previously[22]. Differential genes were
474 identified using 2.5-fold change and $p<0.001$ as cutoffs. GO terms were identified using PANTHER[66].
475 The high-throughput data have been uploaded to the gene expression omnibus (GEO) server, under series
476 record GSE241332. The data are accessible for download through private token gtyfyseqljufjwz while
477 they remain in private status.

478

479 **Teratoma formation**

480 Feeder-depleted parental and derived Oct4^{C48S} ESC lines were injected subcutaneously into contralateral
481 flanks of 6-8 week-old female NOD.*Rag1*$^{-/-}$*Il2rg*$^{-/-}$ (NRG, Jackson labs) mice to generate teratomas. 1×10^6
482 cells were implanted. After two weeks, mice were humanely euthanized, and tumors were excised and
483 weighed. After fixation, embedding and sectioning, slides were stained with H&E and evaluated in a
484 blinded fashion.

485

486 **Chimeric mouse assays**

487 Targeted or parental ESCs were disassociated into single cell suspension by treatment with 0.05% Trypsin
488 (ThermoFisher). After collection by centrifugation, the cells were resuspended in injection medium
489 consisting of DMEM with HEPES (ThermoFisher) supplemented with 10% FBS (Peak Serum) and 0.1
490 mM 2-mercaptoethanol (Sigma). C57B6 clones were injected into albino B6 blastocysts, while the G4N
491 and C4 lines (and derived lines, B6:129 and 129:129 in origin, respectively) were injected into B6

492 blastocysts. E3.5 blastocysts were used. 10-25 ESCs were injected into each blastocyst using
493 micromanipulators (Eppendorf) and a Leica inverted microscope. After culturing injected blastocysts in
494 KSOM (CytoSpring), approximately 12 to 18 blastocysts were surgically transferred into the uterine horns
495 of day 2.5 pseudopregnant mice or the oviduct of day 0.5 pseudopregnant mice. Mice were scored based
496 on coat and eye color by members of the transgenic core facility at weaning.

497 **Acknowledgements**

498 We thank S. Buckley for critical reading of the manuscript. We thank C. Davey and the University of Utah
499 Health Sciences Center Mouse Transgenic Core and Mutation Generation Core Facilities for assistance
500 with generating CRISPR-edited ESC lines, chimerism assays and karyotyping. We thank B. Dalley and
501 the High-Throughput Genomics core facility, and Q. Li and the Bioinformatics Analysis core facility for
502 assistance with RNA-seq. We thank D. Lum and the university Preclinical Research Resource core facility
503 for assistance with teratoma formation. We thank B. Weaver and B. Cairns for the gift of the C4 ESC line.
504 This work was supported by NIH/NIGMS grant (R01GM122778) to DT.

505

506 **Conflict of interest**

507 The authors declare that they have no conflicts of interest.

508 **Figure legends**

509

510 **Fig.1. Oct4's reprogramming activity vis-à-vis Oct1 maps to the DNA binding domain. a** Mouse Oct1
511 vs Oct4 DNA binding domain alignment. Asparagine 40 mutated in some constructs to lysine is shown in
512 blue. The key Oct1 serine 48 residue (cysteine in Oct4) is shown in blue and boxed in red. **b** Schematic
513 of the reprogramming vector with C-terminal Oct4 FLAG tag. The different domains of Oct4 and Oct1
514 are color coded. **c** Reprogramming activity of C-terminally tagged Oct4 relative to the untagged control.
515 Results show an average of experiments performed in biological triplicate. Error bars show ±standard
516 deviation. Anti-FLAG and anti-Oct4 immunoblots using tagged and untagged constructs transiently
517 transfected into 293T cells are shown at right. **d** iPSC formation assay using target primary fibroblasts
518 with an Oct4-GFP cassette. Reprogramming was assessed using the number of GFP-expressing iPSC
519 colonies. Oct4 was used as a positive control and Oct1 as a negative control. All constructs bore C-terminal
520 FLAG-tags. Experiments were performed in biological triplicate. Error bars show ±standard deviation. **e**
521 Schematic and summary of size and activity of the constructs used in (d). The activity of Oct4 in (d) was
522 set to 100%. C-terminal FLAG-tags are not shown for simplicity. **f** Anti-FLAG immunoblot showing
523 expression of the different constructs when transiently transfected into 293T cells. Cells were transduced
524 with the modified STEMCCA reprogramming vectors similar to the primary Oct4-GFP fibroblasts used
525 in the reprogramming assays. Histone H3 is shown as a loading control.

526

527 **Fig.2. Activity of Oct4 cysteine 48 in reprogramming. a** iPSC formation assay using Oct1-Oct4
528 lentiviral constructs that manipulate cysteine 4. Oct4 was used as a positive control and Oct1 as a negative
529 control. All constructs bore C-terminal FLAG-tags. Experiments were performed in biological triplicate.
530 Error bars show ±standard deviation. **b** Schematic and summary of size and activity of the constructs used
531 in (a). The activity of Oct4 in (a) was set to 100%. **c** Anti-FLAG immunoblot showing expression of the
532 different constructs when transiently transfected into 293T cells. Cells were transduced with the modified

533 STEMCCA reprogramming vectors similar to the primary Oct4-GFP fibroblasts used in the
534 reprogramming assays. Histone H3 is shown as a loading control. **d** Detail from an Oct1:octamer DNA
535 co-crystal structure (PDB ID: 1OCT)[38] showing Ser48 making hydrogen bond contacts with the DNA
536 backbone. **e** Paralog comparison of mouse POU protein amino acid residues corresponding to α-helix 3
537 of the POU_S sub-domain.
538
539 **Fig.3. Oct4 cysteine 48 confers DNA sensitivity to oxidative stress and oxidative stress-induced**
540 **ubiquitylation. a** EMSA using canonical octamer DNA and wild-type of C48S mutant full-length Oct4.
541 293T cells were transiently transfected with plasmids encoding C-terminally Twin-Strep-FLAG-tagged
542 Oct4 constructs. Protein lysates were made and immediately used for EMSA with increasing amounts of
543 diamide (see methods). The larger redox-resistant Oct1 protein that binds wild-type (WT) but not mutant
544 (Mut) probe is highlighted with an arrow. Asterisk shows a redox sensitive band with superior binding to
545 mutant rather than octamer DNA that is presumably nonspecific. Together these proteins serve as loading
546 controls. **b** Quantification of Oct4 band intensity relative to no diamide, which was set to 100%, from
547 three independent experiments. Error bars show ±standard deviation. **c** Oct4 containing a C-terminal
548 streptavidin tag was purified from 293T cell lysates using anti-FLAG beads (see methods). Isolated protein
549 was used in assays similar to (a). **d** Quantification of (c) and two other similar experiments. Error bars
550 show ±standard deviation. **e** Highly exposed Oct4 immunoblot using lysates prepared from 293T cells
551 transfected with constructs expressing C-terminally FLAG/Twin-Strep-tagged Oct4. Subsets of cells were
552 pre-treated with H_2O_2 (1 mM for 2 hr), MG-132 (10 μM for 2 hr) or both. All samples contained MG-132
553 in the lysis buffer to prevent degradation. Histone H3 is shown as a loading control. **f** Purified wild-type
554 and C48S mutant Oct4 was immunoblotted using antibodies against specific ubiquitin linkages. The FK2
555 antibody (Enzo Life Sciences) detects pan-Ub linkages, and the Ub-K48 antibody (Cell Signaling)
556 detecting Ub-K48-specific linkages. **g** Oct4 and control immunoprecipitates from wild-type ESCs were

557 immunoblotted using FK2 antibodies. Cells were treated with 10 μM MG-132 for 2 hr or with 2.5 mM
558 H_2O_2 for 2 hr.
559
560 **Fig.4. Oct4^{C48S} mutant ESCs are phenotypically normal but retain Oct4 expression when**
561 **differentiated. a** *Top:* Schematic for CRISPR-mediated *Pou5f1* (*Oct4*) mutation. A single nucleotide
562 change (arrow) converts the wild-type cysteine to the serine present in Oct1. Asterisk depict two silent
563 point mutations present in the guide RNA to monitor mutagenesis. *Bottom:* DNA sequencing trace of
564 homozygous mutant clone. **b** An example C57BL/6 ESC clone (#46) was immunoblotted for Oct4, Sox2
565 and Nanog. Histone H3 was used as a loading control. **c** Example images of pluripotent colonies from a
566 parental ESC line and derived Oct4^{C48S} mutant ESC line (#46). Images were collected at 10×
567 magnification. Scale bar = 0.2 mm. **d** Pan-Ub and control Oct4 immunoblots of Oct4 immunoprecipitates
568 from parent and Oct4^{C48S} mutant (clone #2) ESC lines. Cells were grown on gelatin without feeder
569 fibroblasts. Normal culture medium was replaced with medium lacking supplemental glutamine and β-
570 mercaptoethanol 2 hr prior to the experiment. Where indicated, cells were treated with 2.5 mM H_2O_2 for
571 a further 2 hr. Oct4 was immunoprecipitated using a rabbit polyclonal antibody (Abcam) and
572 immunoblotted using monoclonal mouse antibodies against pan-Ub (FK2, Enzo Life Sciences) and Oct4
573 (Santa Cruz). **e** Parent or derived Oct4^{C48S} mutant ESCs were differentiated without feeders using retinoic
574 acid. Lysates were collected at the indicated times and probed for Oct4 by immunoblotting. β-actin was
575 used as a loading control. **f** Average quantified Oct4/β-actin levels from three replicate immunoblots
576 conducted with the same mutant ESC clone (#46). Error bars depict ±standard deviation. **g** mRNA from
577 parental and #46 mutant cells differentiated for 2 days with RA, and undifferentiated controls, were
578 transcriptionally profiled using RNA-seq. Significantly increased genes in differentiating C48S mutant
579 cells relative to differentiating parent cells (421 genes) were intersected with the set of genes down

580 regulated in parental cells upon differentiation (1536 genes), to yield 169 genes. **h** PANTHER GO terms
581 associated with 169 genes from (g) identified using the "biological process" ontology.
582
583 **Fig.5. Oct4^{C48S} mutant ESCs differentiate abnormally. a** phase-contrast images of parent or derived
584 Oct4^{C48S} mutant ESCs (clone #46) differentiated using retinoic acid for 12 days. Images were collected at
585 the indicated times at 10× magnification. Scale bar = 0.2 mm. **b** Differentiating cells were trypsinized and
586 counted using trypan blue exclusion using a hemacytometer. Cell counts were averaged with two other
587 biological replicates (n=3) and plotted. Error bars depict ±standard deviation. **c** Proliferation was
588 measured in parent and C48S mutant ESCs differentiating in the presence of RA at days 2, 3 and 4. Cells
589 were loaded with cell trace violet (CTV) at day 2. **d** Flow cytometric TUNEL assay showing Oct4^{C48S}
590 ESCs differentiated using retinoic acid. Cells were differentiated for 2 days and incubated with FITC-
591 labeled dUTP for 1 hour at 37°C prior to flow cytometry.
592
593 **Fig.6. Oct4^{C48S} mutant ESCs form immature teratomas and do not contribute to adult chimeric**
594 **animals. a** 1×10^6 parental or Oct4^{C48S} mutant C57BL/6:129 ESCs (clone G4-N) were injected into
595 contralateral flanks or NRG mice. After two weeks, tumors were harvested, fixed, sectioned, stained with
596 H&E, deidentified and scored pathologically. The percent of cells within the tumor containing
597 undifferentiated components was averaged and plotted. N=5 independent tumors for each condition
598 combined from two separate experiments. For each tumor, values represent the average from a pair of
599 slides. Two nonadjacent sections from each tumor were used. Error bars depict ±standard deviation. *P*-
600 value was generated using an unpaired one-tailed T-test. **b** Representative images are shown. Mature
601 (differentiated) components, e.g. airway epithelium in teratomas derived from parental ESCs are
602 highlighted by yellow arrows. Asterisk indicates area of necrosis. **c** Three different parent (left side) or
603 CRISPR-mutated Oct4^{C48S} ESCs (right side) were injected into recipient blastocysts and implanted in

604 pseudo-pregnant animals. For each line, the proportion of chimeric offspring at weaning is shown as a pie

605 chart. All karyotypes were confirmed to be similar to parent lines (between 72% and 92% normal).

606 **References**

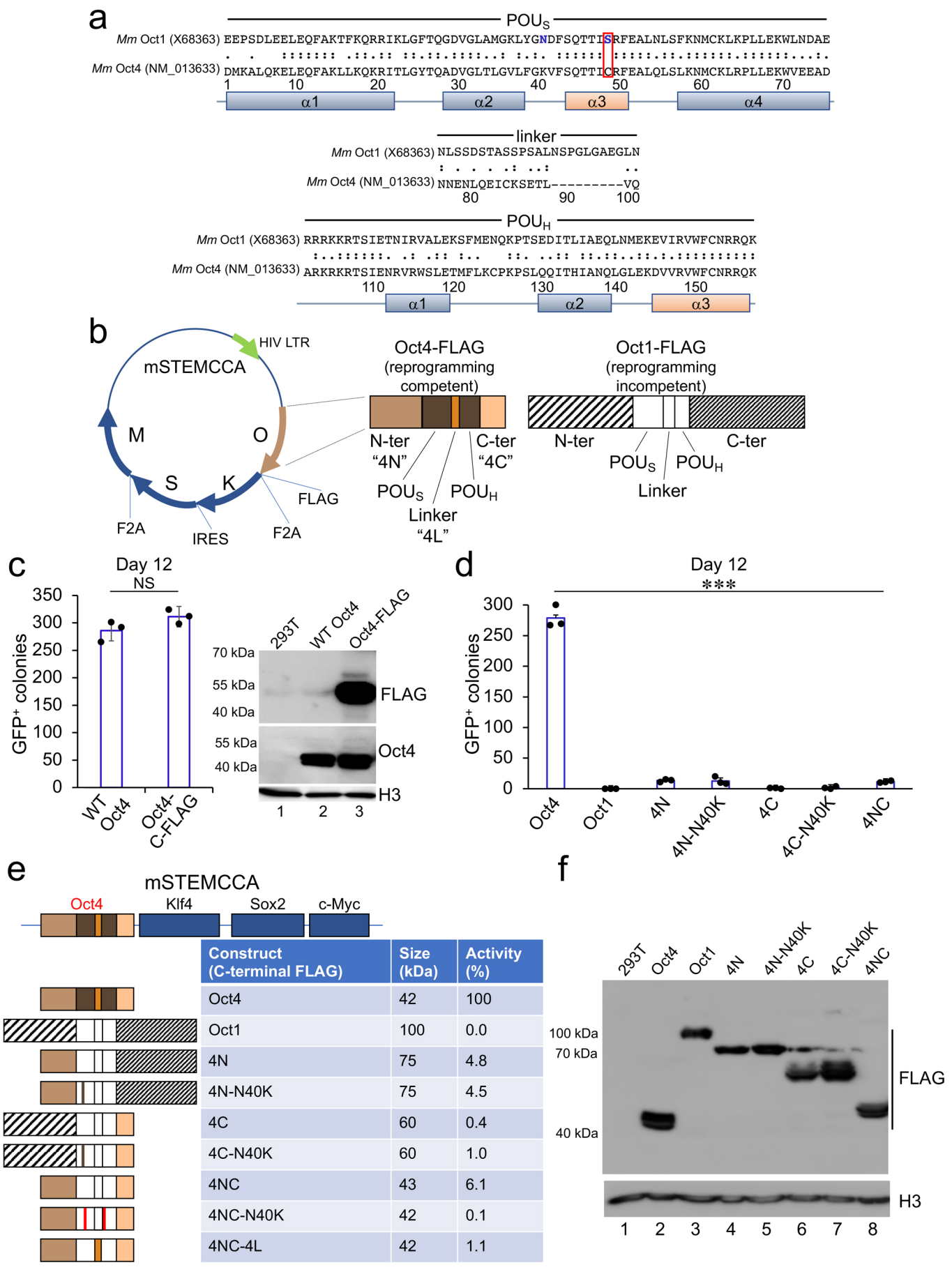
607 1. Abranches E*, et al.* Neural differentiation of embryonic stem cells in vitro: a road map to
608 neurogenesis in the embryo. *PLoS ONE* **4**, e6286 (2009).
609
610 2. Morey L, Santanach A, Di Croce L. Pluripotency and Epigenetic Factors in Mouse Embryonic
611 Stem Cell Fate Regulation. *Mol Cell Biol* **35**, 2716-2728 (2015).
612
613 3. Nichols J*, et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the
614 POU transcription factor Oct4. *Cell* **95**, 379-391 (1998).
615
616 4. Boyer LA*, et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**,
617 947-956 (2005).
618
619 5. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult
620 fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
621
622 6. Li Y*, et al.* Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small
623 molecules. *Cell Res* **21**, 196-204 (2011).
624
625 7. Gao Y*, et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of
626 DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453-469 (2013).
627
628 8. Heng JC*, et al.* The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine
629 somatic cells to pluripotent cells. *Cell Stem Cell* **6**, 167-174 (2010).
630
631 9. Esteban MA*, et al.* Vitamin C enhances the generation of mouse and human induced pluripotent
632 stem cells. *Cell Stem Cell* **6**, 71-79 (2010).
633
634 10. Rais Y*, et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**,
635 65-70 (2013).
636
637 11. Costa Y*, et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency.
638 *Nature* **495**, 370-374 (2013).
639
640 12. Vierbuchen T, Wernig M. Molecular roadblocks for cellular reprogramming. *Mol Cell* **47**, 827-
641 838 (2012).
642
643 13. Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming
644 factors' initial engagement with the genome. *Cell* **151**, 994-1004 (2012).
645
646 14. Buganim Y*, et al.* The developmental potential of iPSCs is greatly influenced by reprogramming
647 factor selection. *Cell Stem Cell* **15**, 295-309 (2014).
648
649 15. Chen J*, et al.* The combination of Tet1 with Oct4 generates high-quality mouse-induced
650 pluripotent stem cells. *Stem Cells* **33**, 686-698 (2015).
651
652 16. Yuan X, Wan H, Zhao X, Zhu S, Zhou Q, Ding S. Combined Chemical Treatment Enables Oct4-
653 Induced Reprogramming from Mouse Embryonic Fibroblasts. *Stem Cells* **29**, 549-553. (2011).

654
655 17. Velychko S, *et al.* Excluding Oct4 from Yamanaka Cocktail Unleashes the Developmental
656 Potential of iPSCs. *Cell Stem Cell* **25**, 737-753 (2019).
657
658 18. Okamoto K, Okazawa H, Okuda A, Sakai M, Muramatsu M, Hamada H. A novel octamer binding
659 transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60**, 461-472 (1990).
660
661 19. Rosner MH, *et al.* A POU-domain transcription factor in early stem cells and germ cells of the
662 mammalian embryo. *Nature* **345**, 686-692 (1990).
663
664 20. Suzuki N, Rohdewohld H, Neuman T, Gruss P, Scholer HR. Oct-6: a POU transcription factor
665 expressed in embryonal stem cells and in the developing brain. *EMBO J* **9**, 3723-3732 (1990).
666
667 21. Shen Z, *et al.* Enforcement of developmental lineage specificity by transcription factor Oct1. *ELife*
668 **6**, e20937 (2017).
669
670 22. Perovanovic J, *et al.* Oct1 cooperates with the Smad family of transcription factors to promote
671 mesodermal lineage specification. *Sci Signal* **16**, eadd5750 (2023).
672
673 23. Tantin D. Oct transcription factors in development and stem cells: insights and mechanisms.
674 *Development* **140**, 2857-2866 (2013).
675
676 24. Ferraris L, *et al.* Combinatorial binding of transcription factors in the pluripotency regions of the
677 genome. *Genome Res* **21**, 1055-1064 (2011).
678
679 25. Feng B, *et al.* Reprogramming of fibroblasts into induced pluripotent stem cells with orphan
680 nuclear receptor Esrrb. *Nat Cell Biol* **11**, 197-203 (2009).
681
682 26. Kim KP, *et al.* Permissive epigenomes endow reprogramming competence to transcriptional
683 regulators. *Nat Chem Biol* **17**, 47-56 (2021).
684
685 27. Nakagawa M, *et al.* Generation of induced pluripotent stem cells without Myc from mouse and
686 human fibroblasts. *Nat Biotechnol* **26**, 101-106 (2008).
687
688 28. Shi Y, Desponts C, Do JT, Hahm HS, Scholer HR, Ding S. Induction of pluripotent stem cells
689 from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem*
690 *Cell* **3**, 568-574 (2008).
691
692 29. Esch D, *et al.* A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol*
693 **15**, 295-301 (2013).
694
695 30. Jerabek S, *et al.* Changing POU dimerization preferences converts Oct6 into a pluripotency
696 inducer. *EMBO Rep* **18**, 319-333 (2017).
697
698 31. Jin W, *et al.* Critical POU domain residues confer Oct4 uniqueness in somatic cell reprogramming.
699 *Sci Rep* **6**, 20818 (2016).
700
701 32. An Z, *et al.* Sox2 and Klf4 as the Functional Core in Pluripotency Induction Without Exogenous
702 Oct4. *Cell Rep* **29**, 1986-2000 (2019).

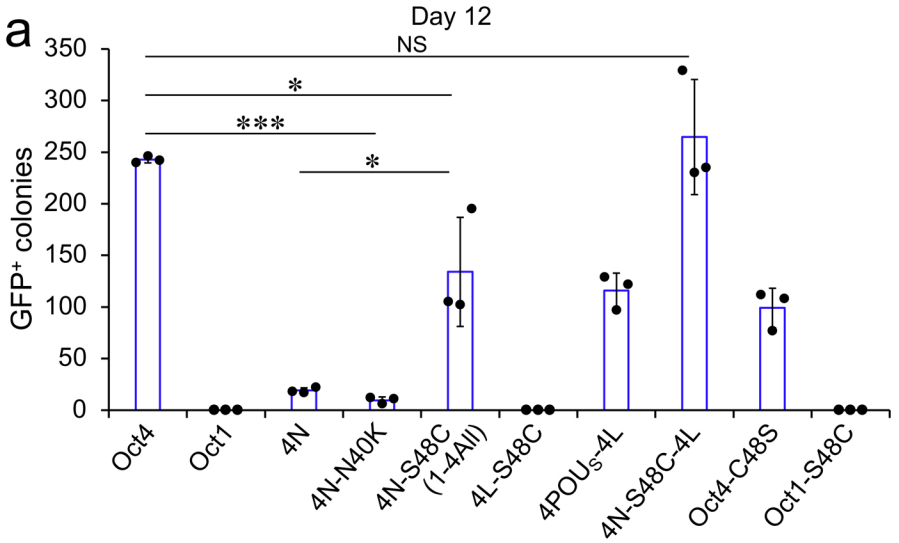
703
704 33. Lengner CJ, *et al.* Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell*
705 *Stem Cell* **1**, 403-415 (2007).
706
707 34. Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G. Induced
708 pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* **27**, 543-549
709 (2009).
710
711 35. Boija A, *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of
712 Their Activation Domains. *Cell* **175**, 1842-1855 e1816 (2018).
713
714 36. Kim KP, *et al.* Reprogramming competence of OCT factors is determined by transactivation
715 domains. *Sci Adv* **6**, eaaz7364 (2020).
716
717 37. Han D, *et al.* A balanced Oct4 interactome is crucial for maintaining pluripotency. *Sci Adv* **8**,
718 eabe4375 (2022).
719
720 38. Klemm JD, Rould MA, Aurora R, Herr W, Pabo CO. Crystal structure of the Oct-1 POU domain
721 bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77**, 21-32
722 (1994).
723
724 39. Michael AK, *et al.* Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science* **368**,
725 1460-1465 (2020).
726
727 40. Marsboom G, *et al.* Glutamine Metabolism Regulates the Pluripotency Transcription Factor Oct4.
728 *Cell Rep* **16**, 323-332 (2016).
729
730 41. Lickteig K, Lamb K, Brigman K, Rizzino A. Effects of oxidation and reduction on the binding of
731 transcription factors to cis-regulatory elements located in the FGF-4 gene. *Mol Reprod Dev* **44**,
732 146-152 (1996).
733
734 42. Guo Y, *et al.* Redox regulation of the embryonic stem cell transcription factor Oct-4 by
735 thioredoxin. *Stem Cells* **22**, 259-264. (2004).
736
737 43. King HW, Klose RJ. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support
738 gene regulatory element function in mouse embryonic stem cells. *Elife* **6**, (2017).
739
740 44. Aksoy I, *et al.* Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and
741 specify endoderm. *EMBO J* **32**, 938-953 (2013).
742
743 45. Wang Z, Oron E, Nelson B, Razis S, Ivanova N. Distinct lineage specification roles for NANOG,
744 OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* **10**, 440-454 (2012).
745
746 46. DeVeale B, *et al.* Oct4 is required ~E7.5 for proliferation in the primitive streak. *PLoS Genet* **9**,
747 e1003957 (2013).
748
749 47. De Los Angeles A, *et al.* Hallmarks of pluripotency. *Nature* **525**, 469-478 (2015).
750

751 48. Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem
752 cells. *Nature* **448**, 313-317 (2007).
753
754 49. Takahashi K*, et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined
755 factors. *Cell* **131**, 861-872 (2007).
756
757 50. Meissner A*, et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells.
758 *Nature* **454**, 766-770 (2008).
759
760 51. Bernstein BE*, et al.* A bivalent chromatin structure marks key developmental genes in embryonic
761 stem cells. *Cell* **125**, 315-326 (2006).
762
763 52. Kang J, Shakya A, Tantin D. Stem cells, stress, metabolism and cancer: a drama in two Octs.
764 *Trends Biochem Sci* **34**, 491-499 (2009).
765
766 53. Britto PJ, Knipling L, Wolff J. The local electrostatic environment determines cysteine reactivity
767 of tubulin. *J Biol Chem* **277**, 29018-29027 (2002).
768
769 54. Marinho HS, Real C, Cyrne L, Soares H, Antunes F. Hydrogen peroxide sensing, signaling and
770 regulation of transcription factors. *Redox Biol* **2**, 535-562 (2014).
771
772 55. Brigelius-Flohe R, Flohe L. Basic principles and emerging concepts in the redox control of
773 transcription factors. *Antioxid Redox Signal* **15**, 2335-2381 (2011).
774
775 56. Delaunay A, Isnard AD, Toledano MB. H2O2 sensing through oxidation of the Yap1 transcription
776 factor. *EMBO J* **19**, 5157-5166 (2000).
777
778 57. Zheng M, Aslund F, Storz G. Activation of the OxyR transcription factor by reversible disulfide
779 bond formation. *Science* **279**, 1718-1721 (1998).
780
781 58. Remenyi A, Pohl E, Scholer HR, Wilmanns M. Crystallization of redox-insensitive Oct1 POU
782 domain with different DNA-response elements. *Acta Crystallogr D Biol Crystallogr* **57**, 1634-
783 1638. (2001).
784
785 59. Mamun MMA*, et al.* Stub1 maintains proteostasis of master transcription factors in embryonic
786 stem cells. *Cell Rep* **39**, 110919 (2022).
787
788 60. Ruetz T*, et al.* Constitutively Active SMAD2/3 Are Broad-Scope Potentiators of Transcription-
789 Factor-Mediated Cellular Reprogramming. *Cell Stem Cell* **21**, 791-805 e799 (2017).
790
791 61. Shen Z, Formosa T, Tantin D. FACT Inhibition Blocks Induction But Not Maintenance of
792 Pluripotency. *Stem Cells Dev* **27**, 1693-1701 (2018).
793
794 62. Shakya A*, et al.* Pluripotency transcription factor Oct4 mediates stepwise nucleosome
795 demethylation and depletion. *Mol Cell Biol* **35**, 1014-1025 (2015).
796
797 63. Somers A*, et al.* Generation of transgene-free lung disease-specific human induced pluripotent
798 stem cells using a single excisable lentiviral stem cell cassette. *Stem Cells* **28**, 1728-1740 (2010).
799

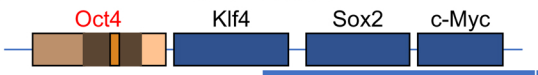
800 64. Kang J*, et al.* A general mechanism for transcription regulation by Oct1 and Oct4 in response to
801 genotoxic and oxidative stress. *Genes Dev* **23**, 208-222 (2009).
802
803 65. Tantin D, Gemberling M, Callister C, Fairbrother W. High-throughput Biochemical Analysis of
804 in-vivo Location Data Reveals Novel Distinct Classes of POU5F1(Oct4)/DNA complexes.
805 *Genome Res* **18**, 631-639. (2008).
806
807 66. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the
808 PANTHER classification system. *Nat Protoc* **8**, 1551-1566 (2013).



Construct (C-terminal FLAG)	Size (kDa)	Activity (%)
Oct4	42	100
Oct1	100	0.0
4N	75	4.8
4N-N40K	75	4.5
4C	60	0.4
4C-N40K	60	1.0
4NC	43	6.1
4NC-N40K	42	0.1
4NC-4L	42	1.1

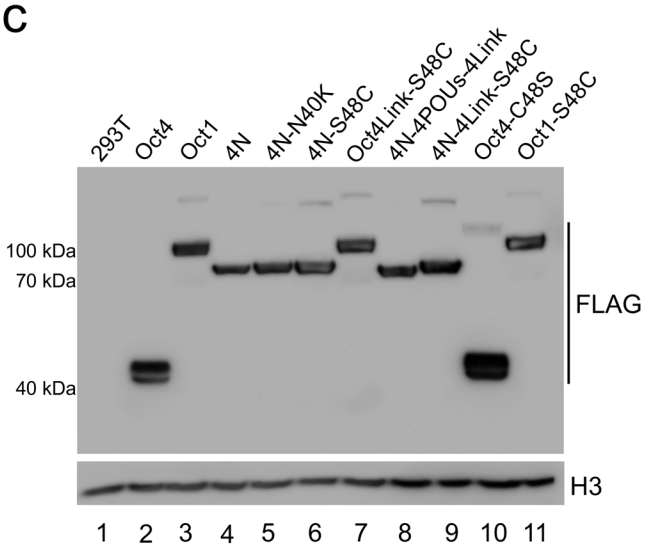


b

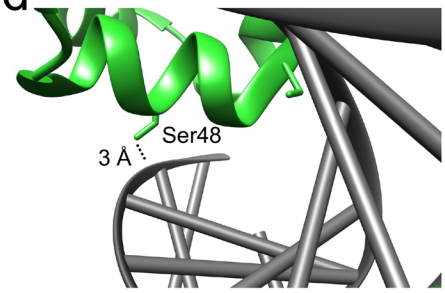


Construct (C-terminal FLAG)	Size (kDa)	Activity (%)
Oct4	42	100
Oct1	100	0.0
4N	75	7.8
4N-N40K	75	4.0
4N-S48C “1-4All”	75	55.2
4L-S48C	98	0
4N/POUs/L	72	47.8
4N/L-S48C	73	109
Oct4-C48S	42	40.8
Oct1-S48C	100	0

c



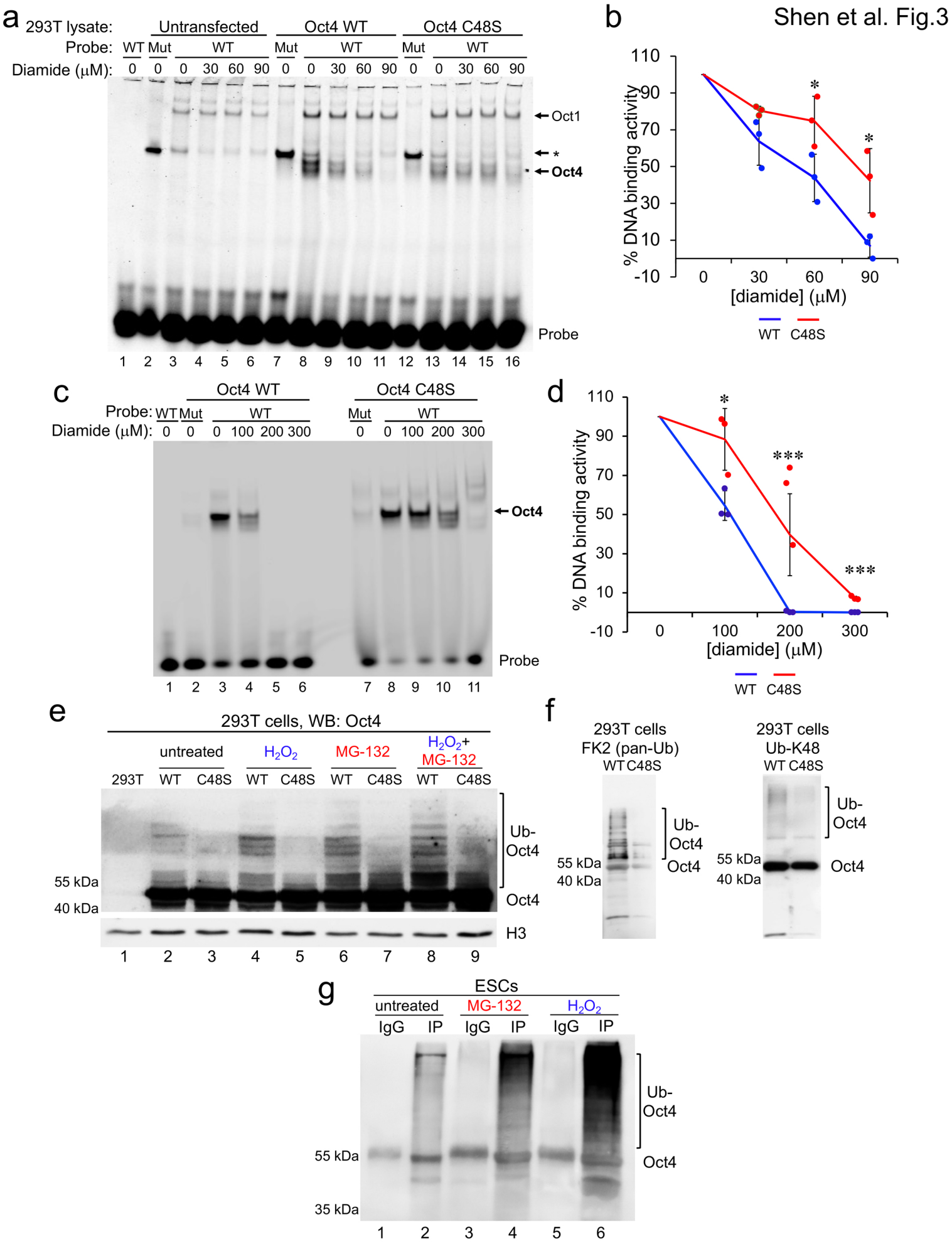
d

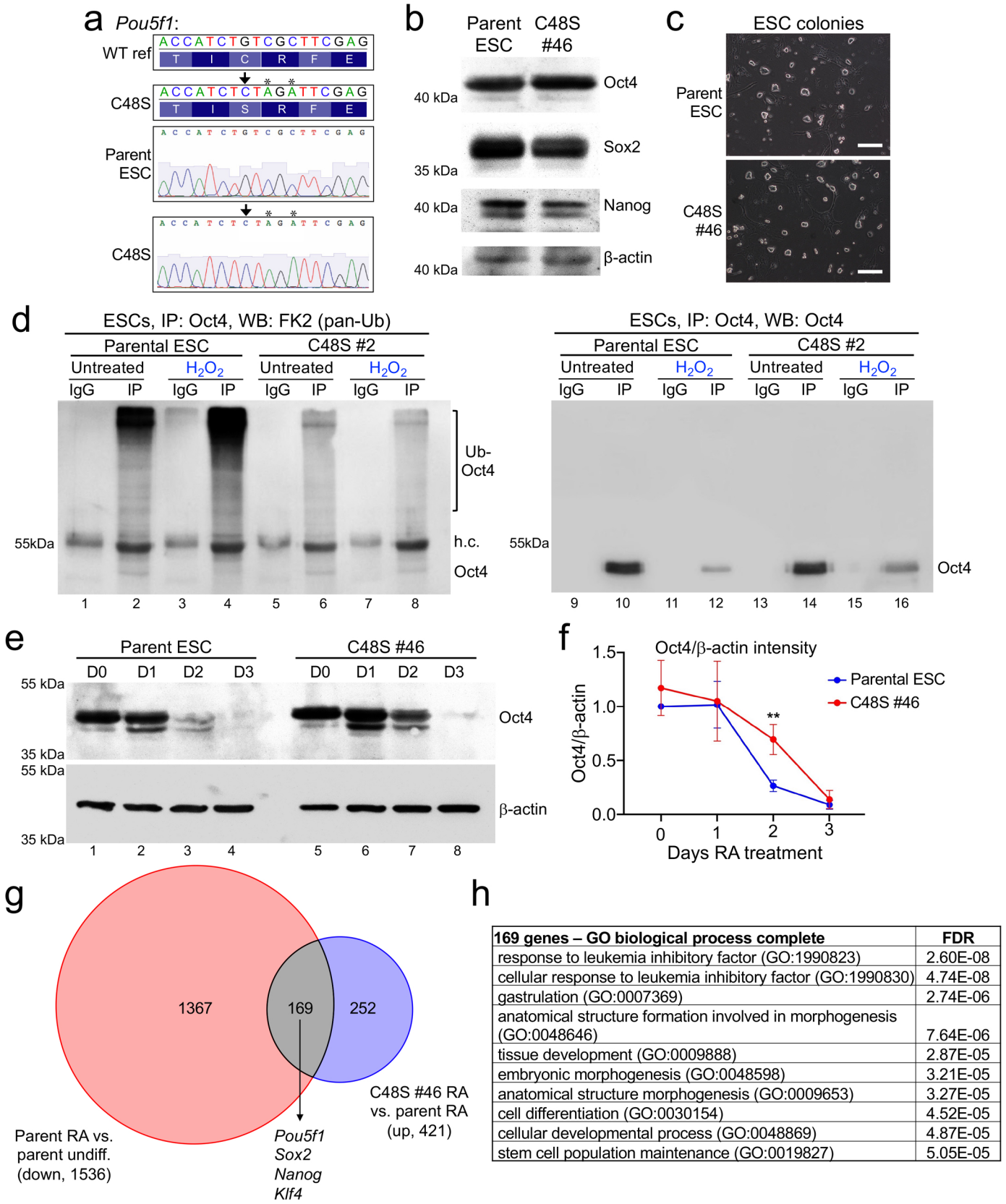


e

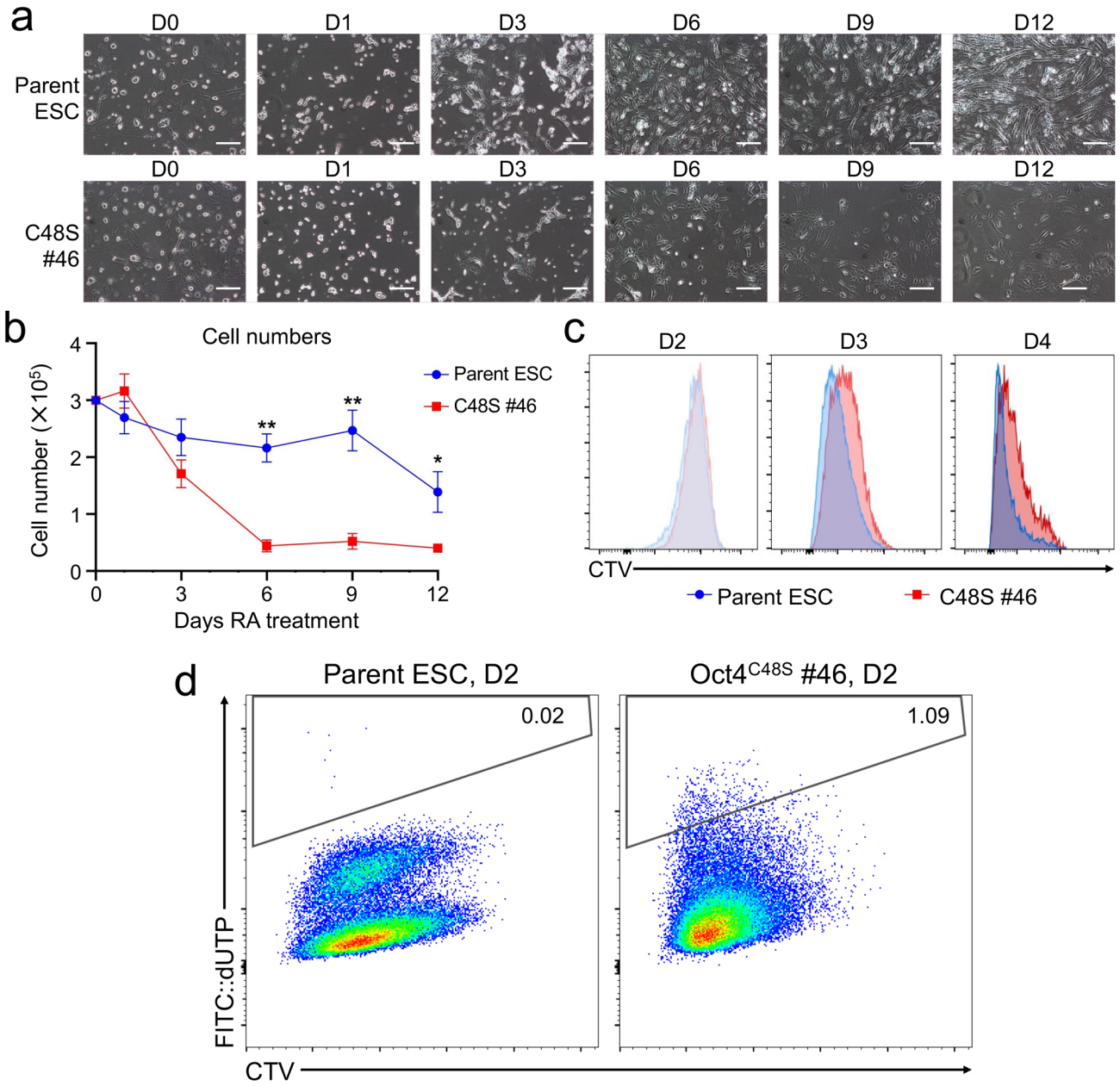
FSQTTISRFE *Mm* Oct1/Pou2f1 (NP_035267)
FSQTTICRFE *Mm* Oct4/Pou5f1 (NP_038661)
FSQTTISRFE *Mm* Oct2/Pou2f2 (Q00196-1)
FSQTTICRFE *Mm* Oct6/Pou3f1 (P21952)
FSQTTISRFE *Mm* Oct11/Pou2f3 (XP_011240722)
FSQTTICRFE *Mm* Brn2/Pou3f2 (XP_021017060)
FSQTTICRFE *Mm* Pit1/Pou1f1 (Q00286)

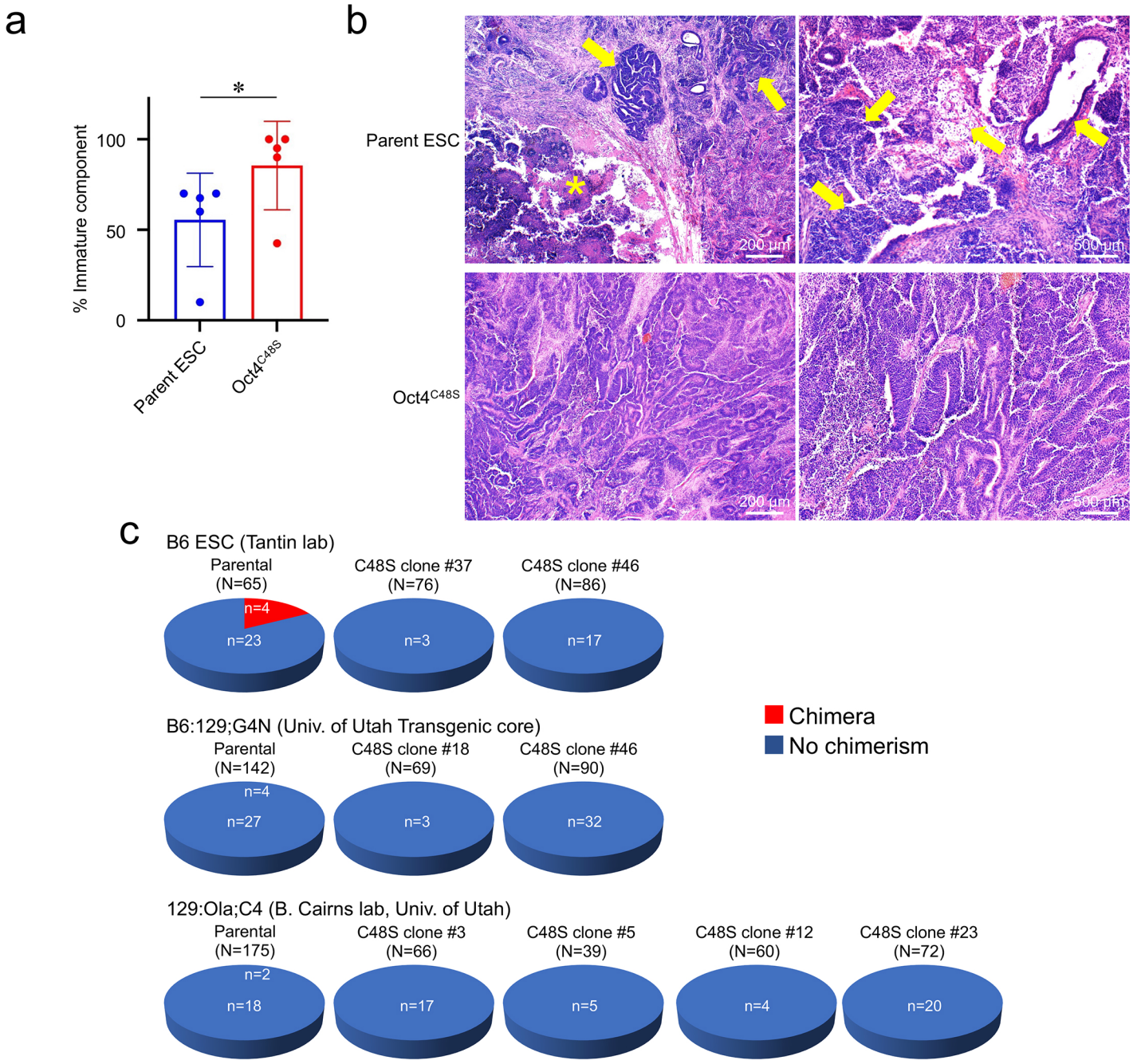
Shen et al. Fig.3





169 genes – GO biological process complete	FDR
response to leukemia inhibitory factor (GO:1990823)	2.60E-08
cellular response to leukemia inhibitory factor (GO:1990830)	4.74E-08
gastrulation (GO:0007369)	2.74E-06
anatomical structure formation involved in morphogenesis (GO:0048646)	7.64E-06
tissue development (GO:0009888)	2.87E-05
embryonic morphogenesis (GO:0048598)	3.21E-05
anatomical structure morphogenesis (GO:0009653)	3.27E-05
cell differentiation (GO:0030154)	4.52E-05
cellular developmental process (GO:0048869)	4.87E-05
stem cell population maintenance (GO:0019827)	5.05E-05


a
D0
D1
D3
D6
D9
D12
Parent ESC
C48S #46
b
Cell numbers
Cell number (X10^5)
Days RA treatment
Parent ESC
C48S #46
**
**
*
c
D2
D3
D4
CTV
Parent ESC
C48S #46
d
Parent ESC, D2
Oct4^C48S #46, D2
0.02
1.09
FITC::dUTP
CTV


a
*
% Immature component
100
50
0
Parent ESC
Oct4C48S
b
Parent ESC
Oct4C48S
200 µm
500 µm
200 µm
500 µm
c
B6 ESC (Tantin lab)
Parental (N=65)
C48S clone #37 (N=76)
C48S clone #46 (N=86)
n=4
n=23
n=3
n=17
B6:129;G4N (Univ. of Utah Transgenic core)
Parental (N=142)
C48S clone #18 (N=69)
C48S clone #46 (N=90)
n=4
n=27
n=3
n=32
129:Ola;C4 (B. Cairns lab, Univ. of Utah)
Parental (N=175)
C48S clone #3 (N=66)
C48S clone #5 (N=39)
C48S clone #12 (N=60)
C48S clone #23 (N=72)
n=2
n=18
n=17
n=5
n=4
n=20
Chimera
No chimerism