Genomic features underlie the co-option of SVA transposons as cis-regulatory elements in human pluripotent stem cells

Samantha M. Barnada[1,2,#], Andrew Isopi[3,4,#], Daniela Tejada-Martinez[1], Clément Goubert[5], Sruti Patoori[1], Luca Pagliaroli[1], Mason Tracewell[1,4], and Marco Trizzino[1,*]

[1] Department of Biochemistry and Molecular Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA
[2] Genetics, Genomics and Cancer Biology PhD Program, Thomas Jefferson University, Philadelphia, PA
[3] Department of Microbiology and Immunology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA
[4] Biochemistry and Molecular Pharmacology PhD Program, Thomas Jefferson University, Philadelphia, PA
[5] Department of Human Genetics, McGill University, Montreal, Quebec, Canada

[#] Co-first authors

* Corresponding author: Marco Trizzino, Department of Biochemistry and Molecular Biology, Sidney Kimmel Medical College, Thomas Jefferson University, 233 S 10th Street, BLSB 826, Philadelphia, PA, 19104. E-mail: marco.trizzino83@gmail.com

Abstract

Domestication of transposable elements (TEs) into functional cis-regulatory elements is a widespread phenomenon. However, the mechanisms behind why some TEs are co-opted as functional enhancers while others are not are underappreciated. SINE-VNTR-Alus (SVAs) are the youngest group of transposons in the human genome, where ~3,700 copies are annotated, nearly half of which are human-specific. Many studies indicate that SVAs are among the most frequently co-opted TEs in human gene regulation, but the mechanisms underlying such processes have not yet been thoroughly investigated. Here, we leveraged CRISPR-interference (CRISPRi), computational and functional genomics to elucidate the genomic features that underlie SVA domestication into human stem-cell gene regulation. We found that ~750 SVAs are co-opted as functional cis-regulatory elements in human induced pluripotent stem cells. These SVAs are significantly closer to genes and harbor more transcription factor binding sites than non-co-opted SVAs. We show that a long DNA motif composed of flanking YY1/2 and OCT4 binding sites is enriched in the co-opted SVAs and that these two transcription factors bind consecutively on the TE sequence. We used CRISPRi to epigenetically repress active SVAs in stem cell-like NCCIT cells. Epigenetic perturbation of active SVAs strongly attenuated YY1/OCT4 binding and influenced neighboring gene expression. Ultimately, SVA repression resulted in ~3,000 differentially expressed genes, 131 of which were the nearest gene to an annotated SVA. In summary, we demonstrated that SVAs modulate human gene expression, and uncovered that location and sequence composition contribute to SVA domestication into gene regulatory networks.

Key words: Transposable Elements, SVAs, iPSCs, NCCITs, YY1, OCT4, CRISPRi

44 **Introduction**

45 Transposable elements (TEs) are mobile DNA sequences that account for over 50% of the human
46 genome, yet there is very limited knowledge on the extent of their impact on genome evolution, function,
47 and disease.

48 Many elegant studies have proposed that TE sequences constantly reshape eukaryotic gene
49 regulation [1–24] yet the underlying mechanisms are largely uncharacterized. Several TE types are active
50 and replication competent in humans, and the genomic dispersal of these elements can affect the
51 regulatory configurations of proximal host genes. For example, TE insertions may introduce novel cis-
52 regulatory elements (CREs = enhancers, promoters, insulators) at the gene locus [8,13,14,16,17].
53 Alternatively, TE insertion can disrupt transcription factor binding sites (TFBS) within pre-existing CREs,
54 thus attenuating or completely repressing nearby gene expression [25]. Additionally, insertion of TEs into
55 coding sequences of genes may disrupt the open reading frames, and modify splicing sites [26–28].
56 Dysregulated gene expression due to TE insertions can lead to disease phenotypes as TE de-repression
57 (i.e. de-methylation of TE sequences) is correlated with many neurological disorders and is a hallmark of
58 multiple cancer types [29–38].

59 Over the last decade, the scientific community has begun to characterize the biological
60 determinants of TE co-option in mammalian regulatory networks [8,13,14,17,20]. SINE-VNTR-Alus (SVAs)
61 are the youngest human TEs. These transposons are composed of a 5' CCCTCT hexamer repeat, an Alu-
62 like element, a variable number of tandem repeats (VNTR), a SINE element derived from an ancestral
63 endogenous retrovirus (HERVK-10), and a poly-A tail [39]. Six main SVA subfamilies have been
64 characterized (SVA-A through -F) and nearly half of the annotated ~3,700 copies are human-specific,
65 including all the SVA-Es and -Fs.

66 Importantly, the SVAs are still actively transposing in the human genome by taking advantage of
67 the L1-LINE machinery, and are among the most epigenetically de-repressed and transcriptionally
68 upregulated TEs across a multitude of cancers and neurological disorders [17,24,26,33]. Given their young
69 evolutionary age, SVAs provide a unique opportunity to elucidate how the human genome is evolving.
70 We, and others, have demonstrated that SVAs are frequently co-opted as functional enhancers and
71 promoters in human and chimpanzee gene regulatory networks [16,17,20,24]. Yet we have an incomplete
72 understanding of the extent of this evolutionary process and the underlying mechanisms. Why are some
73 SVAs recruited as functional enhancers and others are not? What are the molecular properties underlying
74 the regulatory potential of SVAs? Here, we answered these questions by exploring the highly permissive
75 genomic environment of induced pluripotent stem cells (iPSCs) and NCCIT cells. The latter is a human cell

76 line derived from an embryonal carcinoma which exhibits genomic properties comparable to human
77 embryonic stem cells and iPSCs, including widespread de-methylation. These cells are particularly
78 favorable for the study of transposable elements [18]. Compared to iPSCs, NCCITs are easier to manipulate
79 and more suitable to perform Cas9-mediated genome editing experiments.

80 We demonstrated that ~750 SVAs are depleted of repressive histone marks (i.e. H3K9me3) in
81 iPSCs and in NCCITs. We found that the transcriptionally active SVAs are significantly closer to genes and
82 harbor more TFBS than those harboring the repressive H3K9me3 epigenetic modification. Moreover,
83 when comparing the sequence of the de-repressed SVAs with the repressed SVAs, we detected an
84 enrichment for a long DNA motif composed of flanking binding sites for YY1/2 and OCT4 within the de-
85 repressed group. The former is a ubiquitously expressed transcriptional regulator, while the latter is an
86 essential regulator of cell pluripotency. We used Chromatin Immunoprecipitation followed by sequencing
87 (ChIP-seq) to demonstrate that YY1 and OCT4 bind adjacently on many of the active SVAs. We leveraged
88 CRISPR-interference (CRISPRi) to epigenetically repress active SVAs resulting in a loss of YY1 and OCT4
89 binding leading to massive alterations in gene expression.

90

91 **Results**

92

93 **Human-specific SVAs are enriched in areas of de-repressed chromatin**

94 We took advantage of publicly available H3K9me3 ChiP-seq data (paired-end long reads)
95 generated in induced Pluripotent Stem Cells [19] to assess to what extent SVAs are repressed in a
96 pluripotent context. We only retained uniquely mapping high quality reads (Samtools q10 filtering). This
97 analysis revealed that of all the SVAs annotated in the human genome (N=3,734; hg19 assembly),
98 approximately 80% (2,983) are decorated with this repressive histone methylation in iPSCs (Fig. 1A).
99 Conversely, 751 SVAs were depleted of H3K9me3, indicating that they are de-repressed (Fig. 1A).

100 Next, we wanted to compare the iPSCs findings with data generated in NCCITs in our laboratory.
101 As mentioned, NCCITs have pluripotent characteristics but compared to human iPSCs and embryonic stem
102 cells, the NCCITs are significantly easier to manipulate and have high transfection efficiency, which makes
103 them suitable for CRISPR experiments. We thus performed ChIP-seq for H3K9me3 to profile repressed
104 SVA regions in NCCITs. To ensure high-quality read mappability of repetitive regions we sequenced 100
105 bp long Paired-End reads and upon mapping, only retained the uniquely mapping high quality reads
106 (Samtools q10 filtering). Notably, the SVA methylation pattern previously observed in iPSCs was perfectly
107 recapitulated in NCCITs (Fig. 1B). These findings indicate that NCCITs share a similar TE epigenetic

108 landscape with iPSCs making them a suitable model to study stem cell genomics, and support that 751
109 SVAs are de-repressed in pluripotent environments (hereafter "de-repressed SVAs"; Supplementary File
110 S1).

111 We wanted to ensure that the lack of H3K9me3 signal on the de-repressed SVAs was not a
112 consequence of mapping limitations, in that the youngest SVAs may be too similar to one another to allow
113 for unique read mapping. To test this hypothesis, we looked at another histone modification. Specifically,
114 we took advantage of publicly available H3K27ac ChIP-seq data (100bp Paired-end) generated in NCCITs
115 by the Wysocka laboratory [18]. H3K27ac usually decorates active cis-regulatory elements and we
116 surmised that de-repressed SVAs may be transcriptionally active and therefore should exhibit ChIP-seq
117 signal for this histone mark. This analysis revealed that the large majority of the de-repressed SVAs
118 (636/751) are decorated with H3K27ac, which suggests they are in an active regulatory state, and that
119 they are "mappable" with uniquely mapped reads (hereafter "active SVAs"; Fig. 1C).
120 Overall, these data suggest that approximately 20% of all SVAs are located within transcriptionally de-
121 repressed and active chromatin in iPSCs and in stem cell-like NCCITs (Fig. 1A-D).

122 SVAs can be further categorized into six evolutionarily conserved subfamilies (SVA-A through -F),
123 two of which (SVA-E, -F, plus di F1 subgroup) are human-specific. We investigated the subfamily
124 composition of the de-repressed and repressed SVAs and observed that human-specific SVA subfamilies
125 -E and -F (including F1) were significantly enriched in the de-repressed group (Fisher's Exact Test $p < 1.0 \times$
126 10^{-5}; Fig. 1E). In particular, ~40% of the SVA-Es were found in a de-repressed state, as compared to an
127 average of ~19% for all the other families (Fisher's Exact Test $p < 2.2 \times 10^{-16}$; Fig. 1F).
128 In summary, these findings indicate that young, human-specific SVAs are enriched not only in de-
129 repressed chromatin, but also in active regions in NCCIT cells.
130

131 **Sequence location and composition underlies SVA activation**

132 We aimed to investigate the specific genomic features underlying the selective de-repression of
133 SVAs. First, we reasoned that active SVAs preferentially located near coding genes could be co-opted as
134 active cis-regulatory elements. We used the Genecode_v33 annotations and calculated the distance from
135 the nearest transcription start site (TSS) for both repressed and active SVAs. We observed that active SVAs
136 are significantly closer to coding genes when compared to the repressed SVAs (Wilcoxon's Rank Sum Test
137 $p < 2.2 \times 10^{-16}$; Fig. 2A). Namely, the active SVAs are located approximately 10 kb closer to the nearest TSS
138 than the repressed SVAs (Fig. 2A).

139 Next, we surmised that SVA de-repression may be a consequence of gene regulatory activity.
140 Consistent with this hypothesis, we found that relative to the repressed SVAs, active SVAs were
141 significantly closer to TFBS as defined by the Encode Consortium (Wilcoxon's rank-sum test $p = 2.84 \times 10^{-}$
142 ^{7}; Fig. 2B). Further, the active SVAs directly overlap a TFBS three times more frequently than the repressed
143 SVAs (Fig. 2B). These data suggest that the active SVAs have a higher likelihood of exhibiting gene-
144 regulatory activity.
145 To further explore this hypothesis, we performed sequence-based computational motif analysis
146 on the de-repressed SVAs using HOMER with the repressed SVAs as the background control. With this
147 approach, we identified an enrichment for a long motif composed of flanking YY1/2 and OCT4 binding
148 sites in the de-repressed SVAs (Fig. 2C). YY1/2 are ubiquitously expressed transcriptional regulators, while
149 OCT4 is one of the transcription factors essential for pluripotency maintenance. Additionally, we identified
150 an enriched CEBPA motif (Fig. 2C).
151
152 **Repression of conventionally de-repressed SVAs has global consequences on gene expression**
153 Since our results indicate that ~20% of all SVAs may act as functional CREs in NCCITs, we aimed
154 to validate our ChIP-seq data and computational predictions with a functional approach. We utilized
155 CRISPRi to repress the 751 de-repressed SVAs. We leveraged the piggyBac transposon system (Systems
156 Bioscience) to generate NCCIT cells with tetracycline-inducible expression of a functionally dead Cas9
157 (dCas9) fused to a repressive Krüppel associated box (KRAB) domain. We subsequently knocked-in two
158 previously validated [20] single guide RNAs (sgRNAs) that simultaneously target approximately 80% of all
159 the SVAs annotated in the human genome (Fig. 3A). The sgRNAs target the dCas9-KRAB to de-repressed
160 SVAs leading to the deposition of the repressive histone methylation H3K9me3 by the KRAB domain (Fig.
161 3A). This NCCIT cell line exhibiting tetracycline-inducible dCas9-KRAB expression and constitutive dual
162 sgRNA expression is hereafter referred to as SCi-NCCITs (i.e. SVA-CRISPRi-NCCITs). We treated the SCi-
163 NCCITs with doxycycline for 72 hours to robustly induce dCas9-KRAB activation and function (Fig. 3B).
164 Activation of dCas9-KRAB led to the accumulation of repressive H3K9me3 modifications on 620 of the 751
165 SVAs originally de-repressed in NCCITs (Fig. 3C). Importantly, this further demonstrates that reads can be
166 uniquely mapped on this group of SVAs.
167 To assess the impact of global SVA repression on genome-wide gene expression, we performed
168 RNA-seq on the SCi-NCCITs in the presence and absence of doxycycline treatment (three replicates per
169 condition). We identified 3,085 genes as differentially expressed upon induction of SVA repression (FDR <
170 0.05; Supplementary File S2) of which 1,596 were identified as upregulated and 1,489 as downregulated

171 (Fig. 3D). The sgRNAs used for this experiment were originally designed to target a DNA sequence shared
172 by SVAs with the LTR5H family [20]. Given this premise, we restricted our analysis exclusively to genes
173 putatively associated with human SVAs (i.e. considering only the genes that represent the closest gene to
174 an annotated SVA). Overall, 131 of the 3,085 differentially expressed genes represented the closest gene
175 to an SVA (Fig. 3E). This number is significantly higher than expected by chance (Fisher's Exact Test $p <$
176 0.001), suggesting that the expression of most of these 131 genes is likely under the direct control of SVAs.
177 Importantly, 109 of the 131 genes (83.2%) were downregulated upon doxycycline treatment (Fig. 3E),
178 indicating that SVA de-repression is necessary for gene activation. In summary, these data indicate that
179 the SVAs directly regulate the expression of at least 130 genes and that SVA repression may contribute to
180 the differential expression of up to ~2,800 genes, potentially as a cascade effect.
181 We performed Ingenuity Pathway Analysis (IPA) on the 131 SVA regulated genes and found an
182 enrichment for gap junction signaling processes including Gap Junction Signaling, Sertoli-Cell Sertoli Cell
183 Junction Signaling, and Germ Cell-Sertoli Cell Junction signaling (Fig. 3F). These processes are important
184 during gametogenesis and given that NCCIT cells retain pluripotency-like characteristics, we speculate
185 that co-opted SVAs could play a role in regulating germ line development and gametogenesis. However,
186 it is important to note that the NCCIT cell line is derived from an embryonal testicular carcinoma, and this
187 could bias our IPA. Nonetheless, recent studies have corroborated an SVA contribution in germ cells and
188 showed that in human primordial germ cells SVA sites are largely hypomethylated and genes proximal to
189 SVAs are expressed at a higher level compared to embryonic stem cells [20,40,41]. Additionally, SVA
190 transcription and retrotransposition are specifically seen in both human spermatozoa [42] and oocytes
191 [43].
192 We next used IPA to look for top upstream transcriptional regulators of the 131 SVA associated
193 genes. This analysis identified transcription factors previously highlighted by the motif analyses (CEBPA
194 and YY1/2), along with several others including SMARCA1, CREB1, and WT1 (Fig. 3G). While YY1 is an
195 essential ubiquitous developmental regulator [44], CEBPA is a regulator of differentiation in the
196 hematopoietic [45] and adipocyte [46] lineages, and is involved in gametogenesis. SMARCA1 is a
197 chromatin remodeler [47], CREB1 plays a role in both steroidal [48] and non-steroidal [49] female
198 hormonal stimulation, and WT1 is involved in urogenital specification distinctively associated with the
199 differentiation of Sertoli cells [50]. These data suggest a possible SVA contribution to germ line
200 developmental processes and demonstrates the genome-wide impact of SVA repression on gene
201 expression.
202

203 **YY1 and OCT4 mediate SVA regulatory activity**

204 Since our motif analysis revealed that de-repressed SVAs are enriched for adjacent YY1/2-OCT4
205 binding motifs and given that many differentially expressed genes were YY1/2 targets, we performed
206 ChIP-seq for YY1 and OCT4 in SCi-NCCITs with and without doxycycline treatment. As done previously for
207 the histone modification ChIP-seq, we sequenced 100 bp long Paired-End reads to ensure high mappability
208 efficiency.

209 These experiments revealed that under normal conditions (i.e. no doxycycline treatment) YY1 and OCT4
210 bind 288 and 54 SVAs, respectively (Figs. 4A, B). Specifically, we identified two different clusters of YY1
211 binding: one located in the Alu-like region (top cluster), and a second near the start of the SINE element,
212 likely in proximity of the HERVK10-derived promoter in the SINE region (second cluster). OCT4 binding
213 was limited to this second region, whereas we did not detect binding for this transcription factor in the
214 Alu-like region.

215 Importantly, binding of both YY1 and OCT4 was abolished upon SVA repression via CRISPRi (Figs. 4A, B).
216 We observed a significant overlap with 33 SVAs bound by both YY1 and OCT4 in the SINE region. In this
217 case, the binding of the two transcription factors was sequential (i.e. one next to each other) as originally
218 predicted by our motif analysis (see example in Fig. 4C).

219 Finally, we leveraged the nearest TSS approach to determine the closest gene to each of the SVAs
220 bound by YY1 and/or OCT4. Using our RNA-seq data, we investigated whether loss of YY1 and OCT4
221 binding altered the expression of neighboring genes. Notably, 44 of 288 genes located near YY1-bound
222 SVAs and 24 of 54 genes located near OCT4-bound SVAs were differentially expressed upon CRISPRi-
223 induced SVA repression (Fig. 4D). Interestingly, of the 6 genes upregulated upon doxycycline treatment,
224 4 were located near SVAs bound exclusively by OCT4 (Fig. 4D). Overall, 85% of the genes (28/33) located
225 near SVAs that were bound by both YY1 and OCT4 were differentially expressed upon CRISPR-induced SVA
226 repression (Fig. 4D). This suggests that synergistic binding of these two transcription factors on the SVA
227 sequence is important for the regulatory activity driven by these transposable elements (Fig. 4D; Fig. 5).

228

229 **Discussion**

230 SVAs are evolutionarily young transposable elements, which colonized the great ape genomes in
231 the last 10-15 million years. In fact, they are not found in gibbons, whose lineage split from the remaining
232 apes ~17 million years ago [51]. Interestingly, the gibbon genome has been independently colonized by a
233 distinct family of transposons called LAVAs (LINE-AluSz-VNTR-Alu) [52,53]. Notably, SVA and LAVA
234 structures are similar, and both have shown high cis-regulatory potential and massive co-option into gene

235 regulatory networks [16,17,20,53]. However, the mechanisms underlying SVA and LAVA domestication
236 and recruitment into primate gene regulatory networks have not been explored in depth.

237 Here, we aimed at elucidating such mechanisms, focusing specifically on SVAs and stem cells. We
238 show that ~20% of all the human SVAs are found in a de-repressed chromatin state in iPSCs and in NCCITs,
239 with a near perfect overlap between the two cell types. Most of these de-repressed SVAs were also
240 decorated with histone modifications characteristic of active enhancers and promoters (H3K27ac),
241 suggesting that their de-repression is associated with cis-regulatory activity. The youngest SVAs (which
242 are human-exclusive), and especially the SVA-Es, are the most enriched among the active copies. To this
243 end, we speculate that the older SVAs accumulated genetic mutations over time, hampering their
244 regulatory potential. Alternatively, the human genome may be adapted to silence older transposons, and
245 this may have affected copies with higher regulatory potential.

246 According to our data, SVA location and sequence composition are the best predictors of cis-
247 regulatory activity. Active SVAs are, on average, 10 kb closer to genes than the repressed ones. This is
248 likely due to the fact that the chromatin environment near gene loci may be more frequently in an
249 accessible and de-repressed state and is thus more suitable for co-option of novel cis-regulatory elements.
250 Moreover, a shorter distance from the nearest TSS may facilitate the tri-dimensional interaction between
251 the SVA-derived enhancer and the gene promoter. This is also in line with a recent study that
252 demonstrated that eQTL-rich TEs tend to be significantly closer to genes than eQTL-poor TEs [25].

253 We show that active SVAs host a significantly higher number of TFBS than the repressed ones.
254 This coincides with many studies suggesting that TE exaptation into gene regulatory networks is largely
255 driven by the evidence that they propagate TFBS across the genome [54]. We demonstrate that a long
256 motif with flanking YY1/2-OCT4 binding sites is enriched in de-repressed SVA copies relative to repressed
257 SVAs. This may mediate their function in gene regulatory networks of stem and stem-like cells. In fact,
258 OCT4 is one of the four Yamanaka factors [55] essential for pluripotency maintenance. YY1 is one of the
259 major transcriptional regulators in human cells, it is ubiquitously expressed across all cell-types and
260 performs many different functions in transcriptional regulation, including transcriptional activation,
261 repression, as well as mediation of enhancer-promoter looping [56]. To this regard, over 40% of the SVAs
262 de-repressed in iPSCs and NCCITs are bound by YY1, OCT4, or both. Importantly, when both are present,
263 they bind alongside each other as predicted by computational motif analysis. As expected, depositing
264 repressive histone methylation (H3K9me3) on the SVA locus nearly abolishes the binding of these two
265 transcription factors near SVAs, leading to an alteration of nearby gene expression. In fact, even when
266 restricting the analysis to the genes that are located near SVAs, repressing 620 of the 751 SVAs normally

267 active in NCCITs results in 131 of differentially expressed genes. The expression of most (88%) of these
268 genes is attenuated with the repression of the nearby SVA, further supporting the enhancer activity
269 provided by these transposons. These genes include important regulators of cell pluripotency and cell
270 differentiation, such as *MYC*, *MYBL2*, *FUS*, *ITGAX*, *SP4* and several others. We remark that this analysis
271 was very conservative. In fact, given the nature of the gRNAs that we chose, which target both SVAs and
272 LTR5Hs, we exclusively focused on the nearest gene to an SVA transposon, and thus the number of genes
273 affected by distal SVA repression is likely much higher.
274 In summary, in this study we provide further evidence that SVA transposons are an important
275 component of the human gene regulatory networks, specifically in stem and stem-like cells. We propose
276 a potential mechanism underlying this cis-regulatory activity where SVA location and sequence
277 composition regulate this co-option. Additional studies are required to determine if the YY1-OCT4
278 combination is a driver of SVA regulatory activity only in pluripotent cells or, alternatively, in a broader,
279 more universal context.
280
281 **Materials and Methods**
282
283 ***Antibodies and sgRNAs***
284 YY1 ChIP-seq: Cell Signaling Technology D5D9Z/46395S (15ug per ChIP). OCT4 ChIP-seq: Abcam ab181557
285 (15ug per ChIP). H3K27me3 ChIP-seq: Abcam ab8898 (3ug per ChIP). Cas9 Western Blot: Active Motif
286 61757 (1:100). GAPDH Western Blot: Cell Signaling Technology D16H11/5174 (1:1000). Anti-Rabbit IgG,
287 HRP-linked Western Blot: Cell Signaling Technology 7074 (1:10000). Anti-Mouse IgG, HRP-linked Western
288 Blot: Cell Signaling Technology 7076 (1:10000). The two sgRNAs were designed and used in a previous
289 study [20]: sgRNA1: 5' CTCCCTAATCTCAAGTACCC 3' ; sgRNA2: 5' TGTTTCAGAGAGCACGGGGT 3'.
290
291 ***NCCIT Cell Line Culture***
292 The NCCIT cell line (ATCC) was maintained in RPMI media supplemented with 10% tet-free FBS, 1%
293 penicillin-streptomycin solution, and 1% L-glutamine and incubated at 5% CO_2, 20% O_2 at 37°C.
294
295 ***ChIP-Sequencing***
296 All samples from different conditions were processed together to prevent batch effects. Between 10-15
297 million cells were cross-linked with 1% formaldehyde for 5 minutes at room temperature, quenched with
298 125 mM glycine, harvested, and washed twice with 1x PBS. The fixed cell pellet was resuspended in ChIP
299 lysis buffer (150 mM NaCl, 1% Triton X-100, 0.7% SDS, 500 µM DTT, 10 mM Tris-HCl, 5 mM EDTA) and

300 chromatin was sheared to an average length of 200–900 base-pairs, using a Covaris S220 Ultrasonicator.
301 The chromatin lysate was diluted with SDS-free ChIP lysis buffer. 15µg of antibody was used for YY1 and
302 OCT4 and 3µg of antibody for H3K9me3. The antibody was added to 5µg of sonicated chromatin along
303 with Dynabeads Protein G magnetic beads (Invitrogen) and incubated at 4°C overnight. The beads were
304 washed twice with each of the following buffers: Mixed Micelle Buffer (150 mM NaCl, 1% Triton X-100,
305 0.2% SDS, 20 mM Tris-HCl, 5 mM EDTA, 65% sucrose), Buffer 200 (200 mM NaCl, 1% Triton X-100, 0.1%
306 sodium deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, 1 mM EDTA), LiCl detergent wash (250 mM LiCl,
307 0.5% sodium deoxycholate, 0.5% NP-40, 10 mM Tris-HCl, 1 mM EDTA) and a final wash was performed
308 with 0.1X TE. Finally, beads were resuspended in 1X TE containing 1% SDS and incubated at 65°C for 10
309 min to elute immunocomplexes. The elution was repeated twice and the samples were incubated
310 overnight at 65°C to reverse cross-linking, along with the untreated input (5% of the starting material).
311 The DNA was digested with 0.5 mg/ml Proteinase K for 1 hour at 65°C and then purified using the ChIP
312 DNA Clean & Concentrator kit (Zymo) and quantified with QUBIT. Barcoded libraries were made with
313 NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs) and sequenced on an Illumina
314 NextSeq 2000 producing 100 bp paired-end reads.
315
316 ***ChIP-seq Analysis***
317 After removing the adapters with TrimGalore!, the sequences were aligned to the reference hg19, using
318 Burrows-Wheeler Alignment tool, with the MEM algorithm [57]. Uniquely mapping aligned reads were
319 filtered based on mapping quality (MAPQ > 10) to restrict our analysis to higher quality and likely uniquely
320 mapped reads, and PCR duplicates were removed. Peaks were called for each SVA site using the default
321 parameters, at 5% FDR, with default parameters
322
323 ***Generation and Culturing of SCi-NCCIT Stable Cell Lines***
324 A plasmid with a tetracycline-inducible dCas9-KRAB expression cassette flanked by piggyBac
325 recombination sites was obtained from the Wysocka Lab at Stanford University. This plasmid 'p-dCas9-
326 KRAB' confers constitutive puromycin resistance, allowing for selection of stably transduced clones when
327 co-expressed with the piggyBac transposase plasmid ('p-PB-Transposase', Systems Bioscience). The p-
328 dCas9-KRAB and p-PB-Transposase plasmids were co-transfected into NCCIT cells (ATCC) at 70%
329 confluency using a 6:1 ratio of Fugene HD (Promega) for 48 hours. Two days post-transfection, cells were
330 treated with puromycin selective media at a concentration of 1 ug/mL. Stable clones were isolated and
331 dCas9 expression assessed via Western blot.

332 Next, we obtained a piggyBac transposon plasmid containing the two sgRNAs [20] targeting ~80% of all
333 annotated SVAs in humans termed 'p-sgRNA' (Systems Bioscience). This plasmid constitutively confers
334 dual sgRNA expression and geneticin resistance. The p-sgRNA and p-PB-Transposase plasmids were co-
335 transfected into the NCCIT-dCas9KRAB cells as above. Two days post-transfection cell were treated with
336 geneticin selective media at a concentration of 400 ug/mL. Following antibiotic selection, the NCCIT-
337 dCas9KRAB-SVAsgRNA (SCi-NCCITs) cell line was maintained in ATCC-formulated RPMI media
338 supplemented with 10% tet-free FBS, 1% L-glutamine, 1µg/mL puromycin, and 400 µg/mL geneticin and
339 incubated at 5% CO2, 20% O2 at 37°C. The SCi-NCCITs were seeded to 40% confluency and treated with 2
340 ug/mL doxycycline (Sigma Aldrich) for 72 hours. For all molecular and genomic CRISPRi experiments,
341 dCas9-KRAB expression was induced with doxycycline for 72 hours.
342
343 ***Western Blot***
344 Cells were washed three times in PBS and lysed in radioimmunoprecipitation assay buffer (RIPA buffer)
345 (50 mM Tris-HCl pH7.5, 150 mM NaCl, 1% Igepal, 0.5% sodium deoxycholate, 0.1% SDS 500 µM DTT) with
346 protease inhibitors. Approximately 40 µg of whole cell lysate were loaded in Novex WedgeWell 4–20%
347 Tris-Glycine Gel (Invitrogen) and subject to SDS-PAGE. Proteins were then transferred to a Immun-Blot
348 PVDF membrane (ThermoFisher) for antibody probing. Membranes were blocked with a 10% BSA in TBST
349 solution for 30 minutes then incubated with primary antibodies in a 5% BSA in TBST, diluted as above.
350 Next, membranes were washed with TBST and incubated with secondary antibodies, diluted as above.
351 Chemiluminescent signal was detected using the Pierce ECL Plus Western Blotting Substrate
352 (ThermoFisher) and an Amersham Imager 680.
353
354 ***RNA Extraction and Library Preparation for RNA sequencing***
355 Cells were lysed in Tri-reagent (Zymo) and total RNA was extracted using Direct-zol RNA Miniprep kit
356 (Zymo) according to the manufacturer's instructions. RNA was quantified using DeNovix DS-11
357 Spectrophotometer while the RNA integrity was checked on a Bioanalyzer 2100 (Agilent). Only samples
358 with RIN value above 8.0 were used for transcriptome analysis. RNA libraries were prepared using 1µg of
359 total RNA input using NEB- Next® Poly(A) mRNA Magnetic Isolation Module, NEBNext® UltraTM II
360 Directional RNA Library Prep Kit for Illumina® and NEBNext® UltraTM II DNA Library Prep Kit for Illumina®
361 according to the manufacturer's instructions (New England Biolabs). Paired-end 100 bp reads were
362 generated.
363

364 ***RNA-seq Analysis***

365 Reads were aligned to hg19 using STAR v2.567 [58], in 2-pass mode. Bam files were filtered based on
366 alignment quality (q = 10) using Samtools [57]. We used the latest annotations obtained from Ensembl to
367 build reference indexes for the STAR alignment. Adapters were removed with TrimGalore! and Kallisto68
368 [59] was used to count reads mapping to each gene. We analyzed differential gene expression levels with
369 DESeq2 [60].

370

371 ***Statistical and Genomic Analyses***

372 All statistical analyses were performed using BEDTools v2.27.1 [61], DeepTools, and R v4.1.2. Fasta files
373 of the regions of interest were produced using BEDTools v2.27.1 [61]. Shuffled input sequences were used
374 as background. E-values < 0.001 were used as a threshold for significance. Motif analysis of de-repressed
375 SVAs on a repressed SVA background was performed using HOMER [62]. Pathway analysis was performed
376 with Ingenuity-Pathway Analysis Suite (Qiagen Inc.,
377 https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis).

378

379 **Competing interests**
380 The authors declare no competing interests.

381

382 **Acknowledgements**
383 The authors are grateful to Dr. Joanna Wysocka's lab for providing the dCas9-KRAB piggyBac plasmid and
384 in particular to Dr. Raquel Fueyo for providing critical support for the SCi-NCCIT CRISPRi line generation.
385 The authors thank the Genomic Facility at The Wistar Institute (Philadelphia, PA) for the Next Generation
386 Illumina Sequencing.

387

388 **Funding**
389 For this work, M. Trizzino was funded by the National Institute of Health (NIH-NIGMS R35GM138344) and
390 by the G. Harold and Leila Y. Mathers Foundation.

391

392 **Data availability**
393 The original genome-wide data generated in this study have been deposited in the GEO database under
394 accession code GSE192951.

395

396 **Author contributions**
397 MTrizzino designed the project. SMB and AI generated the stable SCi-NCCIT CRISPR line; AI performed a
398 preliminary set of genomic experiments. SMB performed all the genomic experiments. MTrizzino, SMB,
399 and DTM analyzed the data. MTrizzino and SMB wrote the manuscript. CB provided essential advice and
400 contribution for the planning of the CRISPR experiment. LP, SP and MTracewell contributed to some of
401 the experiments. All the authors read and approved the manuscript.

402

403 **References**

404

405 1. McClintock B. The Origin and Behavior of Mutable Loci in Maize. Proc Natl Acad Sci U S A.
406 1950;36: 344–355.

407 2. McClintock B. The significance of responses of the genome to challenge. Science.
408 1984;226: 792–801. doi:10.1126/science.15739260

409 3. Davidson EH, Britten RJ. Regulation of Gene Expression: Possible Role of Repetitive
410 Sequences. Science. 1979;204: 1052–1059. doi:10.1126/science.451548

411 4. Jordan IK, Rogozin IB, Glazko GV, Koonin EV. Origin of a substantial fraction of human
412 regulatory sequences from transposable elements. Trends in Genetics. 2003;19: 68–72.
413 doi:10.1016/S0168-9525(02)00006-9

414 5. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, et al. A distal enhancer and an
415 ultraconserved exon are derived from a novel retroposon. Nature. 2006;441: 87–90.
416 doi:10.1038/nature04696

417 6. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, et al. Species-specific
418 endogenous retroviruses shape the transcriptional network of the human tumor
419 suppressor protein p53. PNAS. 2007;104: 18613–18618. doi:10.1073/pnas.0703637104

420 7. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, et al. Evolution of the
421 mammalian transcription factor binding repertoire via transposable elements. Genome
422 Res. 2008;18: 1752–1762. doi:10.1101/gr.080663.108

423 8. Kunarso G, Chia N-Y, Jeyakani J, Hwang C, Lu X, Chan Y-S, et al. Transposable elements
424 have rewired the core regulatory network of human embryonic stem cells. Nat Genet.
425 2010;42: 631–634. doi:10.1038/ng.600

426 9. Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene
427 regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet.
428 2011;43: 1154–1159. doi:10.1038/ng.917

429 10. Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, et al. Ancient
430 transposable elements transformed the uterine regulatory landscape and transcriptome
431 during the evolution of mammalian pregnancy. Cell Rep. 2015;10: 551–561.
432 doi:10.1016/j.celrep.2014.12.052

433 11. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves Â, Kutter C, et al. Waves of
434 Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple
435 Mammalian Lineages. Cell. 2012;148: 335–348. doi:10.1016/j.cell.2011.11.058

436 12. Jacques P-É, Jeyakani J, Bourque G. The Majority of Primate-Specific Regulatory Sequences
437 Are Derived from Transposable Elements. PLOS Genetics. 2013;9: e1003504.
438 doi:10.1371/journal.pgen.1003504

439 13. Chuong EB, Rumi MAK, Soares MJ, Baker JC. Endogenous retroviruses function as species-
440 specific enhancer elements in the placenta. Nat Genet. 2013;45: 325–329.
441 doi:10.1038/ng.2553

442 14. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-
443 option of endogenous retroviruses. Science. 2016;351: 1083–1087. doi:DOI:
444 10.1126/science.aad5497

445 15. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, et al. Widespread contribution of
446 transposable elements to the innovation of gene regulatory networks. Genome Res.
447 2014;24: 1963–1976. doi:10.1101/gr.168872.113

448 16. Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, et al.
449 Transposable elements are the primary source of novelty in primate gene regulation.
450 Genome Res. 2017;27: 1623–1633. doi:10.1101/gr.218149.116

451 17. Trizzino M, Kapusta A, Brown CD. Transposable elements generate regulatory novelty in a
452 tissue-specific fashion. BMC Genomics. 2018;19: 468. doi:10.1186/s12864-018-4850-3

453 18. Fuentes DR, Swigut T, Wysocka J. Systematic perturbation of retroviral LTRs reveals
454 widespread long-range effects on human gene regulation. Heard E, Weigel D, editors.
455 eLife. 2018;7: e35989. doi:10.7554/eLife.35989

456 19. Ward MC, Zhao S, Luo K, Pavlovic BJ, Karimi MM, Stephens M, et al. Silencing of
457 transposable elements may not be a major driver of regulatory evolution in primate iPSCs.
458 Wittkopp PJ, editor. eLife. 2018;7: e33084. doi:10.7554/eLife.33084

459 20. Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, et al. Hominoid-Specific
460 Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and
461 Control Transcription in Naive Human ESCs. Cell Stem Cell. 2019;24: 724-735.e5.
462 doi:10.1016/j.stem.2019.03.012

463 21. Miao B, Fu S, Lyu C, Gontarz P, Wang T, Zhang B. Tissue-specific usage of transposable
464 element-derived promoters in mouse development. Genome Biology. 2020;21: 255.
465 doi:10.1186/s13059-020-02164-3

466 22. Judd J, Sanderson H, Feschotte C. Evolution of mouse circadian enhancers from
467 transposable elements. Genome Biology. 2021;22: 193. doi:10.1186/s13059-021-02409-9

468 23. Mika K, Marinić M, Singh M, Muter J, Brosens JJ, Lynch VJ. Evolutionary transcriptomics
469 implicates new genes and pathways in human pregnancy and adverse pregnancy
470 outcomes. Rokas A, Perry GH, Stevens A, Wildman DE, Mesiano S, editors. eLife. 2021;10:
471 e69584. doi:10.7554/eLife.69584

472 24. Patoori S, Barnada S, Trizzino M. Young transposable elements rewired gene regulatory
473 networks in human and chimpanzee hippocampal intermediate progenitors. 2021 Nov p.
474 2021.11.24.469877. doi:10.1101/2021.11.24.469877

475 25. Goubert C, Zevallos NA, Feschotte C. Contribution of unfixed transposable element
476 insertions to human regulatory variation. Philosophical Transactions of the Royal Society
477 B: Biological Sciences. 2020;375: 14.

478 26. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH. SVA Elements Are Nonautonomous
479 Retrotransposons that Cause Disease in Humans. The American Journal of Human
480 Genetics. 2003;73: 1444–1451. doi:10.1086/380207

481 27. Cosby RL, Judd J, Zhang R, Zhong A, Garry N, Pritham EJ, et al. Recurrent evolution of
482 vertebrate transcription factors by transposase capture. Science. 2021;371. doi:DOI:
483 10.1126/science.abc6405

484 28. Payer LM, Steranka JP, Ardeljan D, Walker J, Fitzgerald KC, Calabresi PA, et al. Alu insertion
485 variants alter mRNA splicing. Nucleic Acids Research. 2019;47: 421–431.
486 doi:10.1093/nar/gky1086

487 29. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ, et al. Landscape of Somatic
488 Retrotransposition in Human Cancers. Science. 2012;337: 967–971.
489 doi:10.1126/science.1222077

490 30. Li W, Jin Y, Prazak L, Hammell M, Dubnau J. Transposable Elements in TDP-43-Mediated
491 Neurodegenerative Disorders. PLOS ONE. 2012;7: e44099.
492 doi:10.1371/journal.pone.0044099

493 31. Li W, Prazak L, Chatterjee N, Grüninger S, Krug L, Theodorou D, et al. Activation of
494 transposable elements during aging and neuronal decline in Drosophila. Nat Neurosci.
495 2013;16: 529–531. doi:10.1038/nn.3368

496 32. Pugacheva EM, Teplyakov E, Wu Q, Li J, Chen C, Meng C, et al. The cancer-associated
497 CTCFL/BORIS protein targets multiple classes of genomic repeats, with a distinct binding
498 and functional preference for humanoid-specific SVA transposable elements. Epigenetics
499 & Chromatin. 2016;9: 35. doi:10.1186/s13072-016-0084-2

500 33. Anwar SL, Wulaningsih W, Lehmann U. Transposable Elements in Human Cancer: Causes
501 and Consequences of Deregulation. International Journal of Molecular Sciences. 2017;18:
502 974. doi:10.3390/ijms18050974

503 34. Krug L, Chatterjee N, Borges-Monroy R, Hearn S, Liao W-W, Morrill K, et al.
504 Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43
505 model of ALS. PLOS Genetics. 2017;13: e1006635. doi:10.1371/journal.pgen.1006635

506 35. Guo C, Jeong H-H, Hsieh Y-C, Klein H-U, Bennett DA, De Jager PL, et al. Tau Activates
507 Transposable Elements in Alzheimer's Disease. Cell Reports. 2018;23: 2874–2880.
508 doi:10.1016/j.celrep.2018.05.004

509 36. Kong Y, Rose CM, Cass AA, Williams AG, Darwish M, Lianoglou S, et al. Transposable
510 element expression in tumors is associated with immune infiltration and increased
511 antigenicity. Nat Commun. 2019;10: 5228. doi:10.1038/s41467-019-13035-2

512 37. Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM, et al. Transposable elements
513 drive widespread expression of oncogenes in human cancers. Nat Genet. 2019;51: 611–
514 617. doi:10.1038/s41588-019-0373-3

515 38. Ivancevic A, Chuong EB. Transposable elements teach T cells new tricks. PNAS. 2020;117:
516 9145–9147. doi:10.1073/pnas.2004493117

517 39. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, et al. SVA Elements: A Hominid-
518 specific Retroposon Family. Journal of Molecular Biology. 2005;354: 994–1007.
519 doi:10.1016/j.jmb.2005.09.085

520 40. Tang WWC, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, et al. A Unique Gene
521 Regulatory Network Resets the Human Germline Epigenome for Development. Cell.
522 2015;161: 1453–1467. doi:10.1016/j.cell.2015.04.053

523 41. Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, et al. Sperm Methylation
524 Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates. Cell.
525 2011;146: 1029–1041. doi:10.1016/j.cell.2011.08.016

526 42. Lazaros L, Kitsou C, Kostoulas C, Bellou S, Hatzi E, Ladias P, et al. Retrotransposon
527 expression and incorporation of cloned human and mouse retroelements in human
528 spermatozoa. Fertility and Sterility. 2017;107: 821–830.
529 doi:10.1016/j.fertnstert.2016.12.027

530 43. Georgiou I, Noutsopoulos D, Dimitriadou E, Markopoulos G, Apergi A, Lazaros L, et al.
531 Retrotransposon RNA expression and evidence for retrotransposition events in human
532 oocytes. Human Molecular Genetics. 2009;18: 1221–1228. doi:10.1093/hmg/ddp022

533 44. Gordon S, Akopyan G, Garban H, Bonavida B. Transcription factor YY1: structure, function,
534 and therapeutic implications in cancer biology. Oncogene. 2006;25: 1125–1142.
535 doi:10.1038/sj.onc.1209080

536 45. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, et al. Dominant-
537 negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-α (C/EBPα), in
538 acute myeloid leukemia. Nat Genet. 2001;27: 263–270. doi:10.1038/85820

539 46. Umek RM, Friedman AD, McKnight SL. CCAAT-Enhancer Binding Protein: A Component of a
540 Differentiation Switch. Science. 1991;251: 288–292. doi:10.1126/science.1987644

541 47. Clapier CR, Cairns BR. The Biology of Chromatin Remodeling Complexes. Annual Review of
542 Biochemistry. 2009;78: 273–304. doi:10.1146/annurev.biochem.77.062706.153223

543 48. Zubenko GS, Hughes HB. Effects of the G(−656)A variant on CREB1 promoter activity in a
544 neuronal cell line: Interactions with gonadal steroids and stress. Mol Psychiatry. 2009;14:
545 390–397. doi:10.1038/mp.2008.23

546 49. Sirotkin AV, Benčo A, Mlynček M, Harrath AH, Alwasel S, Kotwica J. The involvement of the
547 phosphorylatable and nonphosphorylatable transcription factor CREB-1 in the control of
548 human ovarian cell functions. Comptes Rendus Biologies. 2019;342: 90–96.
549 doi:10.1016/j.crvi.2019.03.002

550 50. Chen M, Zhang L, Cui X, Lin X, Li Y, Wang Y, et al. Wt1 directs the lineage specification of
551 sertoli and granulosa cells by repressing Sf1 expression. Development. 2016; dev.144105.
552 doi:10.1242/dev.144105

553 51. Grabowski M, Jungers WL. Evidence of a chimpanzee-sized ancestor of humans but a
554 gibbon-sized ancestor of apes. Nat Commun. 2017;8: 880. doi:10.1038/s41467-017-00997-
555 4

556 52. Meyer TJ, Held U, Nevonen KA, Klawitter S, Pirzer T, Carbone L, et al. The Flow of the
557 Gibbon LAVA Element Is Facilitated by the LINE-1 Retrotransposition Machinery. Genome
558 Biology and Evolution. 2016;8: 3209–3225. doi:10.1093/gbe/evw224

559 53. Okhovat M, Nevonen KA, Davis BA, Michener P, Ward S, Milhaven M, et al. Co-option of
560 the lineage-specific LAVA retrotransposon in the gibbon genome. PNAS. 2020;117: 19328–
561 19338. doi:10.1073/pnas.2006038117

562 54. Sundaram V, Wysocka J. Transposable elements as a potent source of diverse cis-
563 regulatory sequences in mammalian genomes. Philosophical Transactions of the Royal
564 Society B: Biological Sciences. 2020;375: 20190347. doi:10.1098/rstb.2019.0347

565 55. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and
566 Adult Fibroblast Cultures by Defined Factors. Cell. 2006;126: 663–676.
567 doi:10.1016/j.cell.2006.07.024

568 56. Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, et al. YY1 Is a Structural
569 Regulator of Enhancer-Promoter Loops. Cell. 2017;171: 1573-1588.e28.
570 doi:10.1016/j.cell.2017.11.008

571 57. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
572 Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079.
573 doi:10.1093/bioinformatics/btp352

574 58. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal
575 RNA-seq aligner. Bioinformatics. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635

576 59. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
577 quantification. Nat Biotechnol. 2016;34: 525–527. doi:10.1038/nbt.3519

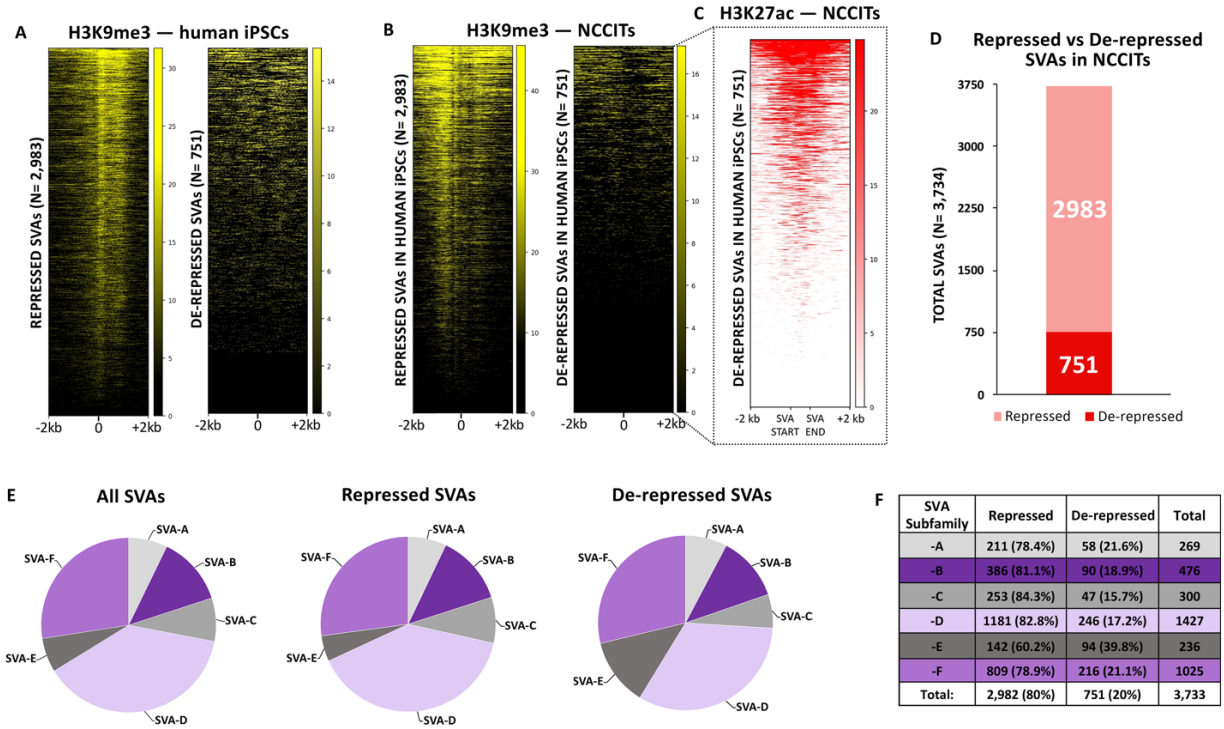
578 60. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
579 seq data with DESeq2. Genome Biology. 2014;15: 550. doi:10.1186/s13059-014-0550-8

580 61. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
581 Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

582 62. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of
583 Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for
584 Macrophage and B Cell Identities. Molecular Cell. 2010;38: 576–589.
585 doi:10.1016/j.molcel.2010.05.004

586

587 **Figures & Figure legends**
588

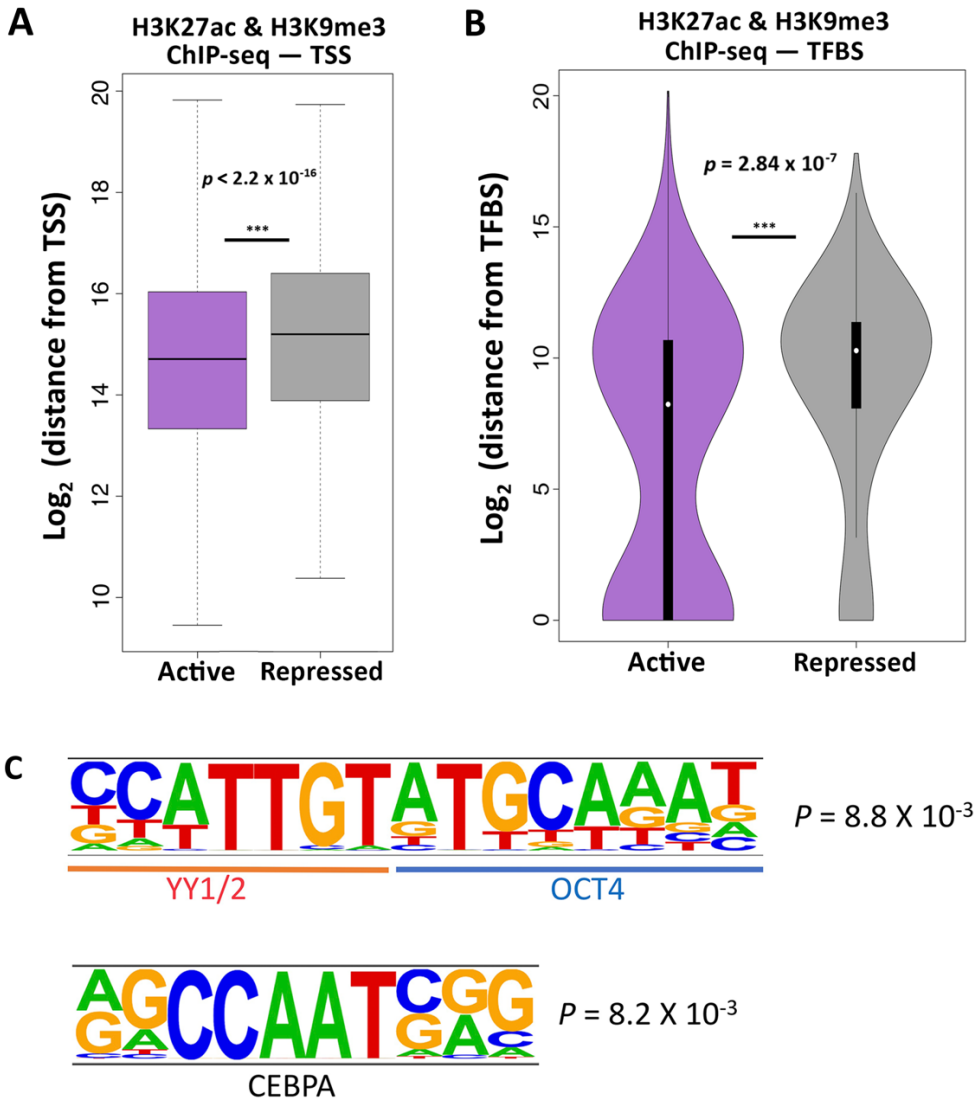


SVA Subfamily	Repressed	De-repressed	Total
-A	211 (78.4%)	58 (21.6%)	269
-B	386 (81.1%)	90 (18.9%)	476
-C	253 (84.3%)	47 (15.7%)	300
-D	1181 (82.8%)	246 (17.2%)	1427
-E	142 (60.2%)	94 (39.8%)	236
-F	809 (78.9%)	216 (21.1%)	1025
Total:	2,982 (80%)	751 (20%)	3,733

589
590

591 **Figure 1** –751 human SVAs are active in human pluripotent cells. (A) Heatmaps depicting H3K9me3 ChIP-
592 seq signal at all human SVAs in iPSCs. (B) ChIP-seq for H3K9me3 in human NCCITs at SVA regions previously
593 classified as repressed and de-repressed in iPSCs. (C) ChIP-seq for H3K27ac in NCCIT cells. The heatmap is
594 centered on the 751 de-repressed SVAs. (D) Barplot representing the number of total repressed and de-

595 repressed human SVAs (i.e. SVAs decorated with H3K9me3 versus lacking H3K9me3). 2982 of the ~3700
596 human SVAs were repressed, while 751 were de-repressed. (E) & (F) Human-specific SVAs, subfamilies
597 SVA-E and -F are enriched within the de-repressed SVA population.



598

599 **Figure 2** – Specific genomic features characterize the de-repressed SVAs. (A) SVAs in an active
600 configuration are approximately 10kb closer to TSS than SVAs in a repressed configuration (Wilcoxon's
601 Rank Sum Test $p < 2.2 \times 10^{-16}$). (B) Active SVAs are significantly closer to, and directly overlap with, TFBS
602 three times more than repressed SVAs (Wilcoxon's rank-sum test $p = 2.84 \times 10^{-7}$). (C) Motif analysis
603 performed on the de-repressed SVAs shows enrichment for consecutive YY1/2 and OCT4 motifs.

604

605

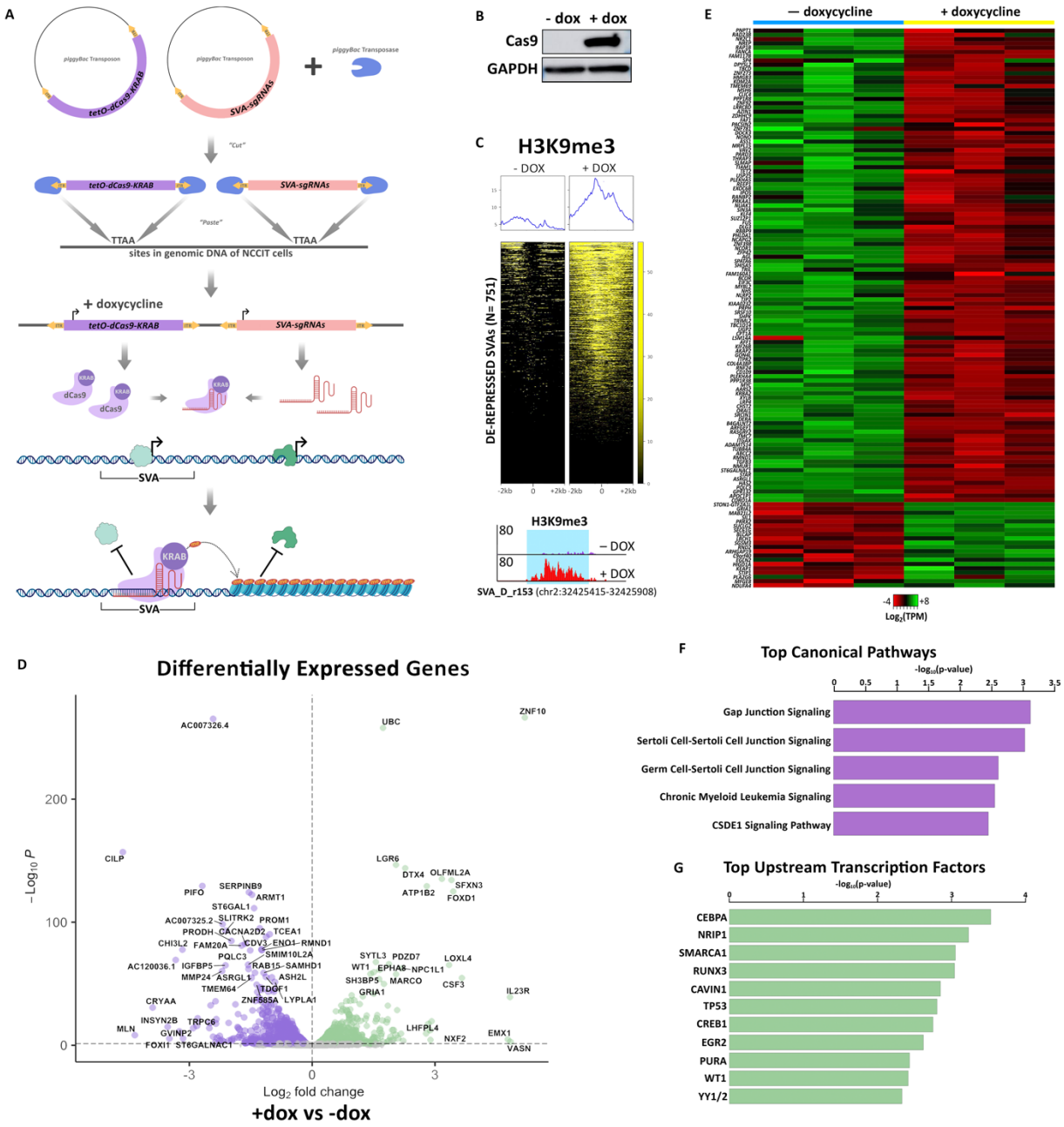
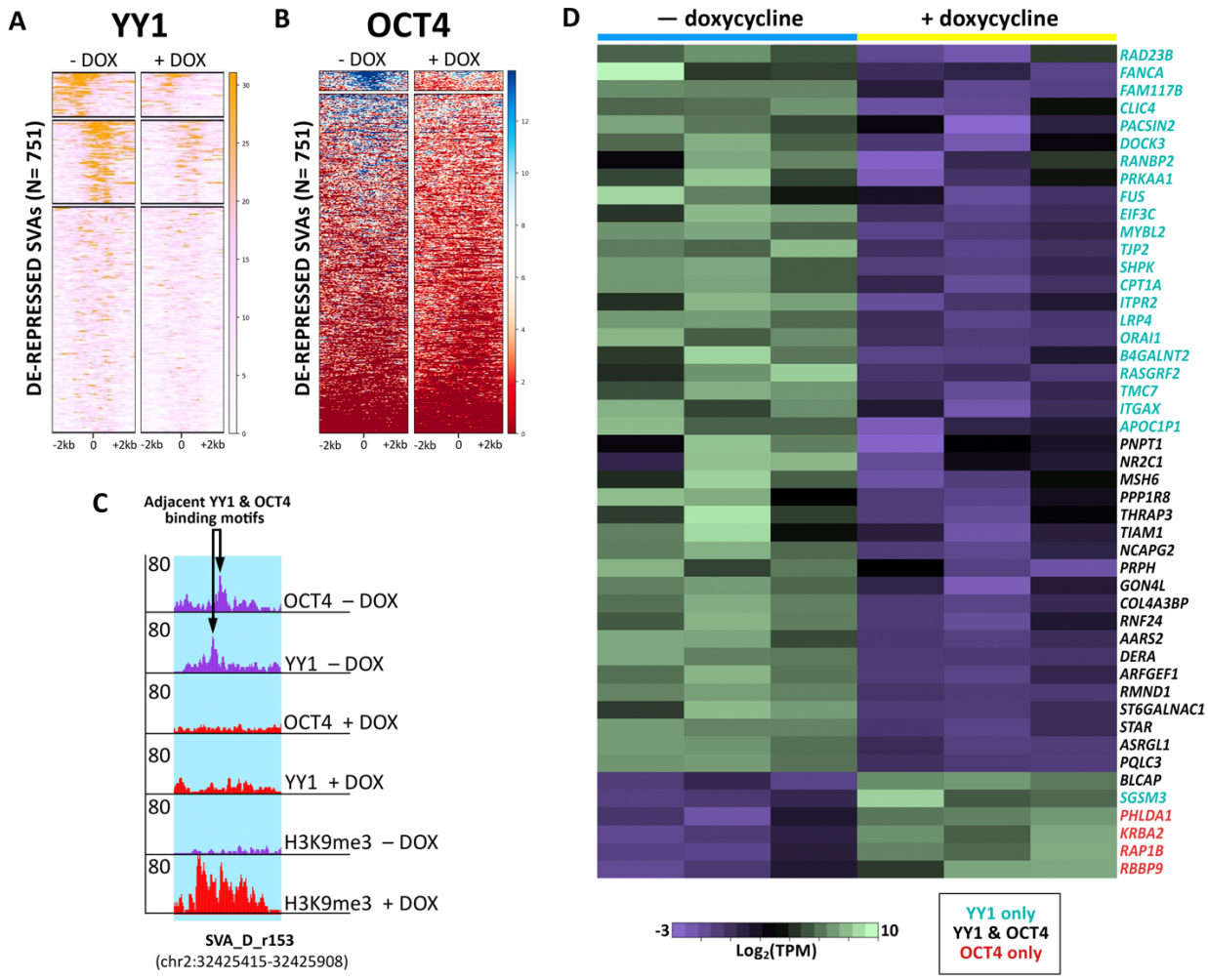


Figure 3 – CRISPRi-mediated repression of conventionally de-repressed SVAs results in aberrant gene expression. (A) Schematic of the generation of the SCi-NCCIT line (created with BioRender.com). (B) Cas9 immunoblot displaying activation of dCas9-KRAB 72 hours post-doxycycline induction. (C) H3K9me3 ChIP-seq heatmap of the SCi-NCCITs shows increased H3K9me3 signal post-doxycycline treatment. Genome browser screenshot displaying increased H3K9me3 before and after treatment with doxycycline at a de-repressed SVA-D. (D) Volcano plot showing genes differentially expressed in SCi-NCCITs after doxycycline treatment (purple = downregulated genes; green = upregulated genes). (E) Heatmap of 131 genes that

614 are differentially expressed post-doxycycline treatment and also represent the nearest gene to an SVA.
615 (F, G) Top canonical pathways and top upstream regulators predicted by IPA for the 131 genes
616 differentially expressed after doxycycline treatment.



617

618 **Figure 4** – Individual and sequential binding of YY1 and OCT4 contributes to SVA regulatory activity. (A)
619 YY1 ChIP-seq signal at the 751 de-repressed SVAs before and after doxycycline treatment. (B) OCT4 ChIP-
620 seq signal at the 751 de-repressed SVAs before and after doxycycline treatment. (C) Genome browser
621 screenshot displaying the sequential binding of YY1 and OCT4 at a de-repressed SVA-D. Upon doxycycline
622 induction, the binding of both transcription factors is lost, while H3K9me3 signal is gained. (D) Heatmap
623 of 47 genes that are differentially expressed upon doxycycline treatment and regulated by YY1 and/or
624 OCT4 (Black: YY1 and OCT4, Teal: YY1 only, Red: OCT4 only).

625

626

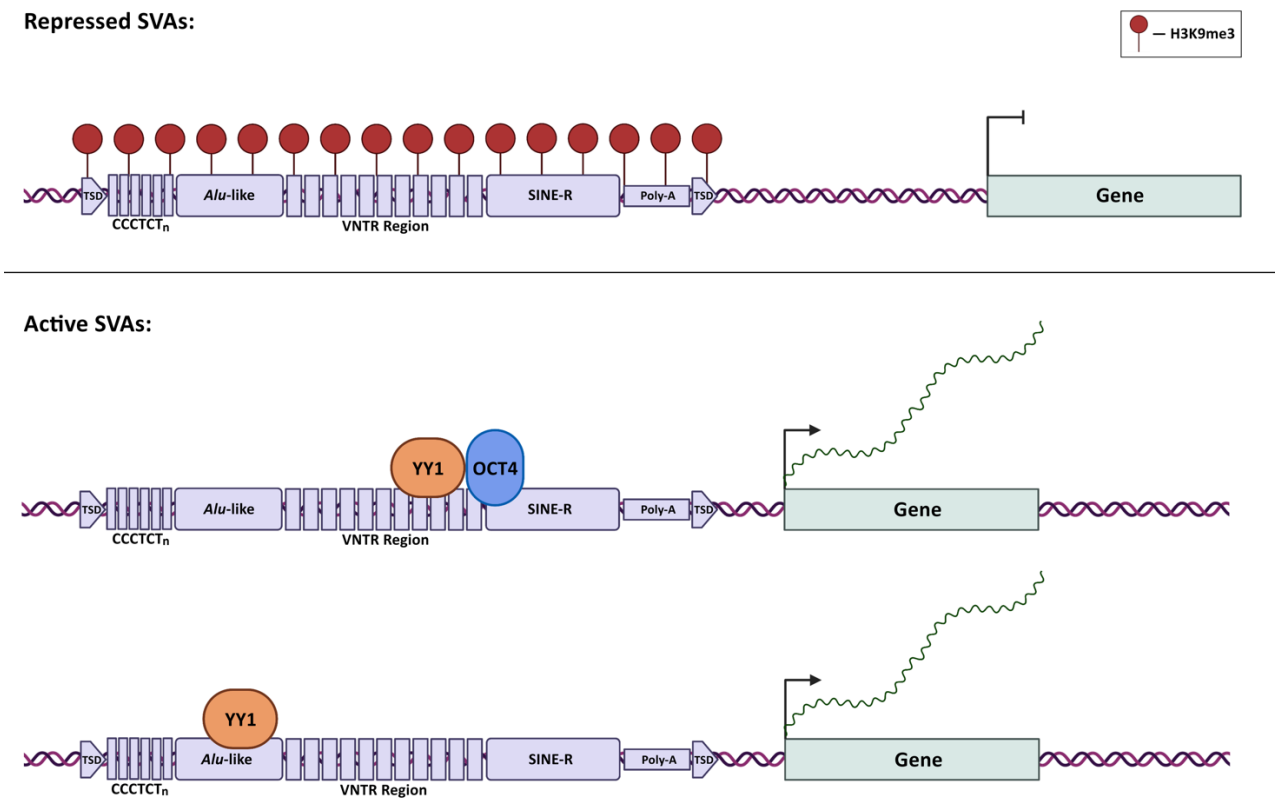


Figure 5 – Model for SVA co-option in human pluripotent stem cell gene regulation. This figure was created with BioRender.com.