B1 SINE-binding ZFP266 impedes reprogramming through suppression of chromatin opening by pioneering factors

Daniel F Kaemena[1], Masahito Yoshihara[2], James Ashmore[1], Meryam Beniazza[1], Suling Zhao[1], Mårten Bertenstam[3], Victor Olariu[3], Shintaro Katayama[2,4,5], Keisuke Okita[6], Simon Tomlinson[1], Kosuke Yusa[7,8] & Keisuke Kaji[1]

[1]MRC Centre for Regenerative Medicine, University of Edinburgh, Edinburgh BioQuarter, 5 Little France Drive, Edinburgh, EH16 4UU, Scotland, UK. [2]Department of Biosciences and Nutrition, Karolinska Institutet, 141 83 Stockholm, Sweden. [3]Computational Biology and Biological Physics, Lund University, Lund, Sweden. [4]Research Programs Unit, Stem Cells and Metabolism Research Program (STEMM), University of Helsinki, 00014 Helsinki, Finland. [5]Folkhälsan Research Center, 00014 Helsinki, Finland. [6]Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan. [7]Stem Cell Genetics, Wellcome Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK. [8]Stem Cell Genetics, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, Japan.

Correspondence: keisuke.Kaji@ed.ac.uk

Abstract

Successful generation of induced pluripotent stem cells (iPSCs) via the overexpression of *Oct4 (Pou5f1)*, *Sox2*, *Klf4* and *c-Myc* (OSKM) highlights the power of transcription factor (TF)-mediated cellular conversions. Nevertheless, iPSC reprogramming is inherently inefficient and understanding the molecular mechanisms underlying this inefficiency holds the key to control cellular identity successfully. Here, we report 16 novel reprogramming roadblock genes identified by CRISPR/Cas9-mediated genome-wide knockout (KO) screening. Of these, disruption of KRAB zinc finger protein (KRAB-ZFP) *Zfp266* strongly and consistently enhanced iPSC generation in several iPSC reprogramming settings, emerging as the most robust roadblock. Further analyses revealed that ZFP266 bound Short Interspersed Nuclear Elements (SINEs) adjacent to OSK binding sites and impedes chromatin opening. This work serves as a resource for better understanding reprogramming mechanisms and proposes SINEs as a critical genetic element that regulates chromatin accessibility at enhancers for efficient pluripotency induction.

Introduction

The reprogramming of somatic cells into iPSCs via the overexpression of *Oct4 (Pou5f1)*, *Sox2*, *Klf4* and *c-Myc* (OSKM) has provided an important tool for medical research and cell therapies[1]. Equally importantly, the generation of fully functional iPSCs indistinguishable from ESCs from somatic cells has clearly demonstrated that cellular identity can be completely converted from one to another by simple overexpression of master transcription factors. This has provided a model system to understand how to control cellular identity. Inhibition of *Trp53* and *Cdkn1a* (*p21*) revealed OSKM-induced apoptosis and senescence as a major roadblock of iPSC generation[2–8]. Knockdown of *Dot1l* and *Suv39h1* demonstrated H3K79me and H3K9me3 as critical epigenetic modifications that impede this cell conversion[9,10]. Thus, identifying genes that act against successful reprogramming offers the foundation to understand critical molecular mechanisms involved in pluripotency induction.

Transposable elements (TEs), which consist of about 40% of mouse and human genome, take part in gene expression regulation as cis-regulatory elements in the genome or non-coding RNAs[11], and importance of

48 many types of TEs in pluripotent cells has been described[12–15]. Knockdown of the long interspersed element
49 1 (LINE1) inhibits mouse ESC self-renewal and induces transition to a 2C state[12]. KLF4 activates transcription
50 of HERVH during reprogramming, and whose down-regulation is critical for exit from the pluripotent state
51 of human iPSCs[13]. KLF4 is also responsible for opening thousands of genomic loci in naïve human ESCs via
52 binding to HERVH, HERVK, and non-LTR retrotransposon SVA subgroups[14]. Similarly, chromatin accessibility
53 of SINEs, which constitute ~25% of TEs, is particularly high in mouse pre-implantation embryos and ESCs[15],
54 however the functional importance of this has not yet been addressed. Krüppel-associated box (KRAB) zinc-
55 finger proteins (ZFPs) are the largest transcription factor family in mouse and human genomes with over
56 300 members, which have evolved to supress rapidly mutating TEs, and about two thirds of KRAB-ZFPs bind
57 to TEs[16]. Nevertheless, only a small number of KRAB-ZFPs that bind to SINEs is known[17].

58 Here, we report an unbiased genome-wide CRISPR KO screen with a library containing 90,230 sgRNAs
59 targeting 18,424 protein coding genes. This screen identified 16 genes as novel reprogramming roadblocks,
60 as well as 8 previously reported inhibitory genes. Of those, KO of the KRAB-ZFP gene *Zfp266* has
61 accelerated the kinetics of reprogramming and improved efficiency of iPSC generation 4- to 10-fold in
62 various reprogramming contexts. We revealed that ZFP266 binds to SINEs adjacent to OSKM binding sites
63 during reprogramming and impedes chromatin opening. Furthermore, replacing its KRAB co-suppressor
64 interacting domain with a co-activator interacting domain converted ZFP266 from a reprogramming
65 inhibitor to a reprogramming facilitator. This indicated that SINEs next to OSKM binding sites during
66 reprogramming were critical genetic elements that modulate efficiency of OSKM-mediated iPSC generation.
67 This work serves as a resource for better understanding of reprogramming mechanisms and highlights
68 SINEs as novel transposable elements (TEs) involved in pluripotency induction.

69

70 **A CRISPR/Cas9-mediated genome-wide KO screen identified 16 novel reprogramming roadblock genes**

71 We have previously generated a Cas9 expressing mouse ES cell line, named Cas9 TNG MKOS, with a *Nanog-*
72 *GFP-ires-Puro* reporter and a doxycycline-inducible *MKOS-ires-mOrange* polycistronic reprogramming
73 cassette in the *Sp3* locus (Supplemental Figure S1A and S1B)[18,19]. Efficient KO by lentiviral sgRNA delivery in

74 both Cas9 TNG MKOS ESCs and mouse embryonic fibroblasts (MEFs), generated through morula
75 aggregation of these ESCs, was confirmed using sgRNA against ICAM1 and CD44, respectively, resulting in
76 >80% loss of protein within 72 hours (Supplemental Figure S1C and S1D). Reprogramming of Cas9 TNG
77 MKOS MEFs following sgRNAs transduction against known roadblock genes, *Trp53* and *Rb1*[3,4,6–8,20], and
78 essential genes, *Pou5f1* and *Kdm6a*[21], reproduced the expected reprogramming enhancement and
79 reduction phenotypes, respectively (Supplemental Figure S1E-S1G), confirming that the CRISPR-based KO
80 system as a powerful tool to investigate gene function in reprogramming. We then performed genome-
81 wide KO screening using a previously published lentiviral sgRNA library[22], with an optimized reprogramming
82 condition consisting of 8 days of reprogramming factor expression followed by 8 days of puromycin
83 selection for *Nanog*-GFP$^+$ iPSCs (Figure 1A and Supplemental Figure S1H-S1K). This condition resulted in an
84 average coverage of ~170 MEFs/sgRNA/screening replicate. Genomic DNA from flow-sorted *Nanog*-GFP$^+$
85 iPSCs was then collected in triplicate, and integrated sgRNAs were Illumina-sequenced after PCR
86 amplification alongside the original sgRNA plasmid library.

87 The normalized read counts of all sgRNAs and the results of the statistical tests to identify novel roadblock
88 genes with a software, MAGeCK[23], are available in Supplemental Tables S1, S2 and at https://kaji-crispr-
89 screen-updated.netlify.app. Using a false discovery rate (FDR) <0.1 as a cut-off, we identified 24 genes as
90 reprogramming roadblocks (Figure 1B). This included 16 novel candidates as well as 5 previously
91 characterized genes: *Trp53*, *Cdkn1a*, *Dotl1* and AP-1 transcription factor components *Jun* and *Fosl2*[24,25], and
92 3 genes previously uncharacterized yet identified in other screens, *Men1, Gtf2i and Cdk13*[26,27], signifying
93 the robustness of our screen (Figure 1C and 1D). When the top 3 ranked sgRNA for each gene were
94 individually tested, transduction of *Trp53* and *Cdkn1a* sgRNAs produced the largest increase in *Nanog*-GFP$^+$
95 colony numbers (Figure 1E and 1F), although they also significantly increased the number of partially
96 reprogrammed colonies (Figure 1F and Supplemental Figure S2A). Transduction of sgRNAs targeting all
97 other genes, except for the previously reported roadblock *Cdk13,* enhanced *Nanog*-GFP$^+$ colony formation
98 between 2- and 6-fold (Figure 1E and Supplemental Figure S2A), verifying the inhibitory effects of the novel
99 reprogramming roadblock genes. Expression of the validated 23 roadblock genes during reprogramming did
100 not follow any common particular pattern and many of them exhibited consistently low expression,

101 compared to the common housekeeping genes or the reprogramming and pluripotency marker genes[28]
102 (Supplemental Figure S2B). This highlights the advantage of functional screening to identify their inhibitory
103 effects over expression profiling.

104 Of the 16 novel roadblock genes, KO of 8 resulted in a >4-fold enhancement similar to or better than
105 previously reported roadblocks (Figure 1E). We therefore further characterise these 8 novel reprogramming
106 roadblocks; *Fam122a, Zfp266, Bcorl1, Usp28*, *Usp34, Zc3h10, Scaf8* and *Spop* (Figure 1F, blue), alongside
107 the previously reported roadblocks *Trp53, Cdkn1a, Men1, Dot1l* and *Gtf2i* (Figure 1F, green), as 13 top
108 roadblocks.

109

110 ***Zfp266* KO consistently enhances and accelerates the attainment of pluripotency**

111 Reprogramming roadblock function is influenced by multiple elements such as the stoichiometry or
112 expression levels of OSKM, culture conditions, and starting cell types[18]. Thus, we examined the KO effects
113 of our 13 top roadblocks in different reprogramming contexts. We first performed *piggyBac* transposon-
114 based reprogramming with a *MKOS* or STEMCCA (*OKSM*) reprogramming cassette[29,30] (Supplemental Figure
115 S3A). The STEMCCA cassette expresses lower levels of KLF4 protein due to an N-terminal truncation
116 followed by a 2A peptide[31], resulting in inefficient mesenchymal-epithelial transition (MET) and a higher
117 proportion of partially reprogrammed cells[18,32]. *piggyBac* delivery of the MKOS cassette together with each
118 sgRNA against all the 13 roadblocks enhanced reprogramming of *Cas9 Nanog*-GFP MEFs as seen with Cas9
119 TNG MKOS MEFs before (Figure 2A and 2B), despite a markedly lower KO efficiency with the *piggyBac*
120 system compared to lentiviral sgRNA delivery (Supplemental Figure S3B). In STEMCCA-mediated *piggyBac*
121 reprogramming, KO of all, except *Cdkn1a*, *Fam122a* and *Zc3h10,* increased the number of *Nanog*-GFP$^+$
122 colonies. Particularly, the KO of *Zfp266* increased numbers of *Nanog*-GFP$^+$ colonies ~10-fold with almost all
123 colonies expressing *Nanog*-GFP (Figure 2C and 2D), while *Cdkn1a*, *Fam122a* knockout drastically increased
124 only *Nanog*-GFP$^-$ colony numbers (Figure 2C and 2D). When *piggyBac MKOS*+sgRNA vectors were used to
125 reprogram *Cas9* expressing neural stem cells (NSCs), sgRNAs against *Men1, Fam122a, Zfp266* and *Usp34*
126 increased reprogramming efficiency, with KO of *Zfp266* again leading to the greatest enhancement in

127 NANOG^{+} colony formation (~5-fold) (Figure 2E and 2F). When we explored reprogramming kinetics using
128 Cas9 TNG MKOS MEFs by assessing expression changes of reprogramming markers, CD44, ICAM1 and
129 *Nanog*-GFP[33], KO of 5 genes, *Men1*, *Gtf2i, Dot1l, Zfp266* and *Zc3h10,* demonstrated the accelerated
130 reprogramming (Figure 2G, 2H, and Supplementary Figure 3C). In summary, KO of *Zfp266* exhibited the
131 most context-independent and robust reprogramming enhancement amongst all roadblock genes we
132 identified. We therefore investigated further how *Zfp266* impedes the reprogramming process.

133

134 **ZFP266 impedes activation of pluripotency genes via its KRAB domain**

135 *Zfp266* encodes a KRAB-ZF protein with a singular KRAB-A module in the N-terminus and putative DNA
136 binding domain with 12x C2H2 type zinc finger array in the C-terminus (Figure 3A). KRAB domains are
137 known to interact with co-suppressor KAP-1/TRIM28, a scaffold protein that can recruit epigenetic
138 modifiers and promote the formation of heterochromatin and transcriptional repression[34], suggesting that
139 ZFP266 acts as a suppressor. Consistent with the reprogramming enhancement by *Zfp266* KO,
140 overexpression of exogenous *Zfp266* completely hindered reprogramming (Figure 3B). Exogenous
141 overexpression of *Zfp266* mutants either lacking the entire KRAB domain or containing point mutations
142 which disrupt the interaction with KAP-1/TRIM28[35,36] could not abolish *Zfp266* sgRNA-mediated
143 reprogramming enhancement even though the *Zfp266* mutants were resistant to the sgRNA (Figure 3C and
144 3D). This clearly demonstrates that Zfp266 inhibits reprogramming through its KRAB-domain.

145

146 To assess further the function of zfp266, we examined the gene expression changes associated with
147 depletion of *Zfp266* (Supplemental Tables S3). RNA-seq of MEFs 4 days after *Zfp266* sgRNA transduction
148 revealed only 4 differentially expressed genes (DEGs) (FDR<0.05, log2FC<|1|), demonstrating that loss of
149 ZFP266 alone was not sufficient to cause drastic gene expression changes in MEFs (Figure 3E, Supplemental
150 Tables S4). In contrast, the number of DEGs between *Zfp266* KO and wild-type cells rapidly increased during
151 reprogramming from 24 at day 3 to 222 at day 5, and to 1761 at day 7 (Figure 3E, Supplemental Tables S5-
152 S7). The majority of DEGs at day 3 and day 5 were upregulated (75% and 67% respectively, Figure 3E),

153 consistent with the predicted role for *Zfp266* as a transcriptional suppressor. Enhanced up-regulation of
154 pluripotency-associated genes *Piwil2* and *Dppa5a* at day 3, *Nanog, Esrrb, Dppa2*, and *Tcl1, etc* at day 5 was
155 already detected in *Zfp266* KO (Figure 3E). Furthermore, over 60% (11/18), 80% (120/149), 69% (575/835)
156 of up-regulated DEGs in *Zfp266* KO cells at day 3, 5, 7 of reprogramming were genes more highly expressed
157 in ESCs compared to MEFs (FDR<0.05, log2FC<|1|) (Figure 3E, green, Supplemental Tables S8). Gene
158 ontology (GO) enrichment analysis identified 'stem cell population maintenance' in day 5 and 'response to
159 leukaemia inhibitory factor' in day 7 up-regulated DEGs as one of the most enriched terms (Figure 3F),
160 while downregulated DEGs at day 7 were significantly enriched in developmental and differentiation terms
161 (Figure 3F). Principal component analysis (PCA) also indicated that gene expression changes that have
162 occurred in *Zfp266* KO reprogramming at day 5 and 7 reflected an accelerated transition towards a
163 pluripotent state (Figure 3G). Taken together, these data indicate that *Zfp266* KO enhances and accelerates
164 reprogramming by permitting a more efficient activation of pluripotency genes by OSKM.

165

166 ***Zfp266* KO in MEFs results in chromatin opening at the SINE-containing Zfp266 binding sites**

167 One possible mechanism by which ZFP266 impedes reprogramming is that ZFP266 binds and locks the
168 pluripotency loci in MEFs and other differentiated cells. To investigate this possibility, we mapped ZFP266
169 binding sites in MEFs using DamID-seq, which does not required specific antibodies[37,38]. This identified
170 15,119 unique ZFP266 binding sites (Figure 4A), predominantly situated in introns or intergenic regions
171 (Supplementary Figure 4A, Supplemental Tables S9). These ZFP266 binding sites have low chromatin
172 accessibility as measured by ATAC-seq in MEFs, 72 hours after reprogramming as well as in iPSCs[39] (Figure
173 4A). ZFP266 binding sites were predominantly enriched for somatic AP-1 TF motifs (Figure 4B), and little
174 OSKM binding was observed at the same loci in 48hr reprogramming or ESC ChIP-Seq datasets[40,41]
175 (Supplementary Figure 4B and 4C). This disputed the idea that ZFP266 functions as a suppressor at the
176 pluripotency-related gene loci in differentiated cells. Thus, we investigated whether any TE families were
177 enriched in ZFP266 binding sites, since many KRAB-ZFPs are known to bind and supress transcription of
178 TEs[16]. In line with this, we found the 15,119 ZFP266 DamID-seq peaks in MEFs were highly enriched in SINEs

179 with about two thirds (10,523) overlapping with SINEs (Figure 4C and 4D). Of SINE subfamilies, B1 SINEs in
180 particular exhibited both the most significant enrichment and the most abundant overlap with ZFP266
181 binding sites (Figure 4C and 4D). Furthermore, de novo motif analysis of ZFP266 binding sites identified 3
182 long de novo motifs which all corresponded to parts of the B1 SINE consensus sequence (Figure 4E),
183 suggesting ZFP266 might bind B1 SINEs.

184 We next examined how depletion of ZFP266 might affect chromatin accessibility. To this end, we
185 performed ATAC-Seq of *Zfp266* KO MEFs and identified 479 more open regions (MORs) compared to WT
186 MEFs, while only one locus was found to be a more closed (Figure 4F and Supplemental Tables S10).
187 Considering the predicted ZFP266 function as a repressor, we next examined whether ZFP266 binds to the
188 MORs in wild-type MEFs. Although only about 25% (123/479) of *Zfp266* KO MEF MORs overlapped with
189 ZFP266 DamID-seq peaks, non-overlapped MORs also had increased DamID-seq signals albeit at a lower
190 level (Figure 4G), unlike randomly selected control regions with similar chromatin accessibility
191 (Supplementary Figure 4D). This suggests that more than 25% of MORs are likely bound by ZFP266, while
192 they were not identified as a 'peak' with our DamID-seq due to the cut-off criteria and/or technical
193 limitations. In fact, similar to ZFP266 DamID-seq peaks in MEFs, *Zfp266* KO MEF MORs were mainly located
194 in intergenic regions and introns (Supplementary Figure 4E), and enriched in AP-1 TF motifs, with 87% of all
195 *Zfp266* KO MEF MORs containing at least one AP-1 TF motif (Figure 4H). *De novo* motif discovery analysis
196 also identified 5 motifs that overlap with the B1 SINE consensus sequence (Figure 4I), consistent with the
197 fact that SINE was the most enriched repetitive element (Figure 4J) and 92% (1418/1522) of *Zfp266* KO MEF
198 MORs had at least one SINE (Figure 4K). Overall, these data suggest that ZFP266 binds to SINEs in MEFs to
199 keep target loci closed, and removal of ZFP266 allows TFs that binds nearby, like AP-1 TFs, to facilitate
200 chromatin opening (Figure 4L). However, ZFP266 does neither bind to nor regulate pluripotency gene loci in
201 MEFs.

202

203 **Zfp266 KO in reprogramming results in chromatin opening at SINE-containing OSK binding sites**

204 In order to address why *Zfp266* KO results in significant reprogramming enhancement, we performed

205 ATAC-seq 72 hours after reprogramming with and without *Zfp266* KO. Similar to the KO effects in MEF,
206 *Zfp266* KO reprogramming cells exhibited 1522 MORs, majority of which were situated in intergenic regions
207 and introns (Supplementary Figure 5A), and only 86 more closed regions compared to wild-type cells
208 (Figure 5A, Supplemental Tables S11). They were also significantly enriched in SINEs, particularly B1 SINE
209 (Figure 5B), with >90% (1459/1522) of MORs containing at least one SINE (Figure 5C). *De novo* motif
210 discovery analysis also identified motifs that correspond to the B1 SINE consensus sequence as the most
211 significant motifs (Figure 5D). However, these loci hardly overlapped with *Zfp266* KO MEF MORs (Figure 5E),
212 suggesting a context dependency for which loci become more open in the absence of ZFP266. The overlap
213 with MEF ZFP266 DamID-seq peaks was also minimal, with only ~10% of MORs overlapping (Supplemental
214 Figure S5B), and non-overlapped MORs did not have increased DamID-seq signals compared to the control
215 regions with a similar chromatin accessibility (Supplemental Figure S5B). This indicated that upon OSKM
216 expression ZFP266 changes their binding sites, in which ZFP266 regulates chromatin accessibility. TF motif
217 enrichment analysis revealed that KLF, SOX and the OCT4::SOX2 motifs were highly enriched in *Zfp266* KO
218 reprogramming MORs, with particularly KLF family (KLF1, KLF5, KLF4, KLF9, KLF12) motifs identified in >90%
219 of these MORs, while AP-1 TF motifs were also enriched (Figure 5F). Next, we classified *Zfp266* KO
220 reprogramming MORs into two groups using *K*-means clustering (Figure 5G). The first cluster (121 regions)
221 are open in both wild-type and *Zfp266* KO MEFs, and then become more closed upon reprogramming,
222 while *Zfp266* KO cells are more resistant to this closing (cluster 1, Figure 5G). The second cluster contained
223 the majority (1401) of *Zfp266* KO reprogramming MORs, which are closed in both wild-type and *Zfp266* KO
224 MEFs, and become open following reprogramming, an effect which is enhanced when *Zfp266* is knocked
225 out (cluster 2, Figure 5G). Cluster 2 indicates that removal of ZFP266 facilitates OSKM-mediated chromatin
226 opening. In fact, we observed OSK binding in cluster 2 reprogramming MORs with particularly strong KLF4
227 signal using published reprogramming 48-hour ChIP-Seq datasets[41] (Figure 5H). Similar OSK binding was
228 observed in an ESC ChIP-Seq dataset albeit to a lesser extent[40] (Supplemental Figure S5C), and about one
229 third of the cluster 2 loci have an open chromatin state in iPSCs (Supplemental Figure S5D), suggesting that
230 some of the MORs are OSK targets in pluripotent cells. Interestingly, while both OSK binding and motifs
231 were enriched close to the MOR peak summit (within 70 bp) (Figure 5H and 5I), SINEs were depleted from

232 summits and were instead enriched immediately upstream or downstream (~70 bp away from the summit),
233 therefore being located immediately adjacent to OSK motifs and binding sites (Figure 5I and 5J).
234 Furthermore, B1 SINEs within the MORs were likely orientated such that the 5' head sequence was
235 positioned inwards facing towards the peak summit (Figure 5I and 5J). This positional and directional bias
236 within the MOR was exclusive to B1 SINE subfamilies as B2 SINEs exhibited no such bias (Supplemental
237 Figure S5E). Based on these data together with a report that somatic TFs' binding sites drastically change
238 upon OSKM expression[40], we speculated that ZFP266 was relocated to B1 SINEs next to OSK binding sites
239 during reprogramming, where it then impeded chromatin opening.

240

241 **Facilitating chromatin opening at ZFP266 targeted SINEs enhances reprogramming**

242 In order to validate binding of ZFP266 to SINEs, we generated an activator version of *Zfp266* with the KRAB
243 domain replaced by a flexible linker and three transactivating domains VP64, p65 and Rta (VPR)[42] (Figure
244 6A), and performed luciferase reporter assays using HEK293 cells. Clear luciferase expression was observed
245 when the VPR-*Zfp266*, but not BFP or VPR only controls, was co-transfected with a reporter plasmid
246 containing the B1 SINE consensus sequence upstream of a SV40 minimal promoter (Figure 6B). Co-
247 expression of wild-type *Zfp266* alongside VPR-*Zfp266* attenuated this reporter expression (Figure 6C),
248 confirming ZFP266 specifically binds B1 SINEs. We next examined whether VPR-ZFP266 can bind to *Zfp266*
249 KO reprogramming MORs using the *Luciferase* assay. We selected SINE containing MORs in three genes,
250 *B3gnt3*, *Piwil2* and *Snx20*, whose transient up-regulation during reprogramming was significantly
251 augmented by *Zfp266* KO (Figure 6D). These loci are closed in MEFs, open up more in *Zfp266* KO cells upon
252 reprogramming, and are bound by KLF4 at 48 hours of reprogramming (Figure 6E). Each MOR was cloned
253 upstream of a minimal SV40 promoter in both a forward and reverse orientation in a luciferase reporter
254 vector. We found that VPR-*Zfp266* could also enhance luciferase expression from these MOR reporter
255 vectors (Figure 6F), while co-expression of wild-type *Zfp266* would then ablate it (Figure 6G). Deleting B1
256 SINE sequences from the *B3gnt3*, *Piwil2* and *Snx20* MORs diminished VPR-*Zfp266*'s ability to enhance
257 *luciferase* expression, confirming that ZFP266 binds to reprogramming MORs specifically via B1 SINE

258 sequences (Figure 6H-6J). We also confirmed that OSKM expression enhanced luciferase expression from
259 the *B3gnt3* and Snx20 MOR containing reporter vectors in MEFs (Figure 6K), indicating the MORs have
260 OSKM-dependent enhancer activity. In addition, removing B1 SINE sequences from the *Snx20* MOR led to
261 an increase in *Luciferase* expression following OSKM induction (Figure 6L), demonstrating B1 SINEs function
262 to repress reprogramming factor-mediated transactivation presumably via endogenous ZFP266 binding.
263 Finally, overexpression of the VPR-*Zfp266* together with OSKM lead to accelerated and enhanced
264 reprogramming with a robust appearance of *Nanog*-GFP^{+} colonies by day 9 (Figure 6M and 6N). Taken
265 together, we propose a model where 1) ZFP266 binds to B1 SINEs adjacent to OSK binding sites during
266 reprogramming and acts to impede chromatin opening, and 2) KO of *Zfp266* (or recruitment of co-
267 activators to these loci) tips the balance in favour of OSK, allowing them to establish a more open
268 chromatin state to drive gene activation necessary for successful reprogramming (Figure 6O).

269

270 **Discussion**

271 Reprogramming towards iPSCs is a conflict between OSKM transcription factors trying to establish a
272 pluripotent state and somatic factors trying to resist this disruption in cell identity. We have identified 16
273 novel reprogramming roadblock genes whose depletion facilitates OSKM-mediated pluripotency induction.
274 One of the most robust roadblock genes, *Zfp266*, is recruited to OSK bound regulatory elements through
275 the recognition of adjacent B1 SINEs, where it impedes chromatin opening via its KRAB domain. This
276 probably underlays the accelerated up-regulation of pluripotency gene by OSKM in both *Zfp266* KO and
277 VPR-*Zfp266* overexpression reprogramming. In fact, many of *Zfp266* KO MORs at 72 hours of
278 reprogramming have an open chromatin state and bound by OSK in iPSCs/ESCs, and several of them are
279 associated with pluripotency genes or other genes highly expressed in ESCs, including *Pou5f1, Sall4, Zfp42,*
280 *Klf2, Piwil2, Fbxo15*, *Dnmt3l, Tet1/2* (Supplemental Tables S12). These genes were more efficiently up-
281 regulated in *Zfp266* KO reprogramming compared to the control group (Supplemental Tables S3). Loss of
282 ZFP266 in MEFs lead to only 4 DEGs, suggesting that ZFP266 may not play a significant role in a static state
283 but act as a safeguard against drastic changes of cellular states mediated by newly expressed TFs and/or

284 extracellular ques, such as cytokines. During early embryo development, i.e. 2-8 cell stage, ICM and ESCs,
285 SINEs, particularly B1 SINEs, are enriched in the open chromatin regions, but not in subsequent
286 developmental stages[15]. The data from International Mouse Phenotyping Consortium shows that only 6.5%
287 of pups from *Zfp266* heterozygous intercrosses are homozygous for the KO allele, presenting incomplete
288 penetrance and possible roles of ZFP266 during development while there might be compensation
289 mechanisms (https://www.mousephenotype.org/data/genes/MGI:1924769). Further investigation might
290 reveal roles of ZFP266 in B1 SINE region closing and neighbouring gene regulation during embryo
291 development. B1 SINEs within *Zfp266* KO reprogramming MORs displayed a distinct positioning bias
292 enriched at the flanks of MOR summits, and an orientation bias with head towards the summits. A recent
293 publication identified enrichment of B1 SINEs at the flanks of CD8+ T-cell specific enhancers[43]. Our work
294 suggests that this positioning bias could be a much more general feature of B1-SINE linked regulatory
295 elements and provides evidence that SINEs can affect chromatin states via KRAB-ZFPs. The head-to-tail
296 orientation bias of B1 SINEs is an intriguing novel observation of our work. Stably positioned nucleosomes
297 are highly enriched in SINE and LINE retrotransposons in human[44]. Thus, SINEs may have a further
298 architectural or organisational role of chromatin at the regulatory elements, while it could also be
299 influenced by surrounding DNA sequences and/or other proteins as about 20-25% of B1 SINE had the
300 reversed tail-to-head orientation towards the MOR summits. KLF4 has been shown to bind primate/human
301 specific TEs in naive human ESCs and during reprogramming[13,14]. Interestingly, the *de novo* motif identified
302 in *Zfp266* KO reprogramming MORs, but not Zfp266 KO MEF MORs or ZFP266 DamID-seq peaks, has a full
303 KLF4 binding motif with a base switch at the 5' end of B1 SINE consensus sequence, in addition to a partial
304 KLF4 motif identified in nucleosome enriched KLF4 binding sites[45] (Supplementary Figure S5F). While B1
305 SINEs are restricted to rodents, the enrichment of KLF4 binding in *Zfp266* KO reprogramming MORs with
306 SINEs indicates conserved function of KLF4 to regulate gene expression via TE containing regions in
307 reprogramming/pluripotency across species, which is of clear interest for further investigation. It has been
308 reported that ~2/3 of human KRAB-ZFPs bind to TEs genome-wide, and KRAB-ZFPs supress not only TEs, but
309 also expression of genes nearby[16]. Our results suggest a possibility that other KRAB-ZFPs could act as
310 barriers in different TF-mediated cell conversions or differentiation of pluripotent cells to specific cell types,

311 and therefore elimination of those obstacles might realize more efficient cellular identity changes. Our
312 CRISPR/Cas9-mediated genome-wide KO screening also identified several other novel genes whose
313 inhibition of pluripotency induction would have never been predicted from transcriptomic analyses. Further
314 understanding how these genes hamper OSKM-mediated reprogramming will bring us a better
315 understanding of how to control cellular identities.

316

317 **Methods**

318 **Cell culture**

319 ESCs and MEFs were cultured in ESC medium and in MEF medium, respectively, as described previously[18].
320 Reprogramming was performed in reprogramming medium (ESC medium supplemented with 300 ng ml^{-1} of
321 doxycycline (Sigma) and 10 µg ml^{-1} of L-ascorbic acid or 2-Phospho-L-ascorbic acid trisodium salt (Vitamin C)
322 (Sigma). NSCs were cultured in NSC complete medium consisting of DMEM/F-12 Media, 1:1 Nutrient
323 Mixture (Sigma), 1X N2 supplement (Thermo Fisher Scientific), 1X B27 supplement (Thermo Fisher
324 Scientific), 8 mM glucose (Sigma), 100 U ml^{-1} Pencillin/Streptomycin (ThermoFisher Scientific), 0.001% BSA
325 (ThermoScientific), 0.05 mM β-mercaptoethanol (Thermo Fisher Scientific), supplemented with 10 ng ml^{-1}
326 mouse EGF (Peprotech) and 10 ng ml^{-1} human FGF2 (Peprotech).

327 **Plasmids**

328 Plasmids used in this work are summarized in Supplemental Table S13. The plasmids and their sequences
329 are available upon request.

330 **Generation of Cas9 TNG MKOS ESC line and Cas9 TNG MKOS MEFs**

331 The *Rosa26* targeting vector carrying EF1α-hCas9-ires-neo cassette (Addgene #67987) was electroporated
332 into TNG MKOS ESC line[18,22]. Correct targeting was confirmed by Southern blotting using KpnI and MscI
333 digested genome DNA for a 5' and 3' probe, respectively. The 5' and 3' probes were generated from PCR
334 amplicon using the following primers, 5' forward CAAGTGCTCCATGCTGGAAGGATTG, 5' reverse
335 TGATTGGGGAGGATCCAGATGGAG, 3' forward GGATTGCACGCAGGTTCTCCG 3' reverse

336 CGCCGCCAAGCTCTTCAGCAA and genome DNA (for 5' probe) or the targeting vector (for 3' probe) as a
337 template. Cas9 TNG MKOS MEFs were isolated from E12.5 chimeric embryos generated via morula
338 aggregation and the proportion of transgenic MEFs from each embryo was assessed measuring % of
339 mOrange^{+} cells after exposing 1/10th of the dissociated cells to Dox for 2 days as describe previously[18].

340 **sgRNA screen**

341 The sgRNA library (Addgene #67988) was prepared as describe before[22]. 9 x 10^{6} high contribution (>98%
342 mOrange^{+} 2 days after addition of dox) TNG MKOS Cas9 MEF plated across 90 wells of 6-well culture plates
343 were exposed to lentiviral sgRNA library at MOI=2 for 4 hrs. We used MOI=2 (infection efficiency ~86%) in
344 order to increase coverage of the sgRNA library, presuming the scarcity of reprogramming relevant genes
345 and the negligible probability of the same neutral sgRNAs being repeatedly present in combination with
346 relevant sgRNAs. After viral containing media was removed, the cells were cultured in 3 ml of
347 reprogramming medium. Medium was replaced once 3 ml a day for the first 3 days, and then twice 4 ml a
348 day from day 4 of reprogramming. From day 8, the media was switched to ESC medium supplemented with
349 puromycin (1 µg ml^{-1}) and medium was replenished twice a day with 4 ml / well until day 16. Puromycin
350 resistant, *Nanog*-GFP^{+} cells were then sorted using the FACS AriaII (BD Biosciences) and stored at -80°C as
351 cell pellets before extraction of genomic DNA. Screening was performed in triplicate. Genomic DNA from
352 3x10^{7} sorted GFP^{+} cells was extracted using the Blood & Cell Culture DNA Maxi Kit (Qiagen). Amplification
353 of sgRNA regions from the extracted genome and the original sgRNA plasmid library, and Illumina
354 sequencing was performed as described before[46]. sgRNA read count data was analysed with MAGeCK
355 (version 0.4.4)[23] and genes with enriched and depleted sgRNAs were detected using the test command
356 (default parameters).

357 **Cas9 TNG MKOS MEF reprogramming**

358 0.25 x 10^{4} Cas9 TNG MKOS MEFs were mixed with 9.5 x 10^{4} WT MEFs (129 strain) and seeded in gelatine-
359 coated wells of 6-well plates. Cells were transduced with sgRNAs lentiviruses at an MOI of 3 with 8 µg ml^{-1}
360 Polybrene (Merck-Millipore) for 4 hours and then reprogramming was initiated by addition of
361 reprogramming medium. On day 14, whole well colony images were taken using the Celigo S Cell

362 Cytometer (Nexcelom) and the number of *Nanog*-GFP$^+$ and *Nanog*-GFP$^-$ colonies were counted. The images
363 shown for illustration were stitched using Celigo S Cell Cytometer and processed using ImageJ.

364 ***piggyBac* reprogramming of MEFs with sgRNA expression and/or *Zfp266* cDNAs**

365 *Nanog*-GFP MEFs with or without Cas9 expression from the *Rosa* locus isolated from E12.5 embryos, or wild
366 type MEFs were plated at 1.5x10^5 cells per well in a gelatine-coated 6-well plate. 24 hrs later co-
367 transfection of a Dox-inducible *piggyBac* transposon vector carrying the *tetO-MKOS-ires-mOrange* or *tetO-*
368 *STEMCCA-ires-mOrange* cassette with sgRNA expression cassette, *PB-CA-rtTA* vector with/without carrying
369 a *P2A*-linked *Zfp266* cDNAs, and *pCMV-hyPBase* was performed using 750 ng each DNA and 9 µl of
370 FugeneHD (Promega) as per manufacturer's instructions[18,47,48]. 24 hrs later reprogramming was initiated
371 with reprogramming medium. Medium was changed every 2 days. For colony counting, whole well colony
372 images were taken on day 14 using the Celigo S Cell Cytometer (Nexcelom) and colonies were counted with
373 ImageJ.

374 ***piggyBac* reprogramming of NSCs with sgRNA expression**

375 A GFP sgRNA vector was delivered into Cas9 and GFP expressing NSCs using nucleofection with the SG Cell
376 Line 4DNucleofector X Kit (Lonza) as per manufacturer's instructions[49]. GFP$^-$ NSCs were sorted using the
377 FACS AriaII (BD Biosciences) and plated at clonal density. Individual clones were picked and genotyped to
378 confirm GFP KO. NSCs were reprogrammed by nucleofection of a Dox-inducible *piggyBac* transposon vector
379 carrying the *tetO-MKOS-ires-mOrange* cassette with/without a sgRNA expression cassette, *PB-CA-rtTA*
380 vector and *pCMV-hyPBase*. 2x10^5 NSCs for essential gene expression were nucleofected with 750 ng each of
381 the above-mentioned plasmids using SG Cell Line 4DNucleofector X Kit (Lonza) as per manufacturer's
382 instructions. Cells were then plated on a layer of wild type MEF feeder cells seeded the day before at a
383 density of 1 x10^5 cells per well in a gelatin-coated 6-well plate. One day post-nucleofection, reprogramming
384 was initiated with NSC complete medium supplemented with 100 U ml^{-1} human LIF, 0.3µg ml^{-1} of
385 doxycycline (Sigma) and 10 µg ml^{-1} of L-ascorbic acid or 2-Phospho-L-ascorbic acid trisodium salt (Sigma)
386 (sigma). After 6 days, the medium was switched to serum-free N2B27-based medium (containing
387 DMEM/F12 medium supplemented with N2 combined 1:1 with Neurobasal® medium supplemented with

388 B27; all from Thermo Fisher Scientific), MEK inhibitor (PD0325901, 0.8 µM, Axon Medchem), GSK3b
389 inhibitor (CHIR99021, 3.3 µM, Axon Medchem), 1 µg ml^{-1} of doxycycline (Sigma) and 10 µg ml^{-1} of L-ascorbic
390 acid or 2-Phospho-L-ascorbic acid trisodium salt (sigma). At day 16 of reprogramming, immunofluorescence
391 for NANOG was performed as follows: cells fixed with 4% paraformaldehyde for 10 minutes on day 14 were
392 permeabilized in 0.1% Triton-X in PBS for 1 hours, blocked in 5% BSA in PBS with 0.1% Tween20 for 1 hour
393 at room temperature, and then stained in blocking solution with a primary antibody for NANOG (eBioMLC-
394 51, Thermofisher Scientific) overnight at 4 °C. The next day, an AlexaFluor488 conjugated secondary
395 antibody (A-21208, Invitrogen) was applied in blocking solution for 45 minutes at room temperature before
396 washing and imaging. Whole well images were taken using the Celigo S Cell Cytometer (Nexcelom) and
397 colonies were counted with ImageJ.

398 **CD44, ICAM1, *Nanog*-GFP expression analysis during reprogramming**

399 Cells harvested at different time points of reprogramming were stained in FACS buffer for 30 min at 4°C and
400 washed with FACS buffer prior to acquisition with LSR Fortessa (BD Biosciences) cytometer. The following
401 antibodies from eBioscience were used: ICAM1-biotin (Clone: 13-0541; Dilution: 1/100), CD44-APC (Clone:
402 17-0441; Dilution 1/300), streptavidin-PE-Cy7 (Clone: 25-4317-82; Dilution: 1/1500). Dead cells were
403 excluded using LIVE/DEAD™ Fixable Near-IR Dead Cell Stain Kit (ThermoFisher Scientific, Dilution: 1/1500).
404 Data were analysed using Flowjo v10.

405 **RNA-Seq**

406 **Sample Preparation.** For Wt and *Zfp266* KO MEF samples, 1 x 10^5 Cas9 TNG MKOS MEFs were transduced
407 with either a non-targeting control sgRNA or *Zfp266* sgRNA lentivirus at an MOI of 3 with 8 µg ml^{-1}
408 polybrene (Merck-Millipore) for 4 hours. After additional 96 hours culture in MEF media, the cells were
409 harvested for RNA extraction. For reprogramming samples, 0.25 x 10^4 Cas9 TNG MKOS MEFs were mixed
410 with 9.5 x 10^4 WT MEFs (129 strain) and seeded in gelatine-coated wells of 6-well plates. Cells were
411 transduced with either a non-targeting control sgRNA or *Zfp266* sgRNA lentivirus at an MOI of 3 with 8 µg
412 ml^{-1} Polybrene (Merck-Millipore) for 4 hours, before being recovered for 24 hours in MEF media. After 24
413 hours reprogramming was initiated by addition of reprogramming medium. Cells were harvested at day 3,

414 day 5 and day 7 of reprogramming, respectively, and 1 x 10^5 of mOrange$^+$ OSKM expressing cells were
415 sorted with the FACS AriaII (BD Biosciences) per sample. *Nanog*-GFP$^+$ iPSCs were harvested at day 15, and
416 sorted with the FACS AriaII (BD Biosciences) into 96-well plates. Sorted iPSCs were cultured in ESC medium
417 with puromycin (1 µg ml^{-1}) to select for transgene independent clones and KO of *Zfp266* was confirmed by
418 genotyping. *Zfp266* KO ESCs were generated by transfecting Clone J ESCs with a *Zfp266* sgRNA plasmid
419 expressing BFP. Single BFP$^+$ ESCs were then sorted with the FACS AriaII (BD Biosciences) into 96-well plates
420 48 hours after transfection. Clones which became BFP- negative (i.e. shed the sgRNA plasmid) were
421 selected and KO of *Zfp266* was confirmed by genotyping. 1 x 10^5 iPSCs or ESCs were used for RNA
422 preparation. Cells were homogenised with the QIAshredder kit (Qiagen) and total RNA was extracted from
423 all samples using the RNeasy Plus Micro Kit (Qiagen). Libraries were prepped with the NEB Ultra II stranded
424 mRNA Library prep kit (NEB). RNA-Seq libraries were sequenced with NextSeq, 75SE.

425 **Read processing.** For each sequencing run, a quality control report was generated using FastQC and
426 Illumina TruSeq adapter sequences were removed using Cutadapt[50]. Sequencing runs from the same
427 biological sample were then concatenated and mapped to the GRCm38 reference genome using STAR[51].

428 **Differential analysis**. For each biological sample, aligned sequencing reads were first assigned to genomic
429 features (e.g., genes) using Rsubread[52] and a count table was generated. Differential expression analysis
430 was then performed with DESeq2[53], and statistically significant genes (e.g., FDR < 0.05 and
431 log2FoldChange > 1) were identified using the standard workflow. Importantly, although the data
432 represents a control and treatment time-series experiment, we opted to combine the factors of interest
433 into a single factor for easier comprehension. Gene ontology analysis for differentially expressed genes was
434 performed using the goseq package[54].

435 **Downstream analysis**. For exploratory analysis and visualization, a batch-corrected and regularized log
436 matrix of expression values was used. The count table was first transformed to stabilize the variance across
437 the mean using the rlog function from DESeq2 and then unwanted batch effects (e.g., library preparation
438 date) were removed using the removeBatchEffect function from limma[55].

439 **DamID-seq**

440 **Sample Preparation.** 1 x 10^5 WT MEFs (129 strain) were nucleofected with either *PGK-mO-Dam* or *PGK-mO-*
441 *Dam-Zfp266* plasmids using the P4 Primary Cell 4D-Nucleofector X Kit (Lonza). 5 replicates were performed
442 in total. Cells were recovered in MEF media for 48 hours before 3 x 10^4 - 1.6 x 10^5 GFP+ cells per sample
443 were sorted with the FACS AriaII (BD Biosciences). Genomic DNA was isolated with Quick-gDNA™
444 MicroPrep (ZymoResearch) and 32 ng genomic DNA/sample was used for DamID-seq library preparation as
445 previously described[37]DamID libraries were sequenced with NextSeq, 40PE.

446 **Read processing.** For each sequencing run, a quality control report was generated using FastQC and
447 Illumina Nextera adapter sequences were removed using Cutadapt. Sequencing runs from the same
448 biological sample were then concatenated and mapped to the GRCm38 reference genome using BWA[56].
449 Uninformative and spurious alignments were subsequently filtered using a combination of SAMtools[57] and
450 BEDtools[58] commands. Specifically, reads mapped to the mitochondrial chromosome and reads mapped to
451 blacklisted regions were filtered.

452 **Peak calling.** For each biological sample, aligned sequencing reads were assigned to genomic features (e.g.,
453 DpnII restriction fragments) using Rsubread and a count table was generated. Statistically significant
454 regions of Dam-fusion protein binding (e.g., FDR < 0.05 and log2FoldChange > 1) were detected using the
455 callPeak command from Daim[37]. For further details, please refer to the original manuscript describing the
456 Daim software[37]. The regions were then annotated and analysed for gene and genome ontology
457 enrichment using the annotatePeaks command from HOMER[59].

458 **Downstream analysis.** Heatmaps of read coverage at Dam-fusion binding regions were produced using the
459 computeMatrix and plotHeatmap commands from deepTools[60]. When plotting heatmaps, a total of 5 peaks
460 identified exactly over *Zfp266* exons (chr9:20495068-20521417) were removed from the *Zfp266* DamID
461 peak regions due to the high signal intensity caused by the *PGK-mO-Dam-Zfp266* plasmid. *De novo* motif
462 discovery and was performed using the MEME-ChIP tool from the MEME suite (version 5.1.1)[61]. Motif
463 enrichment analysis was performed using findMotifsGenome command from HOMER[59] as DamID-seq's
464 large perk size was not optimal for the MEME-ChIP tool. Genome browser images of peak regions and read
465 coverage were composed using the Integrative Genomics Viewer[62].

466 **ATAC-seq**

467 **Sample Preparation.** Cas9 TNG MKOS MEFs were thawed, plated and transduced in the same manner as
468 samples prepared for RNA-Seq. After 24 hours reprogramming was initiated by addition of reprogramming
469 medium for reprogramming samples, while MEF samples were maintained in MEF media. Cells were
470 harvested 96 hours after sgRNA transduction (which was 72 hours after OSKM induction for reprogramming
471 samples) and sorted with the FACS AriaII (BD Biosciences). Cells were then processed for ATAC-Sequencing
472 according to the Omni-ATAC protocol[63]. Briefly, 5 x 10^4 sorted Cas9 TNG MKOS MEFs per sample were
473 washed with cold 1x PBS then pelleted before the supernatant was discarded. Cell pellets were then gently
474 resuspended in 50 μl of lysis buffer (48.5 μl resuspension buffer, 0.5 μl 10% NP-40 (Sigma) 0.5 μl 10%
475 Tween-20 (Sigma), 0.5 μl 1% Digitonin (Promega) (resuspension buffer: 500 μl 1M Tris-HCl, pH7.5
476 (ThermoFisher), 100 μl 5M NaCl (Sigma), 150 μl 1M $MgCl_2$ (Sigma), 49.25 ml nuclease-free H_2O) and
477 incubated on ice for 3 minutes. Then, 1 ml of wash buffer (990 μl resuspension buffer, 10 μl Tween-20
478 (Sigma)) was added to the tubes before they were gently inverted and then centrifuged for 10 minutes at
479 500 x g, at 4 °C. Supernatants were then carefully aspirated. Nuclei pellets were then resuspended in 50 μl
480 of transposition mix (2.5 μl Tn5 transposase, 25 μl 2x TD buffer (both Illumina), 0.5 μl 1% Digitonin
481 (Promega), 0.5 μl 10% Tween-20 (Sigma), 16.5 μl 1x PBS, 5μl nuclease-free H_2O) and incubated in a
482 thermomixer at 37 °C, 1000 rpm for 30 minutes. Transposed DNA was then purified with the Zymo DNA
483 Clean and Concentrator-5 Kit (Zymo Research) and eluted in 21 μl nuclease-free H_2O. All purified DNA (~20
484 μl) was used for PCR amplification with NEBNext High Fidelity 2x MasterMix (NEB) and optimum cycle
485 number was determined by qPCR, as per the protocol. Amplified DNA was then purified with double-sided
486 bead purification using AMPure XP magnetic beads (Beckman Coulter). Library concentration was
487 determined with Qubit (ThermoFisher) and fragment size/quality with TapeStation (Agilent). ATAC libraries
488 were sequenced with NextSeq, 40PE.

489 **Read processing.** For each sequencing run, a quality control report was generated using FastQC and
490 Illumina Nextera adapter sequences were removed using Cutadapt[50]. Sequencing runs from the same
491 biological sample were then concatenated and mapped to the GRCm38 reference genome using BWA[56].

492 Duplicate reads caused by PCR amplification were subsequently identified using the MarkDuplicates
493 command from Picard (https://broadinstitute.github.io/picard/). Uninformative and spurious alignments
494 were next filtered using a combination of SAMtools[57] and BEDtools[58] commands. Specifically, reads mapped
495 to the mitochondrial chromosome, reads mapped to blacklisted regions, reads marked as duplicates, and
496 reads not properly paired (e.g., reads that aren't FR orientation or with an insert size greater than 2 kb)
497 were filtered.

498 **Peak calling.** For each biological sample, statistically significant regions of chromatin accessibility (e.g., FDR
499 < 0.1) were detected using the callpeak command from MACS2[64] (https://github.com/macs3-
500 project/MACS). For downstream analyses, a consensus set of peaks was created by taking the union across
501 all biological samples with the multiinter command from BEDtools[58].

502 **Differential analysis.** For each biological sample, aligned sequencing reads were first assigned to genomic
503 features (e.g., consensus set of peaks) using Rsubread[52] and a count table was generated. Differential
504 accessibility analysis was then performed with DESeq2[65] and statistically significant peaks (e.g., FDR < 0.05
505 and log2FoldChange > 1) were identified using the standard workflow.

506 **Downstream analysis.** Heatmaps of read coverage at chromatin accessibility regions were produced using
507 the computeMatrix and plotHeatmap commands from deepTools[60]. K-means clustering was used to
508 partition the regions into two distinct categories of reprogramming MORs. Genome browser images of
509 peak regions and read coverage were composed using the Integrative Genomics Viewer[62]. Peaks were
510 annotated against mm10 with annotatePeaks.pl from the HOMER suite (version 4.11)[59]. *De novo* motif
511 discovery and enrichment analysis of MORs were performed using the *Zfp266* KO samples' narrowpeak
512 summits within MORs with the MEME-ChIP tool from the MEME suite (version 5.1.1)[61]. The number of SINE
513 elements around peaks were counted using the BEDTools window command in a window of ±500 bp from
514 the summits of the peaks. ATAC-seq data of iPSCs were retrieved from GSE98124[39].

515 **Luciferase Reporter Assays**

516 The pGL3 reporter plasmid containing the SV40 early promoter (Promega) was used for all luciferase
517 reporter assays along with an internal control Renilla plasmid (Promega). Luciferase activity was measured

518 with the GloMax 96-microplate luminometer (Promega) using the Dual-Glo Luciferase Assay System
519 (Promega). For assays performed in HEK293 cells, 0.5-1x10^4 cells were plated per well of a 96-well plate 24
520 hours prior to transfection. Transfection mixes were prepared as follows; 100 ng pGL3 reporter plasmid, 0.5
521 ng Renilla plasmid and 100 ng overexpression plasmid (BFP/VPR/VPR-*Zfp266*/Wt *Zfp266*) were mixed in
522 Opti-MEM I Reduced Serum Medium (Gibco) up to 100μl. Fugene HD Transfection Reagent (Promega) was
523 then added at a ratio of 3:1 (reagent:DNA) and 5-10 μl was added to each well of cells. Luciferase activity
524 was measured 48 hours after transfection. For assays performed in MEF/reprogramming cells, 1x10^4 TNG
525 MKOS MEFs were plated per well of a 96-well plate 24 hours prior to transfection, either in MEF media or
526 ES media +dox (300 ng/ml) to induce OSKM expression. Transfection mixes were prepared as such; 1 μg
527 pGL3 reporter plasmid, 10 ng Renilla plasmid were mixed in Opti-MEM I Reduced Serum Medium up to 100
528 μl. Fugene HD Transfection Reagent was then added at a ratio of 4:1 (reagent:DNA) and 20μl was added to
529 each well of MEFs/reprogramming cells. Luciferase activity was measured 48 hours after transfection.

530

531 **END NOTE**

532 **Acknowledgements**

533 We thank I. Chambers for providing TNG ESC line, F. Rossi and C. Cryer for assistance with flow cytometry,
534 Biomed unit staff for mouse husbandry, the Wellcome Sanger Institute sequencing facility for gRNA
535 sequencing, EMBL GeneCore for RNA-seq, ATAC-seq and DamID-seq, A. Soufi and members of the Kaji and
536 Soufi labs for comments on the manuscript. Some of the computations for this work were enabled by
537 resources in project SNIC 2017/7-317 provided by the Swedish National Infrastructure for Computing (SNIC)
538 at the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). This work was
539 supported by European Research Council (ERC) grants ROADTOIPS (no. 261075) and MRC senior non-
540 clinical fellowship (MR/N008715/1) funded for K.K. We also thank the generous support from Baillie Gifford
541 for the collaboration between CiRA and MRC CRM, from Japan Agency for Medical Research and
542 Development (AMED) for CiRA. K.Y. was supported by the Wellcome Trust (206194). D.F.K., J.A. and M.Y.
543 was supported by the BBSRC (EASTBIO doctoral training partnership), Principal's Career Development

544 scholarship from the University of Edinburgh, and Japan Society for the Promotion of Science (JSPS)
545 Overseas Research Fellowships, respectively. V.O. and M.Bertenstam were supported by the Swedish
546 Foundation for Strategic Research (A3 04 159p). V.O. was also supported by the Swedish Research Council
547 (Vr 621-2008-3074).

548 **Author Contribution**

549 D.F.K. designed and performed sgRNA screen, validation and characterization of the roadblock genes. M.A,
550 J.A., S.K. and S.T. contributed to the analyses of the gRNA sequencing, RNA-seq, ATAC-seq, ChIP-seq and
551 DamID-seq data sets. M.B. and S.Z. provided technical support. M.Bertenstam and V.O. generated the
552 screening data website. K.O. advised and supervised D.F.K. K.Y. provided the gRNA library, the Rosa26-Cas9
553 targeting vector, and advised on the screen. K.K. conceived the study, supervised experiment design and
554 data interpretation, and wrote the manuscript with D.F.K.

555 **Author Information**

556 The authors declare no competing financial interests. Correspondence and requests for materials should be
557 address to K.K. (keisuke.kaji@ed.ac.uk).

558

559

560 **References**

561 1. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult
562 Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).

563 2. Zhao, Y. *et al.* Two Supporting Factors Greatly Improve the Efficiency of Human iPSC Generation. *Cell*
564 *Stem Cell* **3**, 475–479 (2008).

565 3. Banito, A. *et al.* Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev.*
566 **23**, 2134–2139 (2009).

567 4. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway.

568 *Nature* **460**, 1132–1135 (2009).

569 5. Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming.
570 *Nature* **460**, 1140–1144 (2009).

571 6. Li, H. *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460**, 1136–1139 (2009).

572 7. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell
573 genomic integrity. *Nature* **460**, 1149–1153 (2009).

574 8. Utikal, J. *et al.* Immortalization eliminates a roadblock during cellular reprogramming into iPS cells.
575 *Nature* **460**, 1145–1148 (2009).

576 9. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483**, 598–
577 602 (2012).

578 10. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and Impediments of the Pluripotency
579 Reprogramming Factors' Initial Engagement with the Genome. *Cell* **151**, 994–1004 (2012).

580 11. Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol. 2018 191*
581 **19**, 1–12 (2018).

582 12. Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity.
583 *Cell* **174**, 391-405.e19 (2018).

584 13. Ohnuki, M. *et al.* Dynamic regulation of human endogenous retroviruses mediates factor-induced
585 reprogramming and differentiation potential. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12426–12431
586 (2014).

587 14. Pontis, J. *et al.* Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic
588 Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* **24**, 724-735.e5
589 (2019).

590 15. Lu, J. Y. *et al.* Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated
591 Regulation. *Cell Rep.* **30**, 3296-3311.e5 (2020).

592 16. Yang, P., Wang, Y. & Macfarlan, T. S. The Role of KRAB-ZFPs in Transposable Element Repression and
593 Mammalian Evolution. *Trends Genet.* **33**, 871–881 (2017).

594 17. Wolf, G. *et al.* Krab-zinc finger protein gene expansion in response to active retrotransposons in the
595 murine lineage. *Elife* **9**, 1–22 (2020).

596 18. Chantzoura, E. *et al.* Reprogramming Roadblocks Are System Dependent. *Stem Cell Reports* **5**, 350–
597 364 (2015).

598 19. Ruetz, T. *et al.* Constitutively Active SMAD2/3 Are Broad-Scope Potentiators of Transcription-Factor-
599 Mediated Cellular Reprogramming. *Cell Stem Cell* **21**, 791-805.e9 (2017).

600 20. Kareta, M. S. *et al.* Inhibition of Pluripotency Networks by the Rb Tumor Suppressor Restricts
601 Reprogramming and Tumorigenesis. *Cell Stem Cell* **16**, 39–50 (2015).

602 21. Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic
603 reprogramming. *Nature* **488**, 409–413 (2012).

604 22. Tzelepis, K. *et al.* A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets
605 in Acute Myeloid Leukemia. *Cell Rep.* **17**, 1193–1205 (2016).

606 23. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale
607 CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).

608 24. Liu, J. *et al.* The oncogene c-Jun impedes somatic cell reprogramming. *Nat. Cell Biol.* **17**, 856–67
609 (2015).

610 25. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nat. 2012*
611 *4837391* **483**, 598–602 (2012).

612 26. Yang, C. S., Chang, K. Y. & Rana, T. M. Genome-wide Functional Analysis Reveals Factors Needed at
613 the Transition Steps of Induced Reprogramming. *Cell Rep.* **8**, 327–337 (2014).

614 27. Michlits, G. *et al.* CRISPR-UMI: Single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat.*
615 *Methods* **14**, 1191–1197 (2017).

616 28. O'Malley, J. *et al.* High-resolution analysis with novel cell-surface markers identifies routes to iPS
617 cells. *Nat. 2013 4997456* **499**, 88–91 (2013).

618 29. Sommer, C. A. *et al.* Induced Pluripotent Stem Cell Generation Using a Single Lentiviral Stem Cell
619 Cassette. *Stem Cells* **27**, 543–549 (2009).

620 30. Kaji, K. *et al.* Virus-free induction of pluripotency and subsequent excision of reprogramming factors.
621 *Nature* **458**, 771–775 (2009).

622 31. Reinhardt, A., Kagawa, H. & Woltjen, K. N-Terminal Amino Acids Determine KLF4 Protein Stability in
623 2A Peptide-Linked Polycistronic Reprogramming Constructs. *Stem Cell Reports* **14**, 520–527 (2020).

624 32. Kim, S. Il *et al.* KLF4 N-terminal variance modulates induced reprogramming to pluripotency. *Stem*
625 *Cell Reports* **4**, 727–743 (2015).

626 33. O'Malley, J. *et al.* High-resolution analysis with novel cell-surface markers identifies routes to iPS
627 cells. *Nature* **499**, 88–91 (2013).

628 34. Ecco, G., Imbeault, M. & Trono, D. KRAB zinc finger proteins. *Development* **144**, 2719–2729 (2017).

629 35. Margolin, J. F. *et al.* Krüppel-associated boxes are potent transcriptional repression domains. *Proc.*
630 *Natl. Acad. Sci.* **91**, 4509–4513 (1994).

631 36. Peng, H., Ivanov, A. V., Oh, H. J., Lau, Y. F. C. & Rauscher, F. J. Epigenetic Gene Silencing by the SRY
632 Protein Is Mediated by a KRAB-O Protein That Recruits the KAP1 Co-repressor Machinery. *J. Biol.*
633 *Chem.* **284**, 35670–35680 (2009).

634 37. Tosti, L. *et al.* Mapping transcription factor occupancy using minimal numbers of cells in vitro and in
635 vivo. *Genome Res.* **28**, 592–605 (2018).

636 38. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein–DNA interactions using
637 DamID in mammalian cells. *Nat. Protoc. 2007 26* **2**, 1467–1478 (2007).

638 39. Benchetrit, H. *et al.* Direct Induction of the Three Pre-implantation Blastocyst Cell Types from
639 Fibroblasts. *Cell Stem Cell* **24**, 983-994.e7 (2019).

640 40. Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*
641 **168**, 442-459.e20 (2017).

642 41. Roberts, G. A. *et al.* Dissecting OCT4 defines the role of nucleosome binding in pluripotency. *Nat.*
643 *Cell Biol. 2021 238* **23**, 834–845 (2021).

644 42. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods 2015*
645 *124* **12**, 326–328 (2015).

646 43. Ye, M. *et al.* Specific subfamilies of transposable elements contribute to different domains of T
647 lymphocyte enhancers. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7905–7916 (2020).

648 44. Li, C. & Luscombe, N. M. Nucleosome positioning stability is a modulator of germline mutation rate
649 variation across the human genome. *Nat. Commun. 2020 111* **11**, 1–13 (2020).

650 45. Soufi, A. *et al.* Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate
651 Reprogramming. *Cell* **161**, 555–568 (2015).

652 46. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic
653 screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol. 2013 323*
654 **32**, 267–273 (2013).

655 47. Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells.
656 *Nat. 2009 4587239* **458**, 766–770 (2009).

657 48. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for
658 mammalian applications. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1531–1536 (2011).

659 49. Bressan, R. B. *et al.* Efficient CRISPR/cas9-assisted gene targeting enables rapid and precise genetic
660 manipulation of mammalian neural stem cells. *Dev.* **144**, 635–648 (2017).

661 50. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
662 *EMBnet.journal* **17**, 10–12 (2011).

663 51. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

664 52. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for
665 alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47–e47 (2019).

666 53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq
667 data with DESeq2. *Genome Biol.* **15**, 550 (2014).

668 54. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq:
669 accounting for selection bias. *Genome Biol.* **11**, 1–12 (2010).

670 55. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and
671 microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).

672 56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform.
673 *Bioinformatics* **25**, 1754–1760 (2009).

674 57. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
675 (2009).

676 58. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.
677 *Bioinformatics* **26**, 841–842 (2010).

678 59. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-
679 Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).

680 60. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis.
681 *Nucleic Acids Res.* **44**, W160–W165 (2016).

682 61. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–
683 W49 (2015).

684 62. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol. 2011 291* **29**, 24–26 (2011).

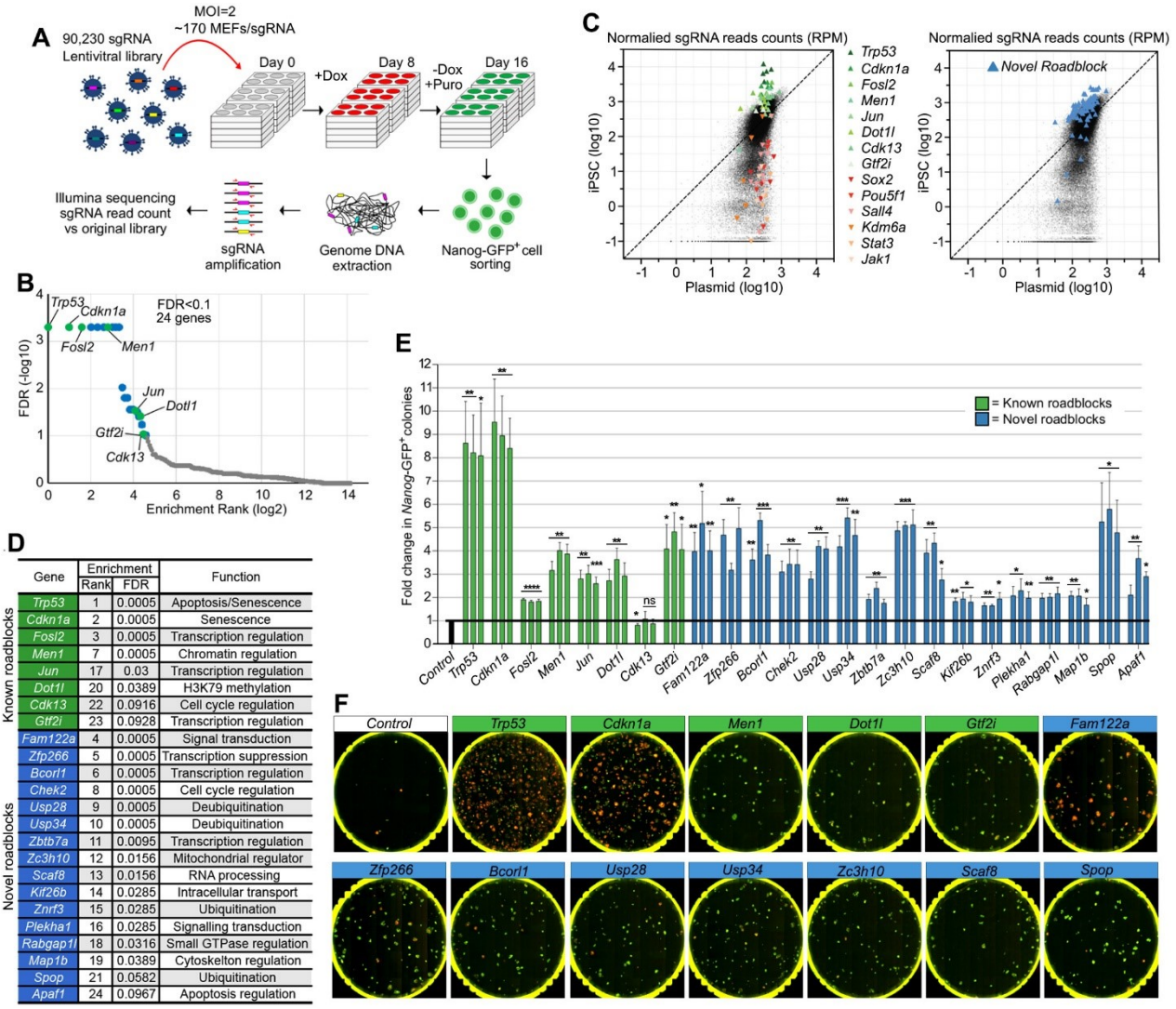
685 63. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation
686 of frozen tissues. *Nat. Methods 2017 1410* **14**, 959–962 (2017).

687 64. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, 1–9 (2008).

688 65. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq
689 data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

690

691



D

	Gene	Enrichment Rank	Enrichment FDR	Function
Known roadblocks	Trp53	1	0.0005	Apoptosis/Senescence
	Cdkn1a	2	0.0005	Senescence
	Fosl2	3	0.0005	Transcription regulation
	Men1	7	0.0005	Chromatin regulation
	Jun	17	0.03	Transcription regulation
	Dot1l	20	0.0389	H3K79 methylation
	Cdk13	22	0.0916	Cell cycle regulation
	Gtf2i	23	0.0928	Transcription regulation
Novel roadblocks	Fam122a	4	0.0005	Signal transduction
	Zfp266	5	0.0005	Transcription suppression
	Bcorl1	6	0.0005	Transcription regulation
	Chek2	8	0.0005	Cell cycle regulation
	Usp28	9	0.0005	Deubiquitination
	Usp34	10	0.0005	Deubiquitination
	Zbtb7a	11	0.0095	Transcription regulation
	Zc3h10	12	0.0156	Mitochondrial regulator
	Scaf8	13	0.0156	RNA processing
	Kif26b	14	0.0285	Intracellular transport
	Znrf3	15	0.0285	Ubiquitination
	Plekha1	16	0.0285	Signalling transduction
	Rabgap1l	18	0.0316	Small GTPase regulation
	Map1b	19	0.0389	Cytoskelton regulation
	Spop	21	0.0582	Ubiquitination
	Apaf1	24	0.0967	Apoptosis regulation

Figure. 1. A genome-wide CRISPR screen identifies reprogramming roadblock genes. A. Schematic diagram of the screening strategy. sgRNA library infected Cas9 TNG MKOS MEFs were cultured in +dox for 8 days then in -dox and +Puro for additional 8 days. Integrated sgRNAs were amplified from *Nanog*-GFP$^+$ cells for Illumina sequencing. **B.** Enrichment FDR ranking with MAGeCK. 24 genes, including 8 previously reported (green) and 16 novel (blue) roadblock genes, were identified using a cut-off of FDR<0.1. **C**. Normalized sgRNA read counts in the initial plasmid library versus mutant iPSC pool. sgRNAs against previously reported roadblock genes (red/orange) and genes essential for reprogramming (green) exhibited expected enrichment/depletion respectively (left). sgRNAs against 16 novel roadblock genes identified in this screen are highlighted in blue (right). **D**. Enrichment rank, FDR, and function of the 24 discovered roadblock genes. **E**. Validation of the screen result with 3 individual sgRNAs per gene. This graph is a summary of 5 data sets shown in Supplemental Figure S2A. Error bars indicate SEM, ****p<0.0001, ***p<0.001, **p<0.01,*p<0.05 based on an unpaired two-tailed t-test **F**. Representative whole-well images of KO reprogramming of 13 top roadblocks from E. Previously reported and novel roadblock genes were labelled in green and and blue, respectively. Red; mOrange, Green; *Nanog*-GFP.

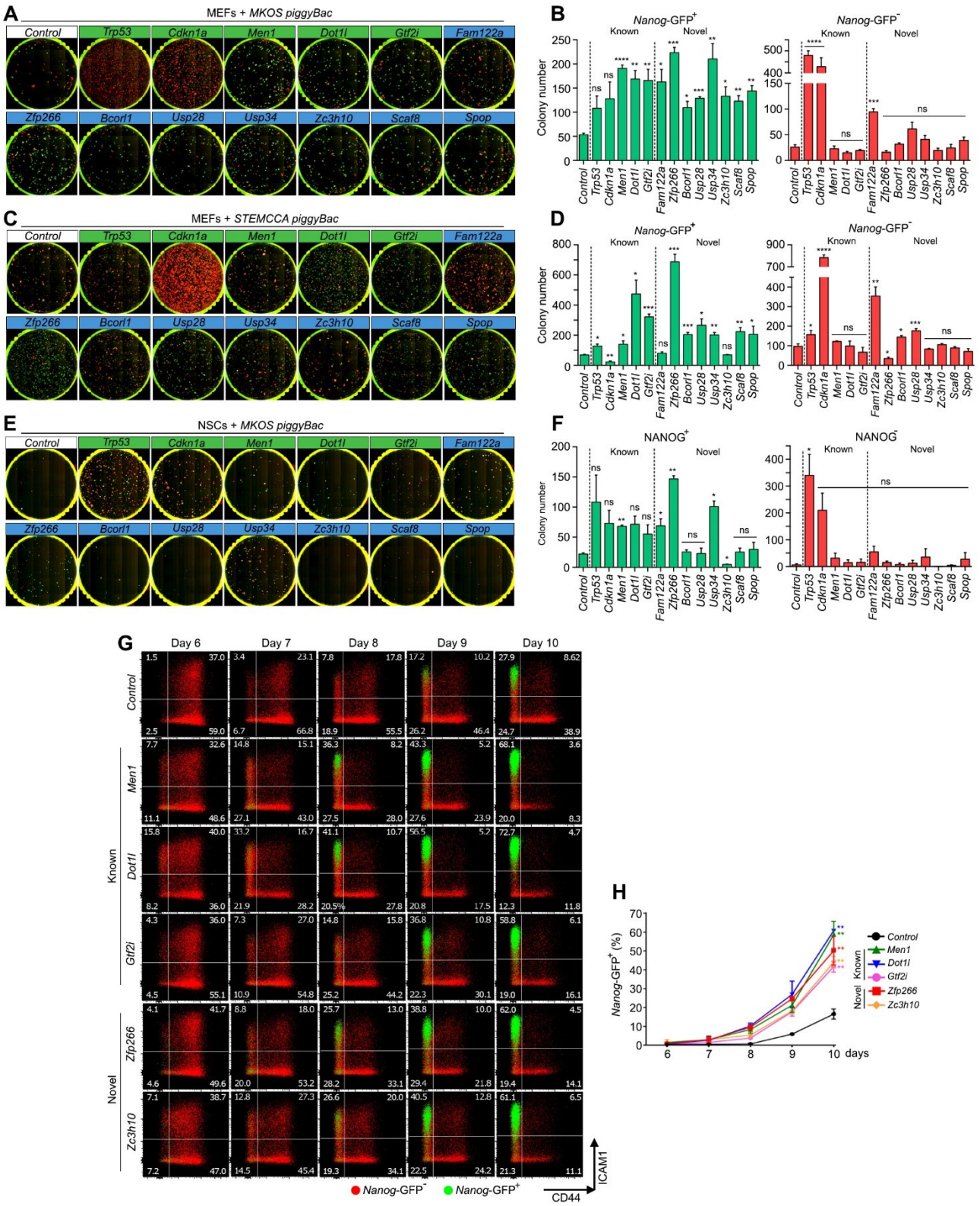


Figure 2. Characterization of the roadblock gene KO in different reprogramming systems and kinetics. A, C. *Cas9* expressing *Nanog*-GFP MEF reprogramming with MKOS (A), STEMCCA (C) *piggyBac* transposons with sgRNA expression at day 15. Red; mOrange, Green; *Nanog*-GFP. **B,D.** *Nanog*-GFP$^+$ and *Nanog*-GFP$^-$ mean colony numbers from A and C. **E.** *Cas9* expressing NSC reprogramming with MKOS *piggyBac* transposons with sgRNA expression at day 15. Green; immunofluorescence for NANOG. **F.** NANOG+ and NANOG- mean colony numbers from E. **G.** Accelerated CD44/ICAM/*Nanog*-GFP expression changes by sgRNA expression against the roadblock genes (n=2). Red; *Nanog*-GFP- cells, Green; *Nanog*-GFP+ cells. **H.** Quantification of *Nanog*-GFP+ cells from day 6 to 10 of reprogramming. The graph represents an average of

716 2 independent experiments. For all graphs error bars indicate SEM, ****$p<0.0001$, ***$p<0.001$,
717 **$p<0.01$,*$p<0.05$ based on an unpaired two tailed t test.

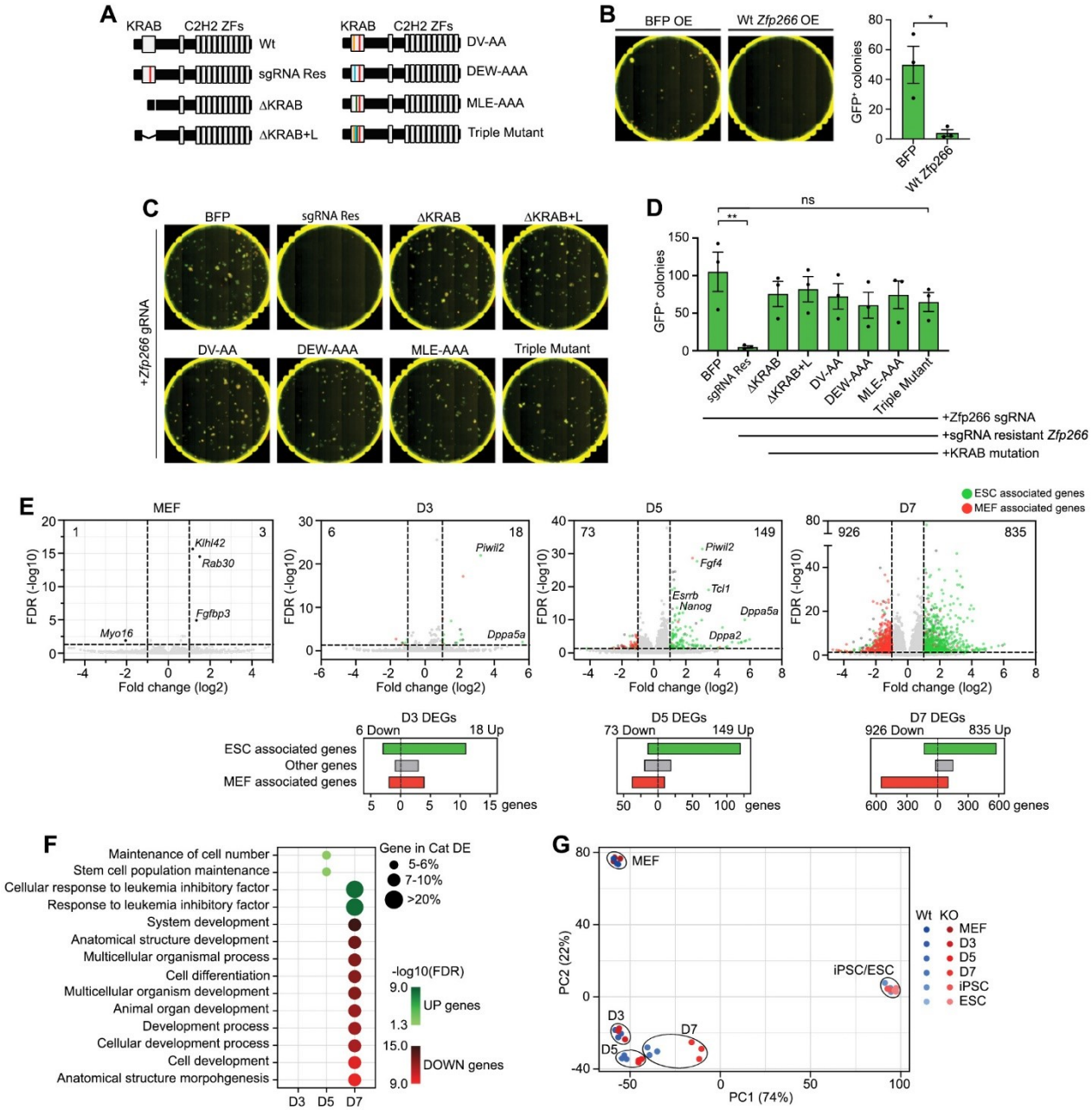


Figure 3. ZFP266 impedes activation of pluripotency genes via its KRAB domain. A. Diagram of *Zfp266* Wt and mutants. A red bar indicates a silent mutation that confers sgRNA resistance. KRAB domain deletion mutants with (ΔKRAB+L) and without a linker (ΔKRAB) do not have the sgRNA target sequence. DV-AA, DEW-AAA, MLE-AAA mutants have alanine substitutions in the indicated critical amino acids in the KRAB domain[28,29]. Triple mutant contains all the alanine substitutions. **B.** *Nanog*-GFP MEF reprogramming with MKOS *piggyBac* transposons and BFP or Wt *Zfp266* cDNA overexpression, imaged at day 15. Red; mOrange, Green; *Nanog*-GFP. Error bars indicate SEM, *p<0.05 based on an unpaired two-tailed t-test. **C.** *Cas9 Nanog*-GFP MEF reprogramming with MKOS *piggyBac* transposons, *Zfp266* sgRNA expression as well as BFP, Wt *Zfp266, or* mutant *Zfp266* cDNA overexpression, imaged at day 15. Red; mOrange, Green; *Nanog*-GFP. **D**. Mean *Nanog*-GFP$^+$ colony numbers of C. Error bars indicate SEM, **p<0.01 based on a one-way ANOVA test. **E.** RNA-Seq volcano plot of *Zfp266* KO vs Wt MEF, day 3, 5 and 7 of reprogramming. Up-regulated and down-regulated genes in KO cells are shown to the right and left of the plot, respectively (cut-off FDR<0.05, fold change >2). ESC- and MEF-associated genes (FDR<0.05, fold change >2 in ESCs vs MEFs) are highlighted in green and red. Graphs below volcano plots show the number of ESC-associated, MEF-associated and other genes within D3, D5, D7 reprogramming differentially expressed genes (DEGs). **F.** GO enrichment analysis of upregulated and downregulated genes in *Zfp266* KO reprogramming. **G.** Principal component analysis of *Zfp266* Wt and KO RNA-Seq samples. Blue dots indicate Wt samples, red dots indicate KO samples, three samples per timepoint.

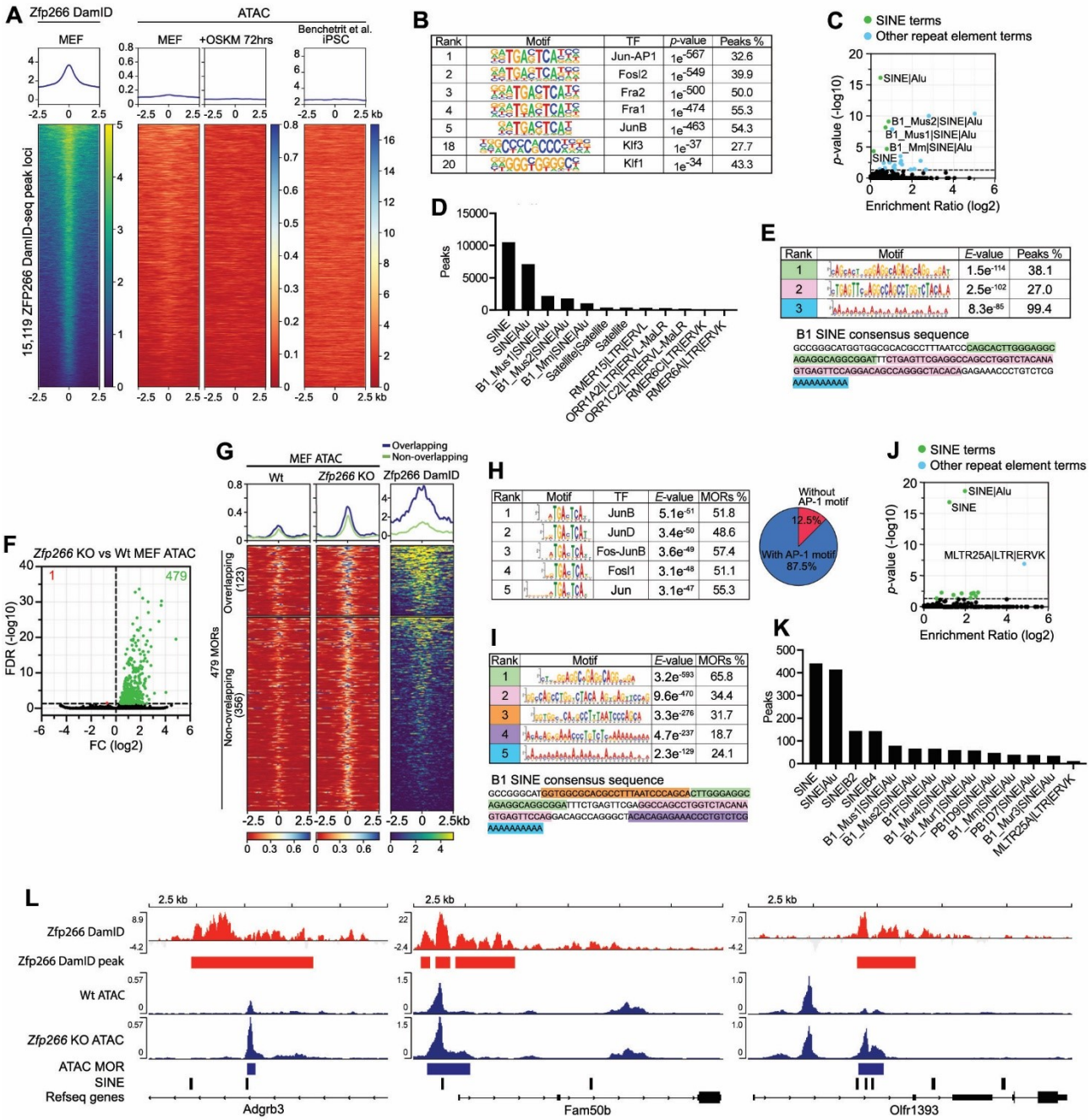


Figure 4. *Zfp266* KO in MEFs results in chromatin opening at the SINE-containing ZFP266 binding sites. A. ZFP266 DamID-seq signals in MEFs, ATAC-seq signals in MEFs, +72hours of reprogramming, and iPSCs at the ZFP266 DamID-seq peak loci. **B.** Motif enrichment analysis with HOMER on ZFP266 DamID-seq peaks. **C.** Significance and fold enrichment ratio of transposable element (TE) families overlap with Dam-ZFP266 peaks. Green dots indicate significantly enriched SINEs, blue dots indicate other significantly enriched TEs. **D.** Number of ZFP266 DamID-seq peaks that overlap with TEs. **E.** *De novo* motif discovery analysis with MEME on ZFP266 DamID-seq peaks. The identified motifs correspond to parts of the B1 SINE consensus sequence, indicated by matching colours. **F.** Volcano plot of *Zfp266* KO vs Wt MEF ATAC-seq. Green and red dots indicate more open regions (MORs) and more closed regions in *Zfp266* KO MEFs, respectively (FDR<0.05). **G.** ATAC-seq and ZFP266 DamID-seq signals in the *Zfp266* KO MEF MORs, overlapped (top) and non-overlapped (bottom) with ZFP266 DamID peaks in MEFs. **H.** Motif enrichment analysis on *Zfp266* KO MEF MORs. Percentages of MORs containing each motif and AP-1 motif are indicated. **I.** *De novo* motif discovery analysis on *Zfp266* KO MEF MORs. The top five most significant motifs correspond to parts of the B1 SINE consensus sequence, indicated by matching colours. **J.** Significance and fold enrichment ratio of transposable element (TE) families overlap with Zfp266 KO MEF MORs. Green dots indicate significantly enriched SINEs, blue dots indicate other significantly enriched TEs. **K.** Number of *Zfp266* KO MEF MORs that

754 overlap with TEs. **L.** Examples of *Zfp266* KO MEF MORs (Blue) with SINE (black), overlapping with ZFP266
755 DamID-seq peaks (red).

756

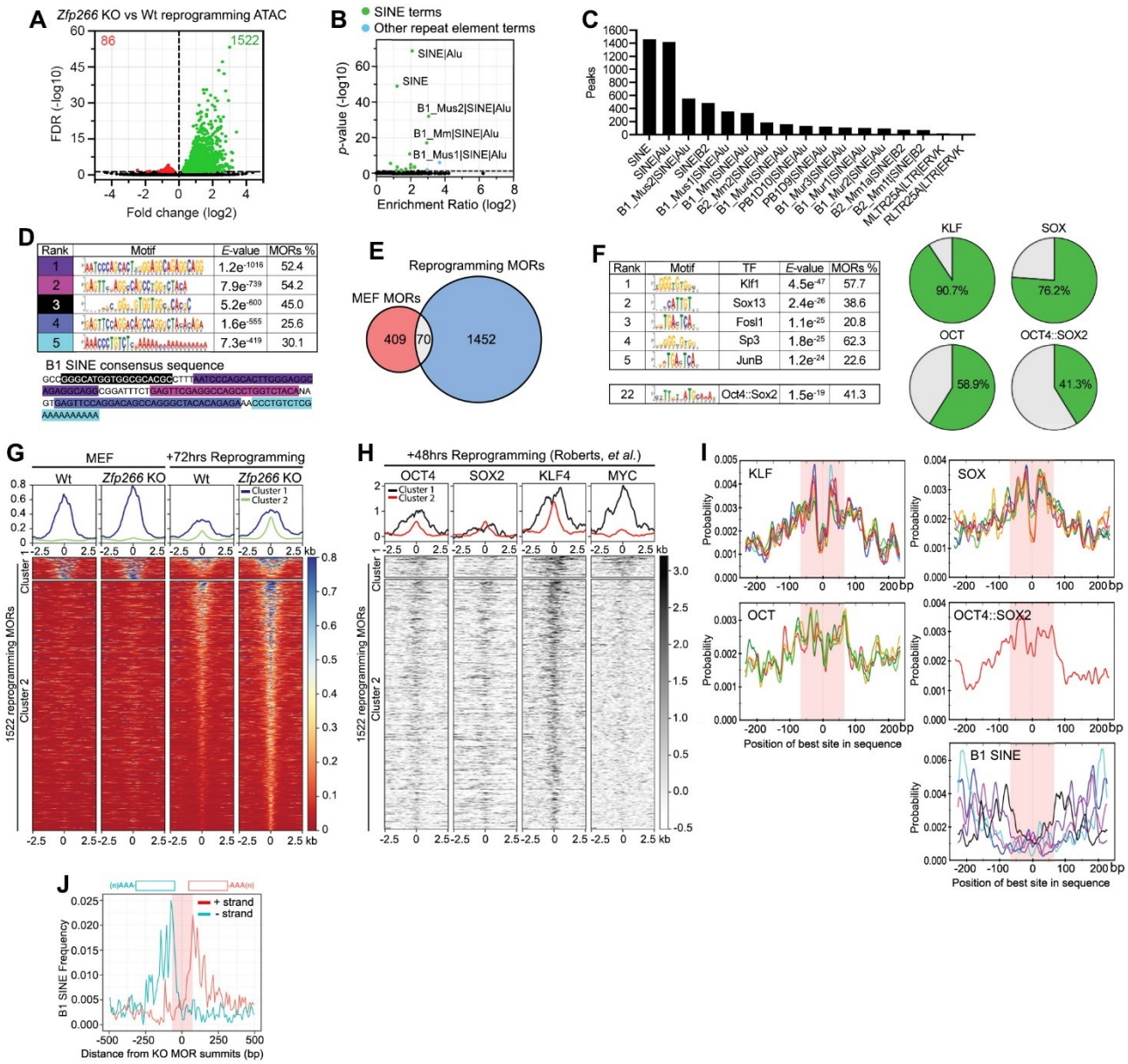
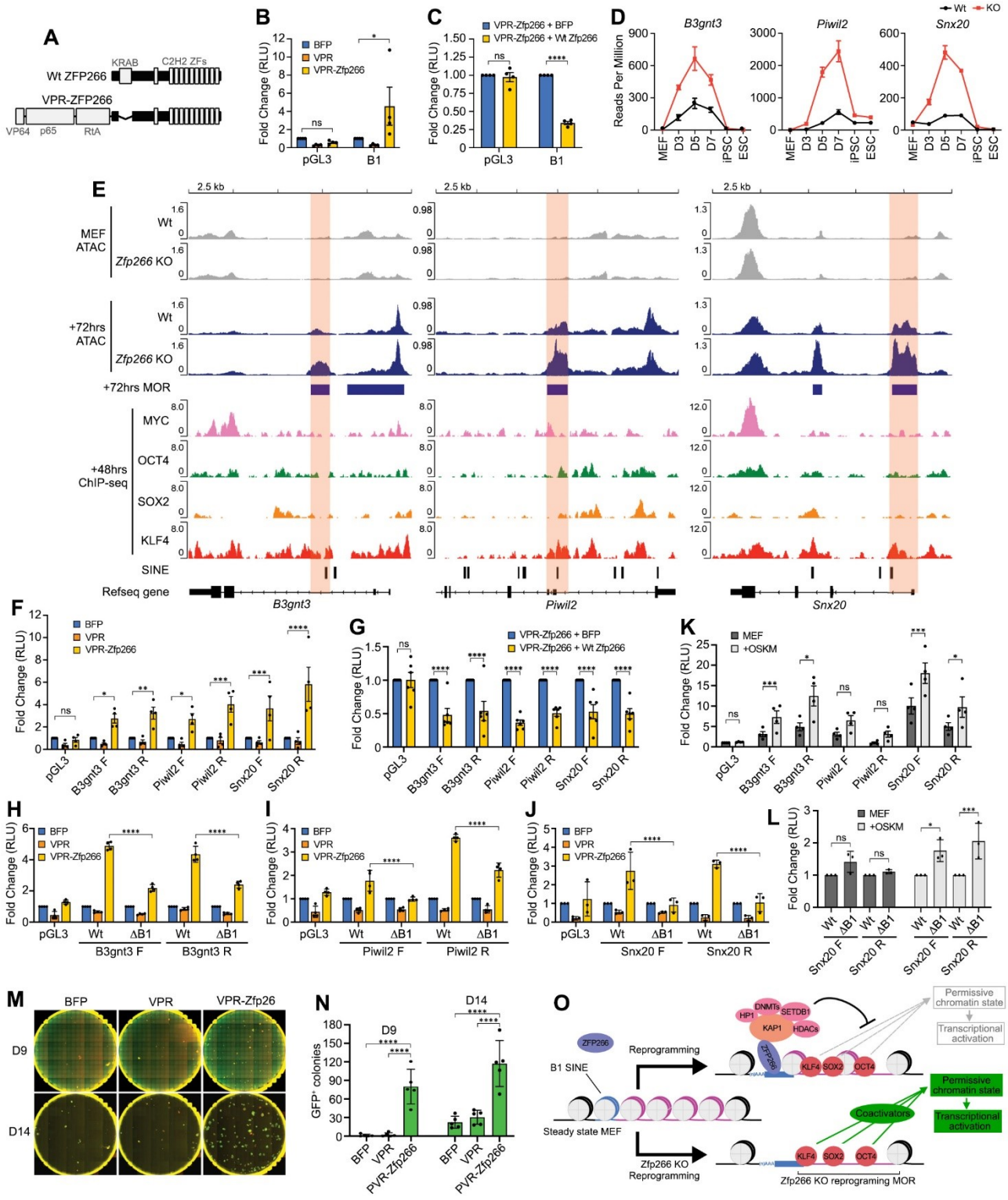


Figure 5. Reprogramming *Zfp266* KO MEFs results in chromatin opening at OSK bound, B1 SINE containing loci. A. ATAC-seq volcano plot of *Zfp266* KO vs Wt reprogramming (+72hours after OSKM induction). Green and red dots indicate more open regions (MORs) and more closed regions in *Zfp266* KO reprogramming, respectively (FDR<0.05). **B.** Significance and fold enrichment ratio of transposable element (TE) families overlap with Zfp266 KO MEF MORs. Green dots indicate significantly enriched SINEs, blue dots indicate other significantly enriched TEs. **C.** Number of *Zfp266* KO reprogramming MORs that overlap with TEs. **D.** *De novo* motif discovery analysis on *Zfp266* KO reprogramming MORs. The motifs correspond to parts of the B1 SINE consensus sequence are indicated by matching colours. **E.** Overlap between *Zfp266* KO MEF MORs and *Zfp266* KO reprogramming MORs. **F.** Motif enrichment analysis with *Zfp266* KO reprogramming MOR peak summits and percentages of MORs with KLF, SOX, OCT (POU) family and OCT4::SOX2 motifs. **G**. Classification of *Zfp266* KO reprogramming MORs based on ATAC-seq signals in MEF and reprogramming 72 hours. **H**. Reprogramming 48 hours OSKM ChIP-Seq heatmap plots at *Zfp266* KO reprogramming MORs. **I**. KLF, SOX, OCT (POU) family, OCT4::SOX2 and SINE motif distribution within *Zfp266* KO reprogramming MORs. 70 bp from the summit is highlighted in pink. The colours of B1 SINE motifs correlate to those in D. **J**. Orientation-biased distribution of B1 SINE elements within *Zfp266* KO reprogramming MORs. Head of B1 SINE tends to locate on the MOR summit side. 70 bp from the summit is highlighted in pink.



775

776 **Figure 6. ZFP266 binds to B1 SINEs in Zfp266 KO reprogramming MORs that impede OSKM-mediated**
777 **gene activation. A.** Diagram of Wt ZFP266 and a synthetic activator version of ZFP266, VPR-ZFP266. **B, C.**
778 Luciferase reporter assay with either an empty reporter vector pGL3 or with a reporter vector with the B1
779 SINE consensus sequence, co-expressed with either BFP, VPR only or VPR-*Zfp266* expression vectors (**B**),
780 with either BFP or Wt *Zfp266* expression vectors in the presence of VPR-ZFP266 (**C**) in HEK293 cells. RLU:
781 Relative Light Units, *p<0.05, ****p<0.0001 based on a two-way ANOVA test. **D.** *B3gnt3*, *Piwil2* and *Snx20*
782 mRNA expression from the *Zfp266* Wt and KO reprogramming RNA-seq data. **E.** ATAC-seq, ChIP-seq signals
783 at the *B3gnt3*, *Piwil2* and *Snx20* MORs, cloned in both forward (F) and reverse (R) directions (relative to
784 gene orientation) for luciferase reporter assays (highlighted in orange). **F-J**. Luciferase reporter assay with
785 an empty reporter vector pGL3 or vectors containing *B3gnt3*, *Piwil2* and *Snx20* MORs co-transfected with

786 either BFP, VPR only or VPR-*Zfp266* expression vectors (**F**), co-transfected with either BFP or Wt *Zfp266*
787 expression vectors in the presence of VPR-ZFP266 (**G**), an empty reporter vector pGL3, vectors containing
788 *B3gnt3* (**H**), *Piwil2* (**I**) and *Snx20* (**J**) MORs with (ΔB1) or without (Wt) B1 SINE deletion co-transfected with
789 either BFP, VPR only or VPR-*Zfp266* expression vectors in HEK293 cells. **K, L**. Luciferase reporter assay with
790 empty reporter vector pGL3 or vectors containing *B3gnt3*, *Piwil2* and *Snx20* MORs (**K**), *Snx20* MOR reporter
791 with (ΔB1) or without (Wt) B1 SINE deletion (**L**), using MEF with or without OSKM expression. *p<0.05,
792 **p<0.01, ***p<0.001, ****p<0.0001 based on a two-way ANOVA test. **M**. Day 9 and day 14 after OSKM
793 induction with overexpression of either BFP, *VPR* only or *VPR-Zfp266*. Red; mOrange, Green; *Nanog*-GFP. **N**.
794 Quantification of *Nanog*-GFP$^+$ colony numbers at day 9 and day 14. ****p<0.0001 based on a one-way
795 ANOVA test. **O**. Mechanistic model of how *Zfp266* KO enhances reprogramming. ZFP266 recruited by OSKM
796 to their target loci binds to adjacent B1 SINE and impedes chromatin opening (top), thus *Zfp266* KO results

797