Isoforms of the TAL1 transcription factor have different roles in hematopoiesis and cell growth

Aveksha Sharma[1†], Shani Mistriel-Zerbib[2†], Rauf Ahmad Najar[1#], Eden Engal[1], Mercedes Bentata[1], Nadeen Taqatqa[1], Sara Dahan[1], Klil Cohen[1], Shiri Jaffe-Herman[1], Ophir Geminder[1], Mai Baker[1], Yuval Nevo[3], Inbar Plaschkes[3], Gillian Kay[1], Yotam Drier[2], Michael Berger[2] and Maayan Salton[1*].

[1]Department of Biochemistry and Molecular Biology, The Institute for Medical Research Israel–Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem 91120, Israel, [2]The Lautenberg Center for Immunology and Cancer Research, The Institute for Medical Research Israel–Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem 91120, Israel and [3]Info-CORE, Bioinformatics Unit of the I-CORE Computation Center, and Institute for Medical Research Israel-Canada, The Hebrew University of Jerusalem, Jerusalem 91120, Israel. † These authors contributed equally to this work.

#Current address: Department of Pediatrics, Lung Biology and Disease Program, University of Rochester School of Medicine and Dentistry, Rochester, New York, USA.

*To whom correspondence should be addressed. Tel: +972 26758814; Email: maayan.salton@mail.huji.ac.il

Running head: TAL1 isoforms, hematopoiesis and cell growth

Keywords: T-ALL, transcription factor, TAL1, hematopoiesis, cell growth

ABSTRACT

T-cell acute lymphoblastic leukemia (T-ALL) protein 1 (TAL1) is a central transcription factor in hematopoiesis. The timing and level of TAL1 expression orchestrate the differentiation to specialized blood cells and its overexpression is a common cause of T-ALL. Here we studied the two protein isoforms of TAL1, short and long, that are generated by the use of alternative promoters as well as by alternative splicing. We analyzed the expression of each isoform by deleting an enhancer or insulator, or by opening chromatin at the enhancer location. Our results show that each enhancer promotes expression from a specific TAL1 promoter. Expression from a specific promoter gives rise to a unique 5′ UTR with differential regulation of translation. Moreover, our study suggests that the enhancers regulate TAL1 exon 3 alternative splicing by inducing changes in the chromatin at the splice site, which we demonstrate is mediated by KMT2B. Furthermore, our results indicate that TAL1-short binds more strongly to TAL1 E-protein partners and functions as a stronger transcription factor than TAL1-long. Specifically TAL1-short has a unique transcription signature promoting apoptosis. Finally, when we expressed both isoforms in mice bone-marrow, we found that while overexpression of both isoforms prevents lymphoid differentiation, expression of TAL-short alone leads to hematopoietic stem cell exhaustion. Furthermore, we found that TAL1-short promoted erythropoiesis and reduced cell survival in the CML cell line K562. While TAL1 and its partners are considered promising therapeutic targets in the treatment of T-ALL, our results show that TAL1-short could act as a tumor suppressor and suggest that altering TAL1 isoform's ratio could be a preferred therapeutic approach.

INTRODUCTION

Tissue-specific programs of gene expression are driven by key transcription factors. T-cell acute lymphoblastic leukemia (T-ALL) protein 1 (TAL1, also known as SCL) is a pivotal transcription factor in hematopoiesis, directing specification of mesoderm differentiation into specialized blood cells. Specifically, TAL1 is expressed in hematopoietic stem cells and erythroid and megakaryocytic lineages, whereas it is silenced in B- and T-cell lineages during lymphocyte development. In addition to hematopoiesis, TAL1 is also an obligatory factor in adult hematopoietic stem cells and for terminal maturation of select blood lineages. Knockdown of TAL1 in mice is lethal at embryonic day 9.5 (E9.5) due to the absence of yolk sac primitive erythropoiesis and myelopoiesis (1-3). TAL1's role in very early hematopoietic development was delineated in chimeric mice harboring homozygous mutant TAL1-/- embryonic stem cells (4, 5). This complete block of hematopoiesis suggests a function in either the first differentiation steps from progenitor cells or the specification of mesodermal cells toward a blood fate. Knocking-down TAL1 later in development (6–8 weeks of age) perturbed megakaryopoiesis and erythropoiesis with the loss of early progenitor cells in both lineages (6).

Aberrant expression of TAL1 is the most common cause of T-ALL (7). TAL1 overexpression is reported to occur through chromosomal translocation, intrachromosomal rearrangement, or a mutation in the enhancer region (8). TAL1 forms a dimer with E-proteins and in this way prevents T-cell differentiation promoted by E-protein dimerization (9). Thus, silencing of TAL1 is a requirement for T-cell differentiation and its expression in this lineage promotes stem cell features and as a result T-ALL.

TAL1's critical role at specific time points during hematopoiesis suggests that its expression is highly regulated. To date, five promoters (10-13) and three enhancers have been identified for TAL1 (14-16). TAL1 promoters were identified by their probable location relative to a transcription start-site (TSS) and a specific chromatin signature associated with promoters, which can be a DHS site, H3K9ac, H3K27ac, or H3K4me3. TAL1 enhancers were identified by a similar chromatin signature minus H3K4me3; their connection to the *TAL1* gene required experimental validation of loop architecture. Each of TAL1 enhancers directs expression to a subdomain of the normal TAL1 expression pattern (17-19). Enhancers -4 (all locations are in Kb relative to promoter 1a, Fig. 1A) and +14 were shown to promote TAL1 expression in hematopoietic stem cells and enhancer +51 was shown to drive TAL1 expression in erythrocytes and megakaryocytes (17-19). The connection of a specific enhancer to TAL1 promoter in a specific cell type is enabled by differential loops in the different cells (16). These

79 loops are established by the DNA binding protein CTCF, which upon dimerization creates a
80 loop. Only when the enhancer and promoter are present in the same loop will a connection
81 occur. Furthermore, insertion of just a few nucleotides at position -8 was shown to create an
82 enhancer in patients with T-ALL (14); this enhancer was named "mutation of the TAL1
83 enhancer" (MuTE). Not much is known about the pairing of the specific promoter-enhancer,
84 and it is clear that this is part of the complex expression of TAL1. This suggests that different
85 enhancers allow for expression of a specific isoform of TAL1 at different developmental
86 stages.

87 TAL1 has four mRNA transcripts (TSS1-4) coding to the TAL1-long protein isoform and TSS5
88 which codes for the TAL1-short protein isoform (Fig. 1A) (10-12). Transcription beginning at
89 TSS1-4 with exclusion of alternative exon 3 will also translate to TAL1-short protein (20) (Fig.
90 1A). Both isoforms harbor the basic helix-loop-helix (bHLH) domain directly binding to
91 DNA and were previously described in zebrafish (21-23), chickens (24) and humans (24,
92 25) (Fig. 1B).

93 Here we aimed to understand the RNA processing regulation that gives rise to TAL1's two
94 protein isoforms and their unique functions. While the isoforms have been studied before,
95 our work is the first to systemically study all aspects of their regulation and function. We
96 manipulated TAL1's enhancers and monitored TAL1 expression as well as its alternative
97 splicing. We took advantage of cells with activated/mutated enhancers and used a
98 CRISPR/dCas9 system conjugated to p300 to change chromatin at TAL1's -60 enhancer.
99 We found that the activated enhancer promotes expression as well as inclusion of TAL1
100 exon 3. Monitoring the chromatin at TAL1 exon 3 led us to speculate that the -60 enhancer
101 affects exon 3 alternative splicing regulation by altering the chromatin at this exon. We further
102 characterized the activity of the promoters and 5′ UTRs and found that while promoter 4 is
103 the strongest, the 5′ UTR transcribed from promoter 5 is the best at promoting translation.
104 Our findings from co-immunoprecipitation (co-IP), chromatin immunoprecipitation followed
105 by sequencing (ChIP-seq) and RNA-seq, indicate that TAL1-short is a stronger binding
106 partner of E-proteins, stronger at binding DNA and promotes expression of apoptotic genes.
107 To study the role of the isoforms in hematopoiesis, we expressed each of the protein isoforms
108 in mice bone marrow and found TAL1-short to reduce survival of hematopoietic stem cells.
109 In addition, using the chronic myeloid leukemia (CML) cell line K562 we showed that TAL1-
110 short promotes erythropoiesis. Finally, we demonstrated that while TAL1 was known to
111 promote cell growth, TAL1-short expression in Jurkat and K562 cell lines stimulates cell death.

112 In summary, our results demonstrate that TAL1 isoforms harbor differential functions in
113 hematopoiesis and cell growth both *in vitro* and *in vivo*.

114 **RESULTS**

115 **TAL1's chromatin landscape and gene expression regulation**

116 We set out to characterize the regulation and function of TAL1 isoforms. The prediction tool
117 ProtParam tool (26) indicated that the half-life of the TAL1-long isoform is 7.2 h and that of
118 the TAL1-short is 1.1 h. The differential half-life and protein length (Fig. 1B) indicate the
119 possibility that these proteins have distinct functions necessitating precise regulation. We
120 began by evaluating the amount of the TAL1 protein and mRNA isoforms in the T-ALL cell
121 lines Jurkat, CutLL1, MOLT4 and CCRF-CEM as well as in the chronic myeloid leukemia
122 (CML) cell line K562 (Fig. 1C and Supp. Fig. S1A-C). In addition, we used available data to
123 study the chromatin in inter- and intra-gene regions at the TAL1 gene location. To explore the
124 regulation giving rise to TAL1 mRNA and protein isoforms, we examined H3K27ac, which
125 indicates open chromatin at promoter and enhancer regions, and H3K4me3 marks, which
126 indicates promoters (Fig. 1D and Supp. Fig. S1C). In the CutLL1 cell line there was no
127 detection of TAL1 protein or mRNA and only weak H3K4me3 marks at promoter 1-4 region
128 (Fig. 1C&D and Supp. Fig. S1A&B). In the MOLT4 cell line, known to harbor the MuTE
129 enhancer, we detected a similar amount of both protein isoforms and H3K27ac marks at an
130 enhancer at the -8 location and promoter 5; and H3K4me3 marks at promoter 1 and 5 (Fig.
131 1C&D and Supp. Fig. S1A&B). The CCRF-CEM cell line harbors the SIL-TAL1 rearrangement,
132 which allows TAL1 expression from the SIR promoter (27). In this cell line, both Tal1 isoforms
133 are present, with Tal1-long being dominant. Furthermore, open chromatin is only observed
134 in the region spanning from promoter 2 to promoter 4 region (Fig. 1C&D and Supp. Fig.
135 S1A&B). As was described before, Jurkat cells overexpress TAL1 as a result of a mutation
136 creating a strong -8 MuTE enhancer. While we could detect TAL1-short, the expression of
137 TAL1-long protein isoform was higher (Fig. 1C). Strong enhancers or super enhancers are
138 marked by wide H3K27ac modifications, as can be seen at the -8 MuTE enhancer where the
139 modification spreads for 15 Kbp from the -8 MuTE enhancer to TAL1 exon 3 (Fig. 1D). In
140 addition we observe broad H3K4me3 modification spanning 7.5 Kbp from TAL1 promoter 1
141 to promoter 5. Broad H3K4me3 modifications are correlated with strong enhancers and
142 increased transcription (28) (Fig. 1D). In the K562 cell line H3K27ac and H3K4me3
143 modification marks an enhancer at +20 location and a broad modification of both histone
144 marks at promoter 1 for 8 Kbp, suggesting strong expression from TSS1-4 that is supported

145 by the protein quantification (Fig. 1C&D and Supp. Fig. S1C). These results point to a unique
146 and complex regulation giving rise to specific amounts of TAL1 isoforms.

147 The mRNA level of TAL1 transcripts did not match the protein amount of the isoforms (Supp.
148 Fig. S1A&B and Fig. 1C). This is due in part to the real-time PCR method that was used that
149 can only compare a specific isoform between samples and not one isoform to the other due to
150 differential primer efficiency. In addition, while specific primers for transcription initiating
151 from promoters 1 to 4 can be designed, such primers could not be designed for promoter 5, as
152 its transcript has no unique sequence. Promoter 5 gives rise to the TAL1-short protein isoform
153 as does exclusion of exon 3 (TAL1-ΔEx3). Hence the detection of TAL1-short protein on the
154 mRNA level is only possible by measuring TAL1-ΔEx3, which does not provide the full
155 picture.

156 **Mutating TAL1 -8 MuTE enhancer and its effect on mRNA processing**

157 To further explore the effect of known TAL1 enhancers on each isoform, we began by focusing
158 on the effect of TAL1 -8 MuTE enhancer in Jurkat cells. We chose to use the Jurkat cell line
159 since its MuTE enhancer is well-characterized (14). We started by performing a 5′ RACE
160 experiment in Jurkat cells to discover the active promoters in this cell line. Our results show
161 transcription initiation from TSS1, 2, 4 and 5 (Supp. Fig. S2A). For this reason, hereon we
162 focused on these TSSs. To study the specific connection between the -8 MuTE enhancer and
163 TAL1 mRNA processing, we used Jurkat cells with deletion of 12 nucleotides of the -8 MuTE
164 enhancer and insertion of exogenous TAL1-long to allow for cell growth (Jurkat Del-12) (14).
165 As reported previously (14), we detected a five-fold reduction in TAL1 total mRNA amount
166 following the impairment of the -8 MuTE enhancer (Fig. 2A and Supp. Fig. S2C). Specifically,
167 the reduction in expression was more pronounced in promoter 2 as compared to promoters 1
168 and 4 (Supp. Fig. S2D). We cannot exclude the possibility that overexpression of TAL1-long
169 in the Jurkat Del-12 cell line may contribute to the observed expression pattern from each
170 promoter. Furthermore, we detected a 50% increase in TAL1-ΔEx3 in mutated -8 MuTE
171 enhancer cells (Fig. 2B). Mutation at TAL1 -8 MuTE enhancer was shown to deplete the
172 H3K27ac from the TAL1 region (14). We speculate that this is the reason for the reduced
173 transcription as well as the change in alternative splicing, which is also known to be regulated
174 by chromatin at the alternative exon (29).

175 **Activation of -60 enhancer and its effect on mRNA processing**

176 We continued checking the effect of the enhancers on the isoform preference by activating
177 the -60 enhancer in HEK293T cells by deleting a CTCF site between TAL1 promoter and -

178 60 enhancer (Supp. Fig. S2E). We selected the HEK293T cell line because it has a previously
179 characterized enhancer and well-defined chromosome conformation. This deletion alters
180 the genome loops, allowing for the -60 enhancer to connect with TAL1 promoters, as
181 demonstrated by the chromosome conformation capture technique 5C, which captures
182 frequently interacting chromatin segments (16). Comparing parental HEK293T to HEK293T
183 ΔCTCF, we found twice the amount of total mRNA and protein of TAL1 (Fig. 2C and Supp.
184 Fig. S2F), as seen previously (16). While we detected more expression from promoters 1, 2
185 and 4 (Supp. Fig. S2G), the increase in TAL1 total mRNA was accompanied by a 50%
186 decrease in TAL1-ΔEx3 (Fig. 2D). We speculate that enhancer activation led to open
187 chromatin at the TAL1 promoter and exon 3, which mediates these changes.

188 To check our hypothesis, we manipulated the chromatin at the TAL1 -60 enhancer by
189 tethering a general enhancer activator, the acetyltransferase p300. To this end, we used a well-
190 characterized nuclease-deficient Cas9 (dCas9) conjugated to the p300 enzymatic core (30). The
191 CRISPR/dCas9 system allows us to tether a chromatin protein core domain to specific
192 chromosome locations using a pool of four gRNAs. A pool of gRNAs was shown empirically
193 to perform better than a single gRNA (30). We selected HEK293T cells for their low expression
194 of TAL1 compared to other cells, such as Jurkat. This feature contributes to a high dynamic
195 range when tethering p300 core to a closed chromatin region. We co-transfected HEK293T
196 cells with dCas9-p300 core (mut) or dCas9-p300 core (WT) and four gRNA plasmids targeting
197 TAL1 -60 enhancer (Supp. Fig. S2E&H). Our results demonstrate that tethering p300's
198 enzymatic core to TAL1 -60 enhancer upregulates TAL1 total mRNA one and a half-fold and
199 gives rise to 50% less TAL1-ΔEx3 (Fig. 2E&F) similar to activation of the -60 enhancer by
200 deleting the CTCF site (Fig. 2C&D). In addition we monitored acetylation in the TAL1 region
201 by performing ChIP with H3 pan acetylation. Our results show that tethering p300 to TAL1
202 -60 enhancer alters acetylation not only at the -60 enhancer as expected, but also at promoter
203 5/exon 3 (Fig. 2G). As we speculate that the acetylation at exon 3 could be the cause of the
204 change in alternative splicing of this exon, we specifically acetylated promoter 5/exon 3 by
205 tethering p300 to this region. We co-transfected HEK293T cells with dCas9-p300 core (mut) or
206 dCas9-p300 core (WT) and four gRNA plasmids targeting promoter 5/exon 3, which led to 9
207 times more TAL1 total mRNA (Fig. 2E). We speculate that the main increase in total mRNA
208 is from transcription beginning from promoter 5. In addition, open chromatin at this region
209 led to 80% less TAL1-ΔEx3. We hypothesize that this could be a result of reduced overall use
210 of promoters 1, 2 and 4 and reduced exclusion of exon 3 by the open chromatin in this region.

211 To gain a deeper understanding of the mechanistic link between the MuTE enhancer, the
212 broad H3K4me3 modification spanning 7.5 Kbp from TAL1 promoter 1 to promoter 5 (Fig.
213 1D), and alternative splicing, we manipulated KMT2B, which is a component of the
214 SET1/COMPASS complexes responsible for methylating H3K4 (31). Specifically, we silenced
215 KMT2B in Jurkat cells by transfecting them with siRNA targeting KMT2B or a negative control
216 (siGFP). Our results demonstrate that silencing KMT2B (Supp. Fig. S2I) led to a 30% exclusion
217 of TAL1 exon 3 (Fig. 2H and Supp. Fig. S2J). These findings provide further insights into the
218 molecular mechanisms underlying the regulation of TAL1 alternative splicing and highlight
219 KMT2B as a regulator of *TAL1* gene alternative splicing

220 The connection between the chromatin at exon 3 and alternative splicing might be via RNAPII
221 elongation kinetics (32-34), but when we slowed elongation of RNAPII in Jurkat cells using a
222 low dose of camptothecin (CPT, 6 uM) expression was reduced but splicing of TAL1 was not
223 changed (Supp. Fig. S2K-M). Another possibility is that the promoter chromatin signature at
224 promoter 5/exon 3 (H3K27ac and H3K4me3, Fig. 1D) promotes inclusion, as was suggested
225 by a genome-wide investigation (35). This study indicated that exons with promoter-like
226 chromatin tend to be included. In addition, these exons were indicated to be in a linear
227 proximity to the gene's active promoter, very similar to TAL1 exon 3 and promoters 1-4 (Fig.
228 1A). This suggests that TAL1 enhancers affect exon 3 alternative splicing regulation by
229 altering the chromatin at this exon.

230 **Characterization of TAL1 promoters and 5′ UTR**

231 To study the characteristics of TAL1 promoters, we considered the effects of the 500 nt before
232 each TSS. Since the 500 nt before TSS 1-3 are overlapping, we cloned the 500 nt before TSS3 as
233 representing all three promoters. For TSS5 we studied the sequence in exon 3 and used exon
234 4 sequence as a negative control. All sequences were cloned upstream to a minimal SV40
235 promoter in a luciferase reporter plasmid and co-transfected to HEK293T cells with renilla as
236 an internal control. We used HEK293T cells because their transfection efficiency is higher
237 compared to Jurkat cells. While exon 4 was similar to the minimal promoter alone, promoters
238 1-3 showed 3 times more luciferase expression, promoter 4 five times more and promoter 5
239 twice the amount of luciferase expression relative to control (Fig. 3A). This result suggests that
240 promoter 4 is stronger at transcription initiation relative to promoters 1-3 and 5.

241 Since TAL1 is a transcription factor and is known to autoregulate its own transcription (9, 36),
242 we wondered whether TAL1 isoforms have an effect on either of the promoters. To this end,
243 we co-transfected each of the promoters cloned before luciferase and renilla as internal control
244 with either empty vector, TAL1-long or TAL1-short (Supp. Fig. S3A&B). Our results show

245 that TAL1-short promotes expression from TSS1-3 while TAL1-long does not (Fig. 3A). This
246 result suggests that TAL1-short has a preference for the TSS1-3 promoters.

247 To study the 5′ UTR characteristics generated by each of the promoters, we cloned the 5′ UTR
248 following promoters 1, 2, 4 and 5 to the region before the luciferase initiation codon. In order
249 to assess the mRNA stability conferred by the 5′ UTR, we measured luciferase total mRNA
250 using real-time PCR relative to renilla as control. Our results did not detect any change in
251 mRNA abundance between the different 5′ UTRs and an empty control (Supp. Fig. S3C),
252 suggesting that the 5′ UTRs do not affect mRNA stability. To study the 5′ UTR effect on
253 translation, we quantified luciferase expression relative to renilla. Our results show that the
254 5′ UTR generated from promoter 4 reduced luciferase expression by 30% in contrast to the 5′
255 UTR generated by promoter 5, which gave rise to three times more luciferase expression (Fig.
256 3B). Interestingly, while promoter 4's sequence had a stronger effect on transcription, its 5′
257 UTR gave rise to reduced translation efficiency. This result indicates that the four TAL1 5′
258 UTRs have different characteristics, suggesting additional regulation to that of transcript
259 level.

260 **TAL1-short binds more strongly than TAL1-long to E-proteins**

261 Our investigation indicates complex mRNA processing, which we speculate is important to
262 give rise to a precise amount of TAL1 protein isoforms at a specific point in time during
263 hematopoiesis. To understand the importance of generating a specific amount of each
264 isoform, we began by studying the function of the protein isoforms in Jurkat cells. TAL1 is a
265 class II basic helix-loop-helix (bHLH) transcription factor that forms an obligate heterodimer
266 with the class I bHLH E-proteins, which include transcription factors E2A (TCF3) and HEB
267 (TCF12). Binding of TAL1 to E-proteins was shown to inhibit T-cell differentiation and to
268 promote T-ALL by creating an oncogenic transcription signature (37, 38). To study the
269 function of each of the protein isoforms, we transfected Jurkat cells with either TAL1-short or
270 -long with a FLAG or GFP tag. We immunoprecipitated each isoform using the tag and
271 checked for TAL1 core complex protein partners. Our results show that TAL1-short
272 immunoprecipitated more of TCF3 and TCF12 relative to TAL1-long (Fig. 4A and Supp. Fig
273 S4A). Binding of TAL1 to E-proteins prevents T-cell differentiation, which is promoted by
274 dimerization of the E-proteins, thus differential binding of TAL1 isoforms to E-proteins could
275 suggest a differential function during hematopoiesis.

276 **TAL1-short is enriched in known TAL1 binding sites**

277 TAL1 is a transcription factor and is known to bind DNA and regulate gene expression with
278 its co-regulators. To check whether each isoform has unique DNA binding characteristics, we
279 conducted chromatin immunoprecipitation followed by sequencing (ChIP-seq) for each of the
280 isoforms. All TAL1 ChIP-seq experiments to date used an antibody recognizing both isoforms
281 (36, 39). To specifically assess the binding of each of the isoforms, we transfected Jurkat cells
282 with either TAL1-short or -long with a FLAG tag and conducted ChIP-seq with anti-FLAG
283 antibody (Supp. Fig. S4B). We first examined specific known TAL1 target genes, including
284 TRIB2, NFKB1 and NKX3.1, and detected TAL1-short binding at sites in the regulatory
285 regions of each gene similar to previous results (Supp. Fig. S4D) (39). Next, we looked genome
286 wide at published ChIP-seq data with an antibody against both isoforms (39) (TAL1-total) and
287 compared it to the binding pattern of each of the isoforms. Our results indicate that binding
288 of TAL1-short is more similar to that of TAL1-total than is TAL1-long at these locations (Fig.
289 4B). The difference between the binding profiles of the isoforms might arise from stronger
290 binding of TAL1-short, and from TAL1-short having more binding sites. This is also
291 manifested in the fact that when considering all TAL1-total peaks, , the level of enrichment of
292 TAL1-short binding is 1.5 fold higher than TAL1-long (Fig. 4C and Supp. Fig. S4D).

293 The relative distribution of TAL1-total bound regions are mainly located in gene bodies and
294 intergenic regions of known protein-coding genes, consistent with the location of enhancer
295 elements (Supp. Fig. S4E). This distribution was similar to that of TAL1-short but in TAL1-
296 long we saw more binding at promoter regions and less at gene bodies (Supp. Fig. S4E).

297 We next sought to identify DNA motifs that were over-represented within 200 bp of the peak
298 of TAL1 binding in TAL1-total compared to each isoform. We identified five transcription
299 factor binding motifs to be enriched in TAL1-total and shared between one or the two
300 isoforms. Four of the transcription factors were previously described to be enriched in TAL1-
301 bound regions: E-box (5′-CAG[CG]TG-3′), GATA (5′-AGATAA-3′), RUNX (5′-TGTGGTC-3′),
302 and motifs recognized by the ETS family of transcription factors (5′-GGAA-3′) (Fig. 4D-H and
303 Supp. Fig. S4F-J). While the ETS motif was enriched in TAL1-total as well as in both isoforms,
304 the rest were only enriched in TAL1-total and TAL1-short (Fig. 4D-H and Supp Fig. S4F-J). In
305 addition, we identified a new motif recognized by the MAF family of transcription factors (5′
306 TGCTGACTCAGCA-3′) known to be expressed in T-ALL and to have a role in TAL1
307 expression (Fig. 4G and Supp. Fig. S4I). This motif was not enriched in TAL1-total but was
308 enriched in both isoforms. We speculate that the enrichment of MAF motif in both isoforms
309 and not in TAL1-total could be due to the independent conditions of these experiments (39).
310 To summarize, while our results agree with previously reported TAL1 binding sites, TAL1-

311 short exhibits stronger binding at multiple binding sites relative to TAL1-long. This result
312 taken together with TAL1-short binding more strongly to E-proteins (Fig. 4A and Fig. S4A)
313 suggest stronger transcription properties of TAL1-short relative to TAL1-long.

314 **Differential targets for TAL1 isoforms**

315 To explore the changes in expression of TAL1 targets by each of the isoforms, we conducted
316 RNA-seq in isoform-expressing cells. First we silenced endogenous TAL1 by stably
317 transfecting Jurkat cells with inducible TAL1 shRNA targeting its 3′ UTR, and so targeting
318 only the endogenous mRNA. Following treatment with tetracycline we observed 50%
319 silencing of total TAL1 mRNA (Supp. Fig S4K). We then transfected empty vector, TAL1-short
320 or TAL1-long and monitored total mRNA amount of the targets. In addition, we analyzed
321 RNA-seq results of Jurkat cells silenced for TAL1 using shRNA (40).

322 Our results identify a similar number of targets for TAL1-short (2043) as previously identified
323 for both isoforms (TAL1-total, 1696), while only 120 targets were attributed to TAL1-long (Fig.
324 4I). Previous work has shown a small overlap (14%) in genes bound by TAL1 and its
325 transcription targets (39-43) (Supp. Fig. S4L). We identified a similar overlap (17%) in our
326 results comparing results of TAL1-short ChIP-seq to RNA-seq (Supp. Fig. S4M). No overlap
327 was identified for TAL1-long, presumably due to small number of peaks and targets. We
328 chose three overlapping targets of TAL1-short for validation using real-time PCR (Supp. Fig.
329 2N). These result strengthen our hypothesis that TAL1-short is a more potent transcription
330 factor.

331 TAL1-total targets were previously shown to have a role in T-cell activation and proliferation
332 (9, 39, 41). Indeed we detect enrichment of these gene sets using the Enrichr tool (44), but only
333 for TAL1-long and not for TAL-short (Supp. Fig. S4O-Q). Interestingly, TAL1-short targets
334 were enriched for genes involved in apoptosis (Fig. 4J and Supp Fig. S4R), suggesting
335 differential cellular functions for each of the isoforms.

336 **Slower growth in TAL1-short overexpressing Jurkat cells**

337 Next, we asked whether this differential expression of apoptotic genes by the isoforms will
338 result in differential cell growth as suggested by research demonstrating that high amounts
339 of TAL1-short in glioblastoma reduced growth (45). To check the effect of the isoforms on
340 Jurkat cell growth, we silenced endogenous TAL1 by stably infecting the cells with inducible
341 TAL1 shRNA targeting its 3′ UTR. We continued by infecting FLAG-TAL1-short followed by
342 IRES expressed GFP or FLAG-TAL1-long/GFP or empty vector. Our results show faster
343 growth of TAL1-long overexpressing cells and slower growth of TAL1-short cells relative to

344 empty vector (Fig. 4K). Furthermore we counted 30% more dead-cells in TAL1-short
345 expressing cells relative to empty vector (Fig. 4L). These results could suggest that promoting
346 the expression of apoptotic genes by TAL1-short and not TAL1-long hinders cell growth.

347 **TAL1 isoforms' role in hematopoiesis**

348 To define the outcome of transcription activity of the two TAL1 isoforms, we tested their
349 influence on hematopoiesis in a mixed bone marrow chimera setting. Enriched HSCs were
350 transduced with retroviruses expressing either FLAG-TAL1-long/dtTomato or FLAG-TAL1-
351 short/GFP. Then both cells were transplanted together in lethally-irradiated syngeneic
352 recipient mice (Fig. 5A). Composition of bone marrow and spleen cells were tested by flow
353 cytometry 13 weeks after bone marrow transplantation. In comparison to WT (GFP^-,
354 $dtTomato^-$), both TAL1-long and TAL1-short-expressing spleen and bone marrow cells
355 showed decreased percentages of B and T cells (Fig. 5B&C and Supp. Fig. S5F&G) and a
356 corresponding increased percentage of $CD11b^+$ cells (Fig. 5D&E and Supp. Fig. S5H&I), as
357 was previously described (46). Similar composition of the $CD11b^+$ subsets, evaluated by Ly6G
358 and Ly6C staining, was observed in WT, TAL1-long and TAL1-short cells (Supp. Fig. S5A-B).
359 In addition, focusing on the differentiation of red blood cells lineage in the bone marrow, we
360 could observe significant decreased and increased percentages in the S0 and S3 stages,
361 respectively, in both TAL1 isoforms expressing cells in comparison to WT cells (Supp. Fig.
362 S5C-E). Overall, these results demonstrate that each of the isoforms is functional in roles of
363 TAL1 in early hematopoiesis described as supporting the myeloid over the lymphoid lineage
364 and enhancing red blood cell maturation (46).

365 **Differential role for TAL1 isoforms in hematopoiesis**

366 Having demonstrated that both TAL1 isoforms are functional in hematopoietic stem cell
367 differentiation, we moved on to ask if they possess differential roles. To this end we evaluated
368 the persistency of the TAL1-expressing cells over time. Blood samples from the mixed bone
369 marrow chimera mice were collected at different time points and the percentages of WT,
370 TAL1-long or TAL1-short cells were measured by flowcytometry (Fig. 5F&G). Three weeks
371 after bone marrow transplantation, the recipient mice had more TAL1-short than TAL1-long
372 expressing leukocytes (Fig. 5F&G). However, while the percentages of the TAL1-long
373 expressing leukocytes remained constant, those of the TAL1-short expressing leukocytes were
374 significantly decreased over time (Fig. 5F&G). These results suggest that TAL1-short and
375 TAL1-long have different effects on hematopoiesis by which TAL1-short but not TAL1-long
376 leads to hematopoietic stem cells exhaustion.

377 **TAL1-short promotes differentiation into erythroid cells**

378 TAL1-short has been shown previously to favor the erythroid lineage in chicken and human
379 cells (24). Our results also support a unique role for TAL1-short in hematopoiesis. To further
380 explore the role of the isoforms, we used K562 cells, which are capable of undergoing
381 erythroid differentiation in vitro when cultured with various inducers, such as hemin. Since
382 K562 cells express high amounts of TAL1 (Fig. 1C), we began by silencing endogenous TAL1
383 by stably infecting the cells with inducible TAL1 shRNA targeting its 3′ UTR. Following
384 treatment with tetracycline, we monitored 50% silencing of total TAL1 mRNA and protein
385 (Supp Fig. S6A&B). We continued by infecting FLAG-TAL1-short followed by IRES expressed
386 GFP or FLAG-TAL1-long/GFP or empty vector and treated cells with 20 μM hemin to induce
387 differentiation (Supp. Fig. S6C). We quantified the mRNA amount of α-, β-, and γ- globin, and
388 SLC4A1, known to be induced during erythrocyte differentiation at various time points. Our
389 results show elevated mRNA expression of α-, β- and γ- globin and SLC4A1 erythroid
390 markers in K562 cells expressing TAL1-short cells relative to empty and TAL1-long expressing
391 cells without induction of differentiation (Fig. 6A and Supp. Fig. S6D&E), or with induction
392 (Fig. 6B and Supp. Fig. S6F&G). TAL1-short's effect on uninduced cells indicates its role as a
393 differentiation factor, a role not taken by TAL1-long.

394 **TAL1-short hamper survival capability in K562 cells**

395 Our *in vivo* results suggest that TAL1-short might hamper the survival capability of
396 hematopoietic cells (Fig. 5F&G). Furthermore, TAL1-short overexpression in Jurkat cells led
397 to slower growth and a higher number of dead cells (Fig. 4K&L). To check the effect of the
398 isoforms on hematopoietic cell survival, we silenced endogenous TAL1 by stably infecting the
399 cells with inducible TAL1 shRNA targeting its 3′ UTR. Following treatment with tetracycline
400 we monitored 50% silencing of total TAL1 mRNA (Supp Fig. S6A). We continued by infecting
401 FLAG-TAL1-short followed by IRES expressed GFP or FLAG-TAL1-long/GFP or empty
402 vector. Our results show slower growth and twice as many dead cells in TAL1-short cells
403 relative to TAL1-long and empty vector (Fig. 6C&D). In addition, induction of erythropoiesis
404 using hemin resulted in similar growth rates (Supp. Fig. 6H&I). Since cells overexpressing
405 TAL1-short show the same level of erythroid differentiation markers, we cannot determine
406 the distinct effect of TAL1-short on differentiation and growth.

407 **Discussion**

408 Previous investigation of TAL1 isoforms indicated their roles in regulation of translation and
409 the generation of four isoforms (24). Using two different sets of antibodies recognizing the C-

410 terminus of the protein, common to all reported isoforms, we found two isoforms (TAL1-short
411 and TAL1-long) with post-translational modifications which affected the migration rate of the
412 protein (Fig. 1C). In contrast to a previous study (24), our results suggest regulation on the
413 mRNA level as the initiator of the two isoforms; however we found that TAL1-short promotes
414 erythroid lineage. In addition to the results in human cell lines, work in zebrafish and chicken
415 indicated that TAL1-short is required for erythroid differentiation while TAL1-long is
416 required for megakaryocytic differentiation (22).

417 Our ChIP-seq results suggest that some of the isoforms bind sites are shared (Fig. 4B). While
418 the location of the binding is similar between the isoforms, the enrichment was 1.5 higher for
419 TAL1-short relative to TAL1-long (Fig. 4C). Although TAL1's binding motif is commonly
420 reported as CAGNTG (47), we did not identify this motif in our analysis of the TAL1-total or
421 TAL1-long ChIP-seq results. However, we did observe this sequence in the TAL1-short ChIP-
422 seq results with a p-value of 1e-93. We speculate that this discrepancy may be due to the
423 complex nature of transcription factors binding and the fact that the ChIP-seq results for
424 TAL1-total were conducted independently (39). Nevertheless, we predict that both TAL1
425 isoforms bind the same DNA motif, as the basic helix-loop-helix (bHLH) domain is similar in
426 both (Fig. 1B). Quantifying the amount of TAL1-short relative to TAL1-long, we noticed that
427 while in MOLT4 the amount was equal in CCRF-CEM, Jurkat and K562 the amount of TAL1-
428 long was a magnitude higher (Fig. 1C). In addition, using the prediction tool ProtParam shows
429 that the expected half-life of TAL1-long is 7 times longer than that of TAL1-short (26). Taken
430 together our results suggest that while TAL1-short might bind more strongly to DNA, the
431 amount of TAL1-long in cells might compensate for its weaker binding enabling it to have a
432 similar effect as the short isoform.

433 Our results point at TAL1-short as a stronger protein partner of E-proteins (Fig. 4A). The
434 TAL1-E-protein dimer does not operate alone and often forms a regulatory complex with
435 other transcription factors, including LMO (LMO1 or LMO2), GATA (GATA1, GATA2 or
436 GATA3) and LDB1 proteins. In addition, RUNX1 and the ETS family of transcription factors
437 often co-occupy the same regulatory elements and coordinately regulate gene expression with
438 the TAL1 complex (39). Using ChIP-seq we identified GATA and ETS motifs at both TAL1
439 isoforms' binding locations, whereas RUNX1 and E-box were only identified for TAL1-short
440 binding sites (Fig 4D-H). While this result can help us reveal some of the complexity of the
441 isoforms' function, we need to take into account that our discovery rate is low as a result of
442 the lower enrichment of TAL1-long. The low discovery rate and complex regulation prevent
443 us from pointing at the exact role of each TAL1 isoform as a transcription factor.

444 Our previous work has demonstrated the connection between enhancer activity and
445 alternative splicing (48). The enhancer was demonstrated to promote transcription elongation
446 kinetics to promote exon inclusion. Here we show that TAL1 enhancers regulate its exon 3
447 alternative splicing. Specifically, deleting the -8 MuTE enhancer promotes exclusion of TAL1
448 exon 3, while activating TAL1 -60 enhancer via deletion of a CTCF site or tethering of dCas9-
449 p300 has the opposite result (Fig. 2B&D). This result indicating enhancer activation with exon
450 exclusion is different than the one in our previous work (48). While we could see increased
451 acetylation at TAL1 exon 3 upon enhancer activation (Fig. 2G), we could not find an effect of
452 RNAPII elongation rate on TAL1 alternative splicing (Supp. Fig. S2K-M). TAL1 -8 MuTE
453 enhancer is considered to be a strong/super enhancer (22). This type of enhancer changes the
454 chromatin not only at the gene's promoter but also at its intragenic region (Fig. 1D). Our
455 findings suggest that the broad chromatin domain marked by H3K4me3 and H3K27ac plays
456 a crucial role in regulating both alternative splicing and gene expression. Specifically, our
457 results highlight KMT2B, which is a component of the SET1/COMPASS complexes
458 responsible for the broad methylation of H3K4, as a regulator of alternative splicing of the
459 *TAL1* gene (Fig. 2H). Thus the close proximity between TAL1 promoter and exon 3 (2.5Kb),
460 as well as the fact that promoter 5 is located in exon 3 (Fig. 1A), forms open chromatin at exon
461 3 promoting inclusion of exon 3 giving rise to TAL1-long.

462 Overexpressing TAL1-long in transgenic mice promotes T-ALL in only 30% of animals after
463 a latent period of more than 14 weeks (37). In line with these results, our work did not lead to
464 T-ALL development after overexpression of either isoform. TAL1 expression is reduced in T-
465 cell development to allow activation of E-proteins. Hence overexpression of both isoforms in
466 mice resulted in a lower number of lymphocytes relative to control (Fig. 5C). While
467 lymphocyte development was affected in the same way by both isoforms, TAL1-short leads
468 to hematopoietic stem cells exhaustion while no change was observed in TAL1-long
469 expressing cells. This is a very interesting result that could have been explained
470 mechanistically by monitoring transcription regulation by the isoforms in those cells.
471 Unfortunately, at 14 weeks the small number of TAL1-short overexpressing cells did not allow
472 us to conduct RNA-seq to further investigate this isoform's targets.

473 TAL1 harbors four active promoters in Jurkat cells and thus four distinct 5' UTRs. Our results
474 point at a complex regulation of these elements orchestrated by its enhancers. In addition, our
475 results show that TAL1-short expression is a cause of Jurkat and K562 cell death while TAL1-
476 long expression does not affect cell survival (Fig. 4K&L and Fig. 6C&D). Furthermore, our
477 results indicate that TAL1-short is a stronger erythropoiesis factor, suggesting that the

478 isoforms have differential roles in hematopoiesis. These results demonstrate the importance
479 of a fine-tuned regulation of TAL1 mRNA processing, which is critical for hematopoiesis and
480 cell survival.

481 **MATERIAL AND METHODS**

482 **Ethics Statement**

483 Mice were maintained and bred under specific pathogen free conditions in the Hebrew
484 University animal facility (IACUC: Authority For Biological and Biomedical Models, NIH
485 approval number: OPRR-A01-5011) according to Institutional Animal Care and Use
486 Committee regulations. Mice experiments were approved by the Hebrew University Ethics
487 committee (approval MD-20-16296-5). At the ethical end point the mice undergo anesthesia
488 with overdose of ketamine/xylazine mixture and are euthanized by cervical dislocation.

489 **Cell lines and plasmids**

490 HEK293T (ATCC Number: CRL-3216) cells were grown in Dulbecco's modified Eagle's
491 medium (DMEM) supplemented with 10% fetal bovine serum (FBS). Jurkat (Clone E6-1),
492 MOLT4, CutLL1, CCRF-CEM, K562 cells (ATCC Number: CCL-243), Jurkat NLO and Jurkat
493 del12 were grown in Roswell Park Memorial Institute's medium (RPMI-1640) supplemented
494 with 10% FBS. HEK293T and HEK293T ΔCTCF were grown in DMEM supplemented with
495 10% FBS. Cell lines were maintained at 37°C and 5% CO_2 atmosphere. Cells were transfected
496 with TransIT-X2 transfection reagent (Mirus) following the manufacturer's instructions. After
497 30 h in culture, plasmid transfected cells were used for experimentation. Silencing KMT2B in
498 Jurkat cells was conducted by transfecting the cells with 20nM of esi KMT2B (Sigma Cat No.
499 EHU076961) using TransIT-X2 transfection reagent (Mirus) for the duration of 72h.

500 TAL1 promoters 1-3, 4, 5 and exon 4 were cloned to pGL4.23 (Promega) upstream to the
501 minimal promoter. TAL1 alternative 5'UTRs TSS1, 2, 4 and 5 were cloned after the minimal
502 promoter and before the firefly luciferase initiation codon.

503 K562 cells were induced into differentiation by incubation with 20 μM hemin and were
504 harvested after the indicated time of induction.

505 pcDNA-dCas9-p300 Core plasmids (D1399Y; plasmid #61358 and plasmid #61357) were
506 purchased from Addgene. pSPgRNA (Addgene, plasmid #47108) was used as the gRNA
507 plasmid. The oligonucleotides containing the target sequences were hybridized,
508 phosphorylated and cloned into the plasmid using BbsI sites. The target sequences are
509 provided in Supplementary Table S1.

510 **Virus production**

511 FLAG-TAL1-short/long were cloned to MIGR1 vector (Addgene, plasmid#27490). HEK293T
512 cells were co-transfected with the viral backbone vector pCMV-VSVG and pCL-Eco
513 packaging vectors using polyethylenimine (PEI) transfection reagent. Virus-containing
514 supernatants were collected 24, 48 and 72 h post-transfection following their centrifugation
515 and collection at 1500 *g* for 10 min at 4°C. Jurkat and K562 cells were infected with retroviruses
516 by the spinoculation method by centrifuging at 800 *g* for 2 h followed by 72 h incubation at
517 37°C.

518 shRNA sequence targeting TAL1 3'UTR (CATAACCACTGAAGGGAAT) was cloned to Tet-
519 pLKO-puro vector (Addgene #8453). For lentivirus production, HEK293T cells were co-
520 transfected with the viral backbone vector, pCMV-dR8.2 dvpr and pCL-Eco packaging vectors
521 using PEI transfection reagent as described above. Jurkat and K562 cells were infected with
522 lentiviruses by the spinoculation method as described above. Infected K562 and Jurkat cells
523 were selected with puromycin. shRNA expression was induced with 5 µg/mL tetracycline.

524 **qRT-PCR**

525 RNA was isolated from cells using the GENEzol TriRNA Pure Kit (GeneAid). cDNA synthesis
526 was carried out with the qScript cDNA Synthesis Kit (QuantaBio). qPCR was performed with
527 the iTaq Universal SYBR Green Supermix (BioRad) on the Biorad iCycler. The comparative Ct
528 method was employed to quantify transcripts, and delta Ct was measured in triplicate.
529 Primers used in this study are provided in Supplementary Table S1.

530 **5′ RACE**

531 cDNA was synthesized using an RT primer at Exon 4 of the TAL1 gene, containing an Illumina
532 adaptor at the 5′ end (gtctcgtgggctcggagatgtgtataagagacagCATCAGTAATCTCCATCTC).
533 This primer, in conjunction with a template switching oligo (TSO;
534 tcgtcggcagcgtcagatgtgtataagagacaGrGrGrG), generates cDNAs containing adaptor sequences
535 at both the 5′ and 3′ ends. The resulting cDNAs are directly amplified using primers for the
536 Illumina adaptor sequences at the cDNA ends.

537 The libraries and sequencing design led to R1 reads showing the 5′ end of transcripts, and R2
538 reads with the TAL1 sequence used for targeted amplification. Cutadapt software (49, 50),
539 version 2.10, was used to select reads from the expected position in the TAL1 gene, by
540 searching for the sequence **CATCAGTAATCTCCATCTC**ATAGGGGGAAGGTCTCCTC
541 (RACE primer in bold, extending into the TAL1 exon) at the beginning of R2 reads. Low
542 quality and adapter sequences were then removed, using cutadapt and fastq_quality_filter

543 (FASTX package, v0.0.14), then processed reads were aligned to the genome. Alignment was
544 done with TopHat, version 2.1.1, allowing for 2 mismatches. The genome version was GRCh38
545 (human), with annotations from Ensembl release 99. The right most position of aligned reads,
546 as TAL1 is on the reverse strand, was used to count abundance at putative TSSs.

547 **Immunoblotting**

548 For immunoblotting, cells were harvested and lysed with RIPA lysis buffer, and 20 µg/µl of
549 the extracts were run on a 10% TGX Stain-free FastCast gel (BioRad) and transferred onto a
550 nitrocellulose membrane. Antibodies used for immunoblotting were anti-TAL1
551 (MBS4381279), anti-TCF3 (CST #12258) anti-TCF12 (CST #11825) , anti-Cas9 [Danyel Biotech
552 #632628 (Clone TG8C1)] and anti-α-Tubulin (CST #3873).

553 **Immunoprecipitation (IP)**

554 Cells were washed twice with ice-cold PBS, harvested and lysed for 30 min on ice in 0.5%
555 NP40, 150 mM NaCl, 50 mM Tris pH7.5, and 2 mM MgCl2 supplemented with protease
556 inhibitor and RNAse inhibitor (Hylabs). Cell supernatants were collected after centrifugation
557 at 16, 000 g for 20 min. Anti-FLAG™ M2 Magnetic Beads (Sigma) were added for 2 h at 4°C.
558 Beads were washed four times, boiled in sample buffer and loaded onto a 10% TGX Stain-free
559 FastCast gel and transferred onto a nitrocellulose membrane and enhanced
560 chemiluminescence using Clarity Western ECL substrate (Bio-Rad) was used to visualize the
561 protein.

562 **Luciferase Reporter Gene Assay**

563 HEK293T cells were transfected using TransIT-X2 transfection reagent, according to the
564 manufacturer's protocol. Cells (15×10^4 cells/well) were seeded in 96-well plates 24 h before
565 transfection. pGL4.23 plasmid including pRL renilla luciferase reporter vector as an internal
566 control, were mixed in OptiMEM reduced serum medium (Gibco) and incubated with
567 transfection reagent for 20 min at room temperature. Transfection complexes were then gently
568 added into individual wells of the 96-well plate. 30 h after transfection, cells were harvested
569 using Promega Dual-Glo® Luciferase Assay System according to the manufacturer's protocol.
570 The relative light units (RLU) were normalized for transfection efficiency using the ratio
571 between firefly and renilla luciferase activity and the fold induction over unstimulated
572 controls was calculated. Each assay was performed in triplicate.

573 **ChIP**

574 Approximately 9 × 10^6 cells per sample were cross-linked for 10 min in 1% formaldehyde at
575 RT and quenched with 0.1 M glycine. Cells were washed twice with cold PBS and lysed with
576 lysis buffer (0.5% sodium dodecyl sulfate [SDS], 10 mM ethylenediaminetetraacetic acid
577 [EDTA], 50 mM Tris–HCl, pH 8 and 1×protease inhibitor cocktail). DNA was sonicated in an
578 ultrasonic bath (Diagenode, Bioruptor pico) to an average length of 200–250 bp. Supernatants
579 were immunoprecipitated overnight with 40 µL of precoated anti-IgG magnetic beads (goat
580 anti-rabbit IgG magnetic beads, NEB) previously incubated with the antibody of interest for
581 4 h at 4°C. The antibodies used were: rabbit anti-H3Ac, acetyl K9 + K14 + K18 + K23 + K27
582 (Abcam, cat. no. ab47915) and rabbit anti-histone H3 (Abcam, cat. no. ab1791). For ChIP-seq
583 supernatant was incubated with 20 µL Anti-FLAG™ M2 Magnetic Beads. Beads were washed
584 sequentially for 5 min each in low-salt (20 mM Tris–HCl pH 8, 150 mM NaCl, 2 mM EDTA,
585 1% Triton X-100, 0.1% SDS), high-salt (20 mM Tris–HCl pH 8, 500 mM NaCl, 2 mM EDTA, 1%
586 Triton X-100, 0.1% SDS) and LiCl buffer (10 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 1%
587 NP-40, 1% Nadeoxycholate) and in Tris-EDTA buffer. Beads were eluted in 1% SDS and 100
588 mM NaHCO3 buffer for 15 min at 65°C and cross-linking was reversed for 6 h after addition
589 of NaCl to a final concentration of 200 mM and sequentially treated with 20 µg proteinase K.
590 DNA was extracted using magnetic beads (Beckman Coulter, Agencourt AMPure XP, cat. no.
591 A63881). Immunoprecipitated DNA (2 out of 50 µL) and serial dilutions of the 10% input DNA
592 (1:5, 1:2.5, 1:1.25, and 1:0.625) were analyzed by SYBR-Green real-time qPCR. ChIP-qPCR data
593 were analyzed relative to input to include normalization for both background levels and the
594 amount of input chromatin to be used in ChIP. The primer sequences used are listed in
595 Supplemental Table S1.

596 **ChIP analysis**

597 ChIP data Fig. 1D: CutLL1: H3K27ac (51) GSM1252938, H3K4me3 (52) GSM732911. MOLT4
598 and CCRF-CEM: H3K27ac (53) GSM2037790, H3K4me3 (54) GSM4703778. Jurkat: H3K27ac
599 (16) GSM1697882, H3K4me3 (55) GSM945267. K562: H3K27ac and H3K4me3 (56) GSM733656.

600 TAL1 isoforms binding sites were identified from two biological replicates using HOMER
601 (v4.11) (57). Peak calling was performed with empty-GFP as background, using the 'factor'
602 mode and fold change ≥2. Otherwise, default parameters were used. Total TAL1 ChIP data
603 was obtained from GEO database: GSE29180 (39) and analyzed using the same pipeline and
604 parameters. Peak annotation and visualization were done using HOMER annotatePeaks tool.

605 **RNA Sequencing and analysis**

606 Total RNA samples for three biological replicates (Approximately 1×10^6 cells) were subjected
607 to sequencing.

608 Raw reads were processed for quality trimming and adaptors removal using fastx_toolkit
609 v0.0.14 and cutadapt v2.10. The processed reads were aligned to the human transcriptome
610 and genome version GRCh38 with annotations from Ensembl release 106 using TopHat v2.1.1
611 (49). Counts per gene quantification was done with htseq-count v2.01 (58). Normalization and
612 differential expression analysis were done with the DESeq2 package v 1.36.0 (59). Pair-wise
613 comparisons were tested with default parameters (Wald test), without applying the
614 independent filtering algorithm. Significance threshold was taken as padj<0.1. In addition,
615 significant DE genes were further filtered by the log2FoldChange value. This filtering was
616 baseMean-dependent and required a baseMean above 5 and an absolute log2FoldChange
617 higher than 5/sqrt(baseMean) + 0.3 (for highly expressed genes this means a requirement for
618 a fold-change of at least 1.2, while genes with a very low expression would need a 5.8 -fold
619 change to pass the filtering).

620 **Cell proliferation**

621 Suspensions of Jurkat or K562 cells in logarithmic growth at a cell density of 1.6×10^5/ml were
622 seeded in 12-well plates. Cell cultures were seeded in triplicates, and were incubated for the
623 time indicated. Cells were counted post trypan blue staining, and the mean value was
624 calculated for each time point relative to day 1.

625 **CPT treatment**

626 To impede the dynamics of transcribing RNAPII, cells were treated with camptothecin (CPT,
627 Sigma) to a final concentration of 6 µM for 6 h.

628 **Mice**

629 C57BL/6J (wild-type), and C57BL/6.SJL (PtprcaPep3b; Ly5.1) (CD45.1) mice were from The
630 Jackson Laboratory.

631 **Bone marrow transplantation**

632 Donor mice were injected IP with 150 mg/kg 5-fluorouracil (5-FU) five days before cell
633 harvesting. Bone marrow cells were extracted as previously described (60). Cells were then
634 transduced with retroviruses expressing FLAG-TAL1-short/GFP or FLAG-TAL1-long/dt-
635 tomato and incubated for 2 days at 37°C. 24 h before transplantation, recipient mice were
636 irradiated (600 rad) using an XRAD-225 machine. The irradiated recipients were then divided

637 into two groups that were injected with 2.5×10^6 bone marrow cells by tail vein injection. Mice
638 were maintained for 4 weeks post-irradiation on water containing neomycin.

639 **Retrovirus generation and bone marrow infection**

640 Virus-containing supernatants were collected 24, 48 and 72 h post-transfection following their
641 concentration in 70000 x g at 4°C using an ultracentrifuge. Infected bone marrow cells were
642 incubated with concentrated viral supernatants for 4 h in bacterial tubes and then transferred
643 back to 6-well plates for additional 48 hours.

644 **Mixed bone marrow chimera transplantation**

645 Donor mice (C57BL/6.SJL (PtprcaPep3b; Ly5.1) (CD45.1) were injected IP with 150 mg/kg 5-
646 fluorouracil (5-FU) five days before bone marrow cells harvesting. The bone marrow (BM)
647 cells were extracted as previously described (61). Cells were then transduced with retroviruses
648 expressing TAL1-short-GFP or TAL1-long-dtTomato and incubated for 48 h at 37°C. 24 h
649 before transplantation, recipient C57BL/6J (wild-type) mice were irradiated (850 rad) using
650 an XRAD-225 machine. The irradiated recipients were injected with equal amounts of bone
651 marrow cells of the two transductions by tail vein injection. Mice were maintained for 3 weeks
652 post-irradiation on water containing neomycin. Mice were monitored by blood tests every 3-
653 4 weeks for thirteen weeks after the bone-marrow transplantation.

654 **Cytokines**

655 The following cytokines were used for bone marrow culture: IL3, IL6 (20 ng/ml), stem cell
656 factor (50 ng/ml), and TPO (100 ng/ml) (Peprotech).

657 **Flow cytometry antibodies**

658 The following antibodies were used: CD90.2 (Thy 1.2) (53-2.1), CD45.1 (A20), LY6C (HK1.4),
659 LY6G(1A8), CD71 (RI7217), Ter119 (TER-119) (Biolegend).

660 **Flow cytometry detection of hematopoietic populations**

661 Blood was collected from recipient mice every 3-4 weeks post-bone marrow transplantation.
662 Only for B, T-cells and granulocytes, red blood cells were removed from the samples by
663 erythrocytes lysis buffer containing NH_4Cl, $KHCO_3$ and EDTA (Sigma). The remaining cells
664 were then re-suspended in the relevant antibody mixture diluted in PBS containing 2% FBS.
665 Cells were then analyzed using a CytoFLEX Flow Cytometer (Beckman Coulter). Analysis
666 was performed by FCS express V.6 (De Novo Software). After 14 weeks the mice were
667 euthanized and bone marrow and spleen were harvested, stained for the relevant antibodies

668 and analyzed in flow cytometry. All FCS file were uploaded to http://flowrepository.org/,
669 ID: FR-FCM-Z69L.

670

671

672 **Figure legends**

673 **Figure legends**

674 **Figure 1: TAL1 gene codes for six mRNA isoforms which translate to two protein isoforms.**
675 (A) Schematic representation of TAL1 mRNA isoforms. Rectangles: exons, black lines: introns;
676 arrow: transcription initiation site, coordinates corresponds to genome build hg19. (B)
677 Schematic representation of TAL1 protein isoforms; grey rectangles represent the protein
678 domain bHLH. (C) CutLL1, MOLT4, CCRF-CEM, Jurkat and K562 whole cell lysate was
679 extracted and subjected to Western blot analysis using TAL1 antibody recognizing both
680 isoforms and an endogenouse control (S2_Data). (D) ChIP-seq tracks for H3K27ac and
681 H3K4me3 at the TAL1 locus in the indicated cell-lines (genome build hg19).

682 **Figure 2: TAL1 enhancers promote the expression of specific isoforms.** (A and B) RNA was
683 extracted from Jurkat NLO cells, which express TAL1 exogenously and from enhancer
684 mutated cells, Jurkat Del-12 and analyzed by real-time PCR for total mRNA amount of TAL1
685 relative to CycloA and hTBP reference genes (S1_Data) (A) and for ΔEx3 relative to
686 endogenous TAL1 total mRNA amount. The proportion spliced in (PSI) represents the
687 amount of mRNA molecules that include exon 3 relative to the total number of mRNA
688 molecules that include or exclude this exon. PSI was calculated according to the formula PSI
689 = ΔCq(inclusion isoform)/ΔCq(total mRNA) (S1_Data) (B). Plots represent the mean of three
690 independent experiments (**P < 0.01, Student's t test). (C and D) RNA was extracted from
691 HEK293T cells and CTCF mutated cells, HEK293T ΔCTCF and analyzed by real-time PCR for
692 total mRNA amount of TAL1 relative to CycloA and hTBP reference genes (S1_Data) (C) and
693 for ΔEx3 relative to TAL1 total mRNA amount. PSI was calculated by ΔEx3 relative to TAL1
694 total mRNA (S1_Data) (D). Plots represents the mean of three independent experiments (****P
695 < 0.0001, Student's t test). (E-G) HEK293T cells were transfected with either dCas9-p300 core
696 (mut) or dCas9-p300 core (WT) with four gRNAs targeted to the TAL1 -60 enhancer or TAL1
697 exon 3 for 30 h. Total RNA was extracted and analyzed by real-time PCR for total mRNA
698 amount of TAL1 relative to CycloA and hTBP reference genes (S1_Data) (E) and for ΔEx3
699 relative to TAL1 total mRNA amount. PSI was calculated by ΔEx3 relative to TAL1 total
700 mRNA (S1_Data) (F). ChIP was performed of H3 pan-acetylated at the -60 enhancer and TAL1

701 exon 3 (S1_Data) (G). Values are expressed as dCas9-p300 core (WT) relative to dCas9-p300
702 core (mut) and horizontal broken lines indicate no change between dCas9-p300 core (WT)
703 relative to dCas9-p300 core (mut). Plots represents the mean of three independent
704 experiments (**P < 0.01, ***P < 0.001, ****P < 0.0001 Student's t test). (H) Jurkat cells were
705 transfected with siKMT2B or a negative control siRNA (siGFP), and RNA was extracted 72
706 hours post-transfection. Real-time PCR was performed for ΔEx3 relative to TAL1 total mRNA
707 amount. PSI was calculated by ΔEx3 relative to TAL1 total mRNA (S1_Data). Plots represents
708 the mean of five independent experiments (**P < 0.01, Student's t test).

709 **Figure 3: TAL1-short promotes transcription from TSS1-3** (A) HEK293T cells were co-
710 transfected with TAL1 promoters: promoters 1-3, 4, 5 and exon 4 as a negative control. The
711 second plasmid was an empty vector, TAL1-short or TAL1-long. After 30 h luciferase activity
712 was calculated relative to renilla. Horizontal broken lines indicate transfection with the empty
713 luciferase. The mean was calculated from three independent biological experiments, each
714 performed with three technical replicates (S1_Data, **P < 0.01, Student's t test). (B) HEK293T
715 cells were co-transfected with TAL1 5′ UTRs: TSS 1, 2, 4 and 5 and empty vector as negative
716 control. After 30 h luciferase activity was calculated relative to renilla. The mean was
717 calculated from three independent biological experiments, each performed with three
718 technical replicates (S1_Data, *P < 0.05, Student's t test) are shown.

719 **Figure 4: TAL1-short promotes transcription of apoptotic genes and is a stronger**
720 **transcription factor compared to TAL1-long.** (A-H) Jurkat cells were infected with empty
721 vector, FLAG-TAL1-short or FLAG-TAL1-long. FLAG was immunoprecipitated and known
722 TAL1 interacting proteins were detected with indicated antibodies (S2_Data) (A). (B-H)
723 FLAG-TAL1-short, TAL1-long and and empty vector's chromatin were immunoprecipitated
724 using FLAG antibody. In addition, TAL1-total ChIP-seq data was analyzed (see methods).
725 Heatmap showing signal enrichment over 5876 TAL1-total binding sites (±1000 kb from the
726 center). Heatmap is sorted by strength of TAL1-short signal centered on ChIP-seq summits.
727 A color scale indicating the relative signal intensity plotted on each heatmap is shown (B).
728 ChIP-seq average signal for TAL1-total, TAL1-short and TAL1-long and empty vector as a
729 function of distance from TAL1-total peaks (S1_Data) (C). Most abundant DNA sequence
730 motifs were identified in TAL1-total, TAL1-short and TAL1-long ChIP-seq peaks. Percentage
731 of loci was calculated by dividing % of target sequences with motifs by % of background
732 sequences with the motif for ETS (D), GATA (E), RUNX(F), MAF (G) and E-box (H) (S1_Data).
733 (I-L) Jurkat cells were infected with inducible shRNA against the 3′ UTR of TAL1. In addition,
734 the cells were infected with either empty vector, TAL1-short or TAL1-long. Following

735 induction with tetracycline for 72 h, RNA was extracted and subjected to sequencing. In
736 addition analysis was performed on available data of TAL1-total (see methods). Venn
737 diagram representing the overlap of TAL1-short, TAL1-long and TAL-total target genes (I).
738 Gene Set Enrichment Analysis (GSEA) for apoptotic pathway for TAL1-short and TAL1-long
739 (S1_Data) (J). Cells were seeded and counted every day following labeling by trypan blue.
740 Live cells are plotted in (S1_Data) (K) trypan blue labeled cells are plotted in (S1_Data) (L).
741 Plots represents the mean of three independent experiments (*<0.05, ***P<0.001 and
742 ****P<0.0001 Student's t test).

743 **Figure 5: TAL1-short but not TAL1-long leads to hematopoietic stem cell exhaustion.** (A)
744 Schematic illustration of the mixed bone marrow chimera experiments. Equal numbers of
745 bone marrow cells from 5FU-treated CD45.1 wild-type mice were transduced with
746 retroviruses expressing either TAL1-short-GFP or with TAL1-long-dtTomato. A mixture of
747 1:1 ratio of the transduced bone marrow cells was then transplanted into lethally-irradiated
748 CD45.2 wild-type recipient mice. Blood samples were taken at different time points to monitor
749 the progression of the bone marrow cell reconstitution using flow cytometry analysis. All mice
750 were sacrificed fourteen weeks after BMT and the spleen and bone marrow were harvested
751 and analyzed. (B) Representative dot plots of Thy1.2 vs. CD19 staining in untransduced (left),
752 dtTomato+ (middle), or GFP+ (right) splenocytes from the recipient mice (S1_Data) (C) Bar
753 graph summarizing results shown in B. (D) Representative flow cytometry histograms of
754 CD11b staining gated on GFP-/dtTomato- (untransduced, left panel), dtTomato+ (Tal1-long,
755 middle panel) or GFP+ (Tal1-short, right panel) splenocytes (S1_Data). (E) Bar graph
756 summarizing results shown in D (S1_Data). (F) Representative dot plots of Tal1-short GFP vs.
757 Tal1-long dtTomato staining in the blood of the recipient mice at different time points over
758 the course of thirteen weeks. (G) Bar graph summarizing results in F (S1_Data). Each mouse
759 is presented with a different shade. In C, two-tailed paired t-test, *<0.05, **<0.01, sd. *(n=6).*

760 **Figure 6: TAL1-short is a stronger differentiation agent. A-D** K562 cells were infected with
761 inducible shRNA against the 3' UTR of TAL1. In addition, the cells were infected with MIGR1
762 plasmid with GFP at the C-terminal followed by an IRES element. Infection was with empty
763 vector, TAL1-short or TAL1-long. Following induction with tetracycline for 72 h, cells were
764 treated with 20 μM hemin to promote erythroid differentiation for the indicated time points.
765 RNA was extracted and analyzed by real-time PCR for total mRNA amount of α β, γ globin
766 and SLCA4 relative to CycloA and hTBP reference genes without hemin (S1_Data) (A) and
767 with treatment (inducing differentiation) for 48 h (S1_Data) (B). Cells were seeded and
768 counted every day following labeling by trypan blue. Live cells are plotted in (S1_Data) (C)

769 and trypan blue labeled cells are plotted in (S1_Data) (D). Plots represents the mean of three
770 independent experiments (*<0.05, **P < 0.01, ***P<0.001 and ****P<0.0001 Student's t test).

771 **Acknowledgements**

772 We thank Roi Gazit and Ido Goldstein for their expert advice throughout this work. We thank
773 Thomas Look for Jurkat NLO and Jurkat Del-12 cells. We thank Richard A. Young for
774 HEK293T parental and HEK293T ΔCTCF cells.

775 **Data availability statement**

776 The data supporting the findings of this study are available within the article and its
777 supplementary material. The data discussed in this publication have been deposited in NCBI's
778 Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series
779 accession number GSE214833 (RNA-seq)
780 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE214833) and accession number
781 GSE216684 (ChIP-seq)
782 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE216684). All FCS file were
783 uploaded to http://flowrepository.org/, ID: FR-FCM-Z69L.

784

785

786 **Funding**

787 This work was supported by the Israel Science Foundation (ISF 462/22), Israeli Cancer
788 Association, Israel Cancer Research Foundation and the Alon Award from the Israeli Planning
789 and Budgeting Committee (PBC). AS's research scholarship is partially funded by TEVA's
790 bioInnovation program. MB's research scholarship is partially supported by the Carole and
791 Andrew Harper Diversity Program, as well as TEVA's bioInnovation program. MB's research
792 scholarship is partially funded by the Kaete Klasuner Scholarship.

793

794 **Conflicts of interest**

795 The authors declare no conflicts of interest

796 **References**

797 1. R. A. Shivdasani, E. L. Mayer, S. H. Orkin, Absence of blood formation in mice
798 lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373**, 432-434 (1995).
799 2. L. Robb *et al.*, Absence of yolk sac hematopoiesis from mice with a targeted
800 disruption of the scl gene. *Proc Natl Acad Sci U S A* **92**, 7075-7079 (1995).

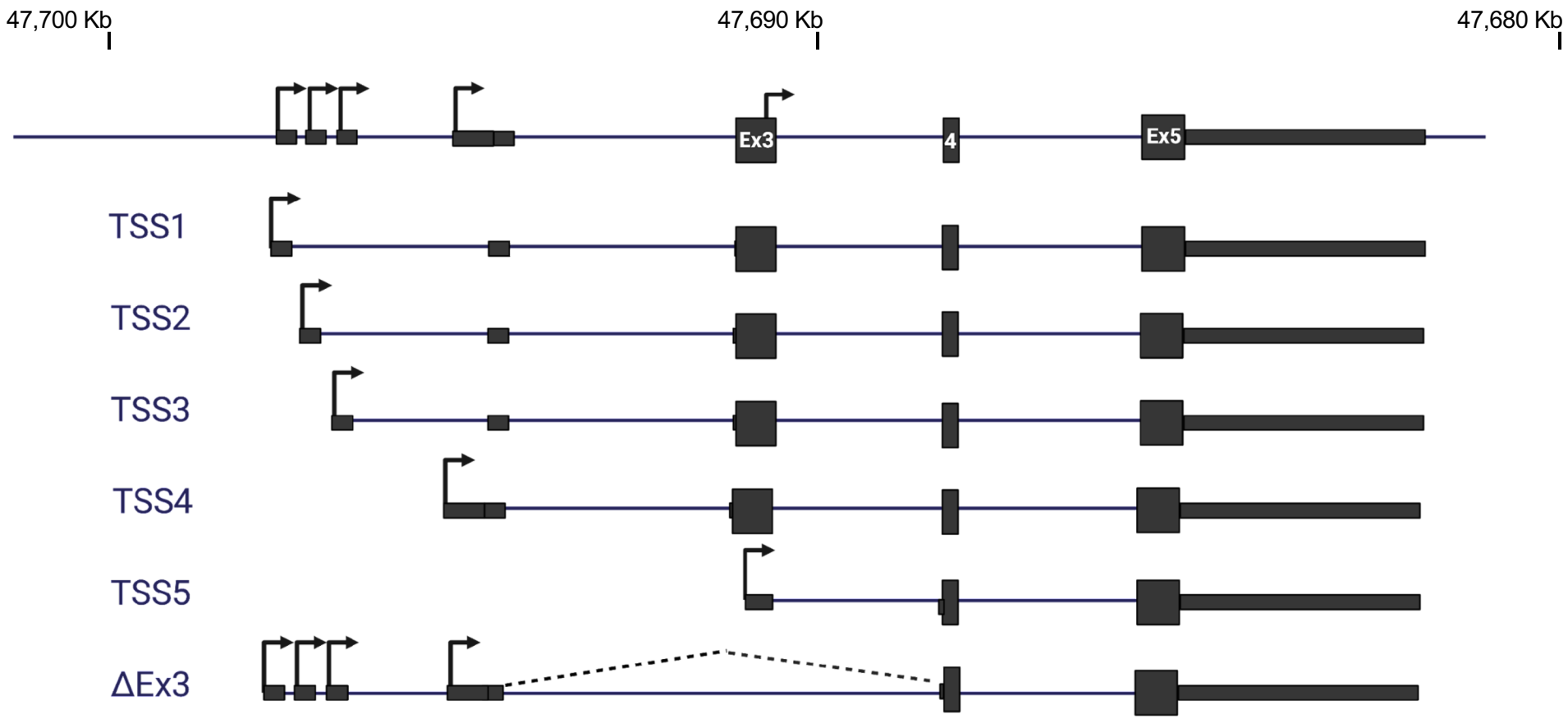
801 3. N. C. Correia, M. L. Arcangeli, F. Pflumio, J. T. Barata, Stem Cell Leukemia: how a
802 TALented actor can go awry on the hematopoietic stage. *Leukemia* **30**, 1968-1978
803 (2016).
804 4. C. Porcher *et al.*, The T cell leukemia oncoprotein SCL/tal-1 is essential for
805 development of all hematopoietic lineages. *Cell* **86**, 47-57 (1996).
806 5. L. Robb *et al.*, The scl gene product is required for the generation of all hematopoietic
807 lineages in the adult mouse. *EMBO J* **15**, 4123-4129 (1996).
808 6. M. A. Hall *et al.*, The critical regulator of embryonic hematopoiesis, SCL, is vital in
809 the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. *Proc*
810 *Natl Acad Sci U S A* **100**, 992-997 (2003).
811 7. C. G. Begley, A. R. Green, The SCL gene: from case report to critical hematopoietic
812 regulator. *Blood* **93**, 2760-2770 (1999).
813 8. A. T. Look, Oncogenic transcription factors in the human acute leukemias. *Science*
814 **278**, 1059-1064 (1997).
815 9. T. K. Tan, C. Zhang, T. Sanda, Oncogenic transcriptional program driven by TAL1 in
816 T-cell acute lymphoblastic leukemia. *Int J Hematol* **109**, 5-17 (2019).
817 10. O. Bernard *et al.*, A third tal-1 promoter is specifically used in human T cell
818 leukemias. *J Exp Med* **176**, 919-925 (1992).
819 11. E. O. Bockamp *et al.*, Lineage-restricted regulation of the murine SCL/TAL-1
820 promoter. *Blood* **86**, 1502-1514 (1995).
821 12. E. O. Bockamp *et al.*, Distinct mechanisms direct SCL/tal-1 expression in erythroid
822 cells and CD34 positive primitive myeloid cells. *J Biol Chem* **272**, 8781-8790 (1997).
823 13. C. Courtes *et al.*, Erythroid-specific inhibition of the tal-1 intragenic promoter is due
824 to binding of a repressor to a novel silencer. *J Biol Chem* **275**, 949-958 (2000).
825 14. M. R. Mansour *et al.*, Oncogene regulation. An oncogenic super-enhancer formed
826 through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373-1377
827 (2014).
828 15. D. Hnisz *et al.*, Super-enhancers in the control of cell identity and disease. *Cell* **155**,
829 934-947 (2013).
830 16. D. Hnisz *et al.*, Activation of proto-oncogenes by disruption of chromosome
831 neighborhoods. *Science* **351**, 1454-1458 (2016).
832 17. B. Gottgens *et al.*, Analysis of vertebrate SCL loci identifies conserved enhancers. *Nat*
833 *Biotechnol* **18**, 181-186 (2000).
834 18. M. Sanchez *et al.*, An SCL 3' enhancer targets developing endothelium together with
835 embryonic and adult haematopoietic progenitors. *Development* **126**, 3891-3904 (1999).
836 19. A. M. Sinclair *et al.*, Distinct 5' SCL enhancers direct transcription to developing
837 brain, spinal cord, and endothelium: neural expression is mediated by GATA factor
838 binding sites. *Dev Biol* **209**, 128-142 (1999).
839 20. S. Jin *et al.*, Splicing factor SF3B1K700E mutant dysregulates erythroid differentiation
840 via aberrant alternative splicing of transcription factor TAL1. *PLoS One* **12**, e0175523
841 (2017).
842 21. F. Qian *et al.*, Distinct functions for different scl isoforms in zebrafish primitive and
843 definitive hematopoiesis. *PLoS Biol* **5**, e132 (2007).
844 22. F. Zhen, Y. Lan, B. Yan, W. Zhang, Z. Wen, Hemogenic endothelium specification
845 and hematopoietic stem cell maintenance employ distinct Scl isoforms. *Development*
846 **140**, 3977-3985 (2013).
847 23. X. Ren, G. A. Gomez, B. Zhang, S. Lin, Scl isoforms act downstream of etsrp to
848 specify angioblasts and definitive hematopoietic stem cells. *Blood* **115**, 5338-5346
849 (2010).
850 24. C. F. Calkhoven *et al.*, Translational control of SCL-isoform expression in
851 hematopoietic lineage choice. *Genes Dev* **17**, 959-964 (2003).

852 25. K. Pulford *et al.*, Expression of TAL-1 proteins in human tissues. *Blood* **85**, 675-684
853 (1995).
854 26. M. R. Wilkins *et al.*, Protein identification and analysis tools in the ExPASy server.
855 *Methods Mol Biol* **112**, 531-552 (1999).
856 27. J. D. Curry, M. T. Smith, Measurement of SIL-TAL1 fusion gene transcripts
857 associated with human T-cell lymphocytic leukemia by real-time reverse
858 transcriptase-PCR. *Leuk Res* **27**, 575-582 (2003).
859 28. K. Chen *et al.*, Broad H3K4me3 is associated with increased transcription elongation
860 and enhancer activity at tumor-suppressor genes. *Nat Genet* **47**, 1149-1157 (2015).
861 29. M. Salton, T. C. Voss, T. Misteli, Identification by high-throughput imaging of the
862 histone methyltransferase EHMT2 as an epigenetic regulator of VEGFA alternative
863 splicing. *Nucleic Acids Res* **42**, 13662-13673 (2014).
864 30. I. B. Hilton *et al.*, Epigenome editing by a CRISPR-Cas9-based acetyltransferase
865 activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510-517 (2015).
866 31. T. H. Beacon *et al.*, The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a
867 mark of essential genes. *Clin Epigenetics* **13**, 138 (2021).
868 32. P. Cramer, C. G. Pesce, F. E. Baralle, A. R. Kornblihtt, Functional association between
869 promoter structure and transcript alternative splicing. *Proc Natl Acad Sci U S A* **94**,
870 11456-11460 (1997).
871 33. G. C. Roberts, C. Gooding, H. Y. Mak, N. J. Proudfoot, C. W. Smith, Co-
872 transcriptional commitment to alternative splice site selection. *Nucleic Acids Res* **26**,
873 5568-5572 (1998).
874 34. S. Naftelberg, I. E. Schor, G. Ast, A. R. Kornblihtt, Regulation of alternative splicing
875 through coupling with transcription and chromatin structure. *Annu Rev Biochem* **84**,
876 165-198 (2015).
877 35. J. Curado, C. Iannone, H. Tilgner, J. Valcarcel, R. Guigo, Promoter-like epigenetic
878 signatures in exons displaying cell type-specific splicing. *Genome Biol* **16**, 236 (2015).
879 36. M. T. Kassouf *et al.*, Genome-wide identification of TAL1's functional targets: insights
880 into its mechanisms of action in primary erythroid cells. *Genome Res* **20**, 1064-1083
881 (2010).
882 37. J. O'Neil, J. Shank, N. Cusson, C. Murre, M. Kelliher, TAL1/SCL induces leukemia
883 by inhibiting the transcriptional activity of E47/HEB. *Cancer Cell* **5**, 587-596 (2004).
884 38. A. N. Goldfarb, K. Lewandowska, Inhibition of cellular differentiation by the
885 SCL/tal oncoprotein: transcriptional repression by an Id-like mechanism. *Blood* **85**,
886 465-471 (1995).
887 39. T. Sanda *et al.*, Core transcriptional regulatory circuit controlled by the TAL1
888 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209-221
889 (2012).
890 40. W. Z. Leong *et al.*, ARID5B as a critical downstream target of the TAL1 complex that
891 activates the oncogenic transcriptional program and promotes T-cell
892 leukemogenesis. *Genes Dev* **31**, 2343-2360 (2017).
893 41. C. G. Palii *et al.*, Differential genomic targeting of the transcription factor TAL1 in
894 alternate haematopoietic lineages. *EMBO J* **30**, 494-509 (2011).
895 42. S. H. Tan *et al.*, TRIB2 reinforces the oncogenic transcriptional program controlled by
896 the TAL1 complex in T-cell acute lymphoblastic leukemia. *Leukemia* **30**, 959-962
897 (2016).
898 43. T. Sanda, W. Z. Leong, TAL1 as a master oncogenic transcription factor in T-cell
899 acute lymphoblastic leukemia. *Exp Hematol* **53**, 7-15 (2017).
900 44. E. Y. Chen *et al.*, Enrichr: interactive and collaborative HTML5 gene list enrichment
901 analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

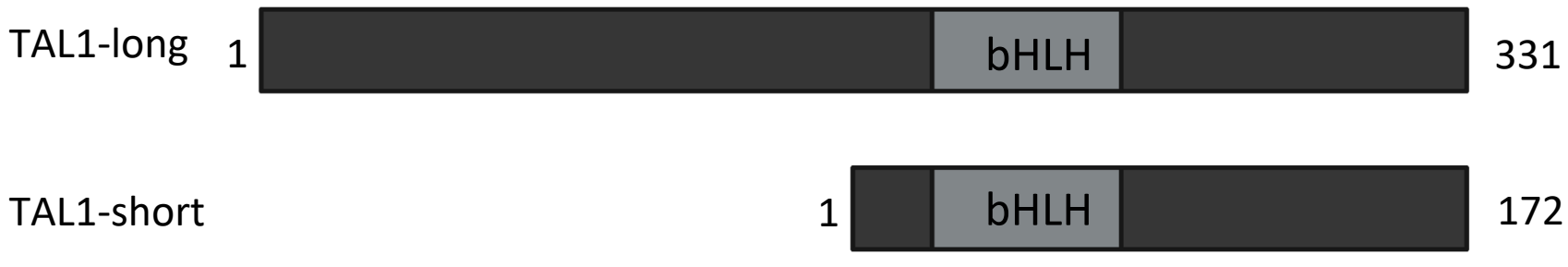
45. S. Guelfi *et al.*, SLUG and Truncated TAL1 Reduce Glioblastoma Stem Cell Growth Downstream of Notch1 and Define Distinct Vascular Subpopulations in Glioblastoma Multiforme. *Cancers (Basel)* **13** (2021).
46. N. Goardon *et al.*, Ectopic expression of TAL-1 protein in Ly-6E.1-htal-1 transgenic mice induces defects in B- and T-lymphoid differentiation. *Blood* **100**, 491-500 (2002).
47. H. L. Hsu *et al.*, Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. *Mol Cell Biol* **14**, 1256-1265 (1994).
48. S. Dahan *et al.*, VEGFA's distal enhancer regulates its alternative splicing in CML. *NAR Cancer* **3**, zcab029 (2021).
49. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
50. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
51. H. Wang *et al.*, NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. *Proc Natl Acad Sci U S A* **111**, 705-710 (2014).
52. H. Wang *et al.*, Genome-wide analysis reveals conserved and divergent features of Notch1/RBPJ binding in human and murine T-lymphoblastic leukemia cells. *Proc Natl Acad Sci U S A* **108**, 14908-14913 (2011).
53. B. J. Abraham *et al.*, Small genomic insertions form enhancers that misregulate oncogenes. *Nat Commun* **8**, 14385 (2017).
54. L. K. Gopi, B. L. Kidder, Integrative pan cancer analysis reveals epigenomic variation in cancer type and cell specific chromatin domains. *Nat Commun* **12**, 1419 (2021).
55. R. E. Thurman *et al.*, The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82 (2012).
56. B. D. Pope *et al.*, Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402-405 (2014).
57. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
58. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
59. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
60. O. M. Siggs *et al.*, The P4-type ATPase ATP11C is essential for B lymphopoiesis in adult bone marrow. *Nat Immunol* **12**, 434-440 (2011).
61. A. Goldshtein *et al.*, Loss of T-cell quiescence by targeting Slfn2 prevents the development and progression of T-ALL. *Oncotarget* **7**, 46835-46847 (2016).

Fig. 1

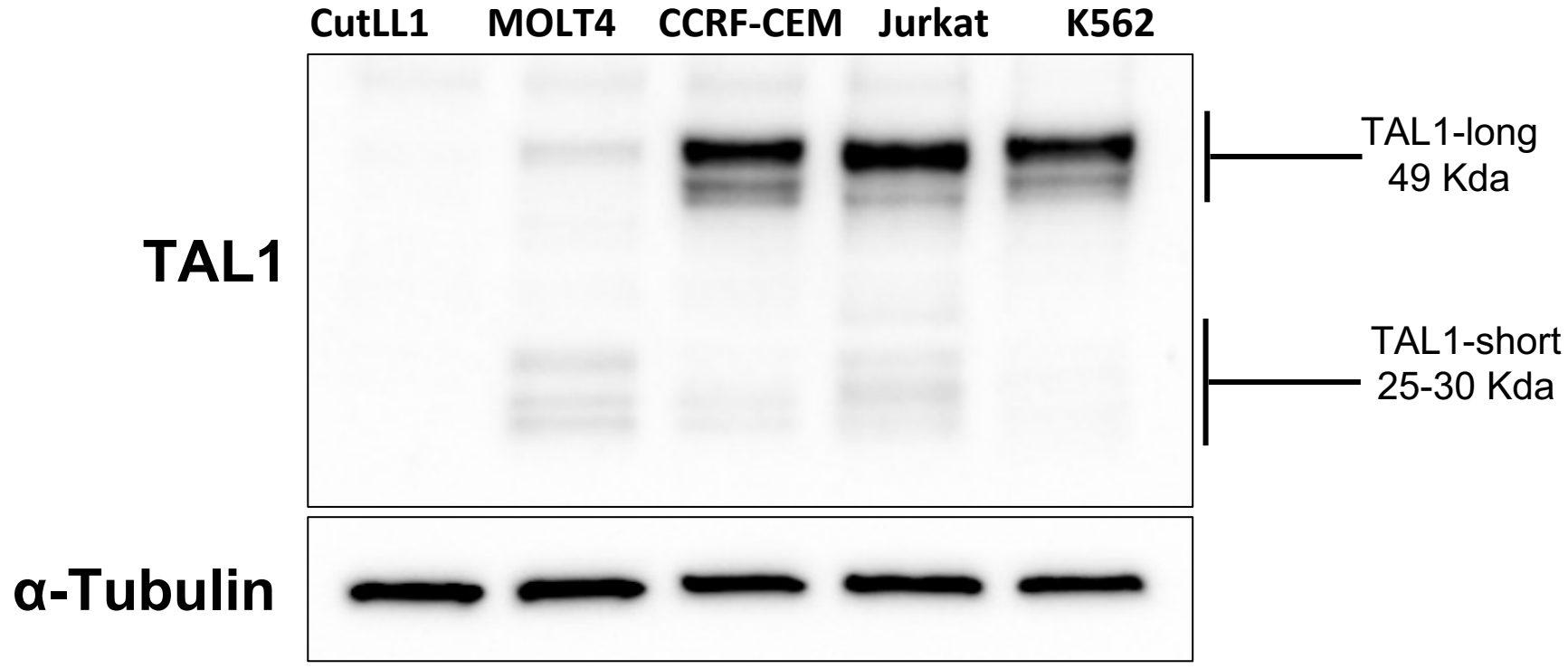
A


47,700 Kb
47,690 Kb
47,680 Kb
Ex3
4
Ex5
TSS1
TSS2
TSS3
TSS4
TSS5
ΔEx3

B


TAL1-long
1
bHLH
331
TAL1-short
1
bHLH
172

C


CutLL1
MOLT4
CCRF-CEM
Jurkat
K562
TAL1
TAL1-long
49 Kda
TAL1-short
25-30 Kda
α-Tubulin

D

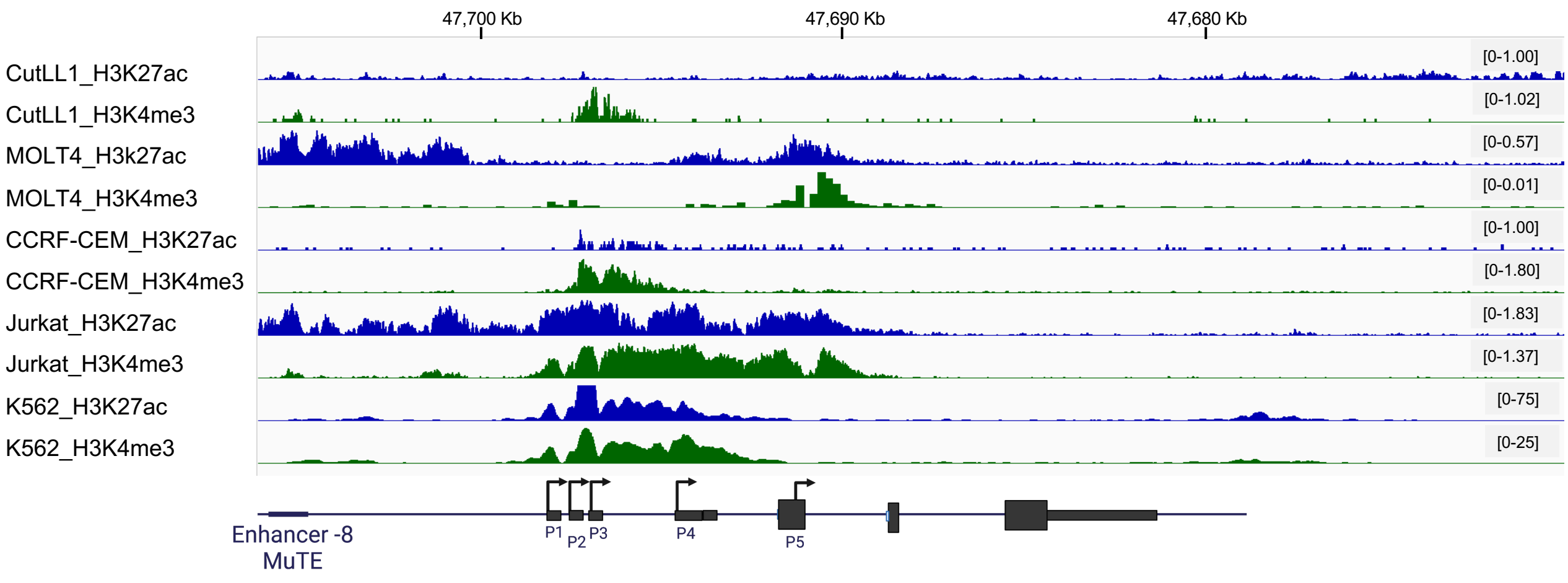
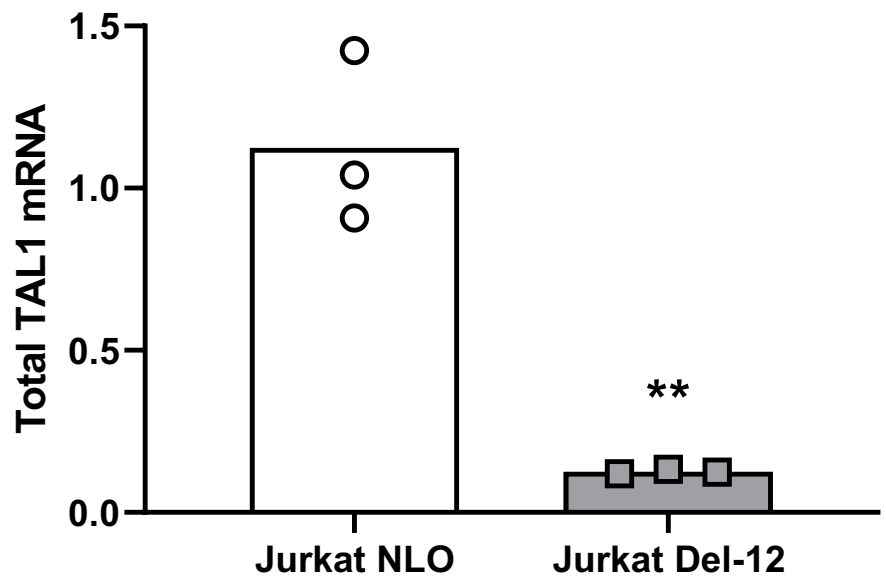
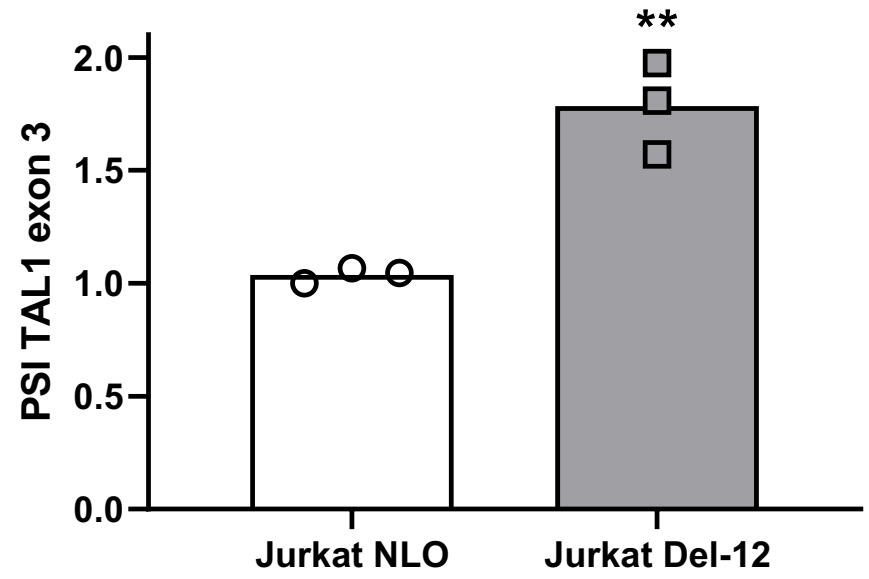

47,700 Kb
47,690 Kb
47,680 Kb
CutLL1_H3K27ac
CutLL1_H3K4me3
MOLT4_H3k27ac
MOLT4_H3K4me3
CCRF-CEM_H3K27ac
CCRF-CEM_H3K4me3
Jurkat_H3K27ac
Jurkat_H3K4me3
K562_H3K27ac
K562_H3K4me3
[0-1.00]
[0-1.02]
[0-0.57]
[0-0.01]
[0-1.00]
[0-1.80]
[0-1.83]
[0-1.37]
[0-75]
[0-25]
Enhancer -8
MuTE
P1
P2
P3
P4
P5

Fig. 2

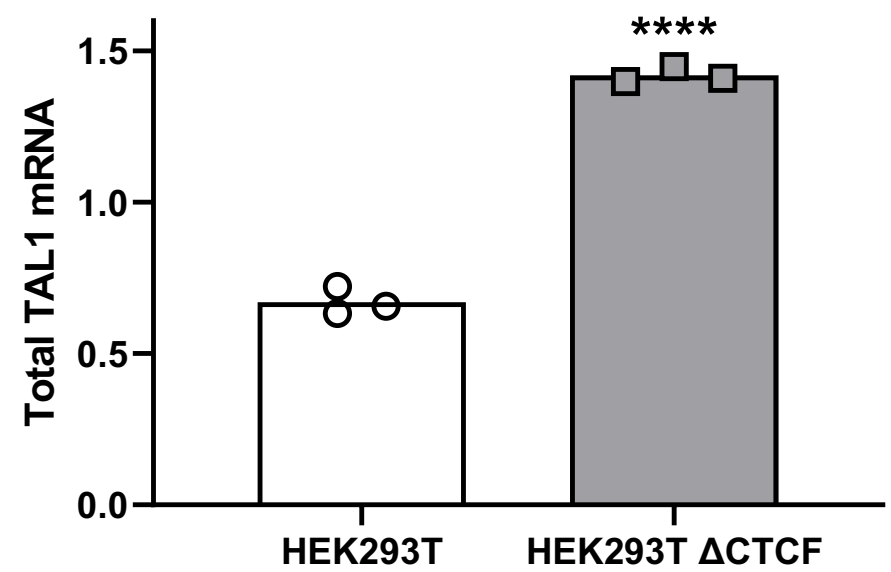
A



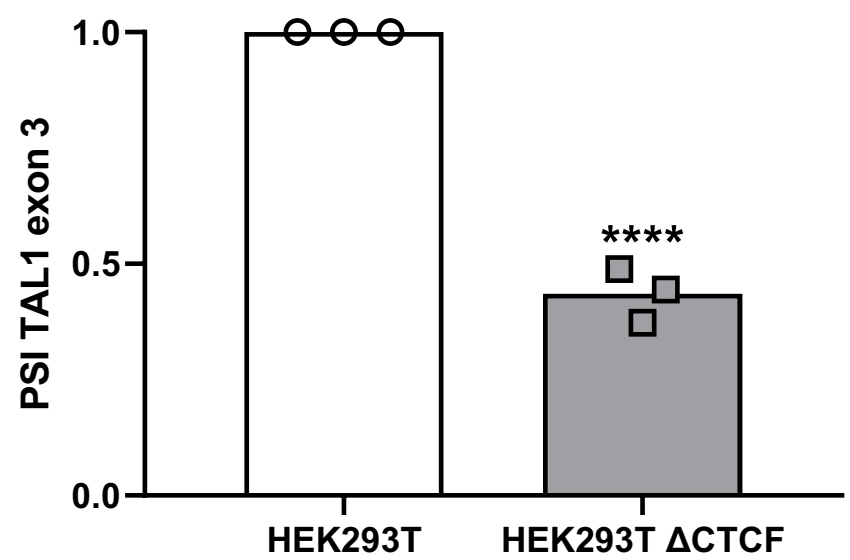
B



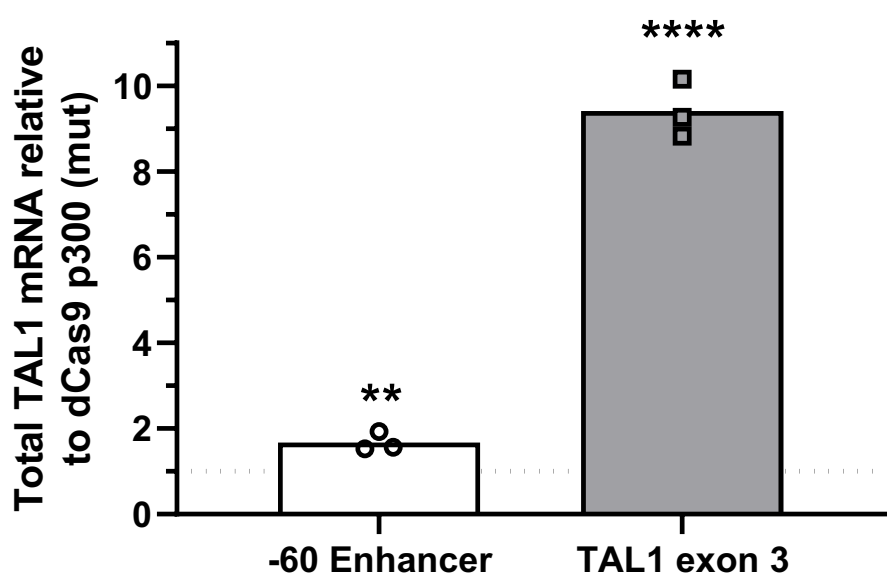
C



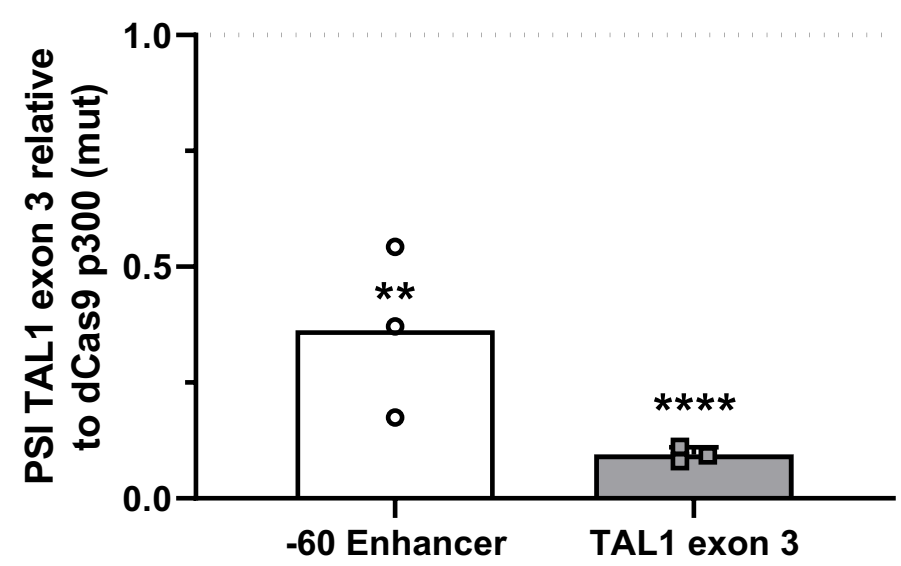
D



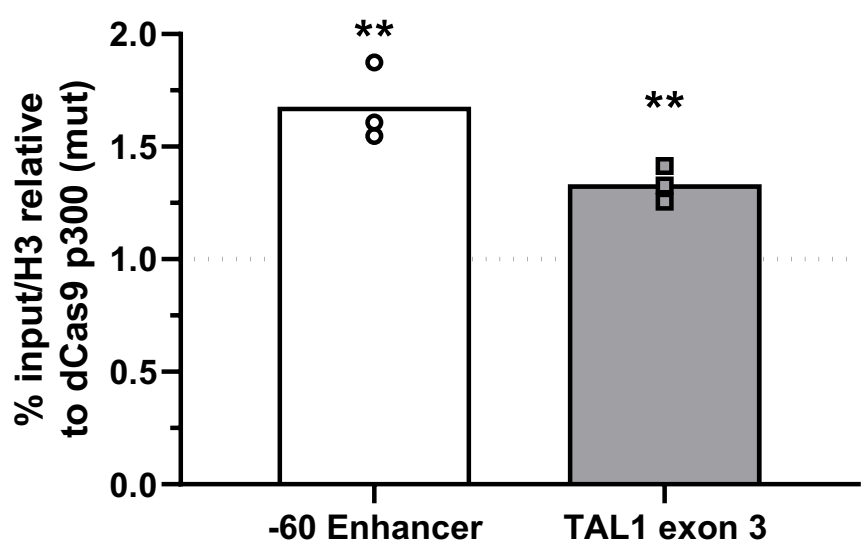
E



F



G



H

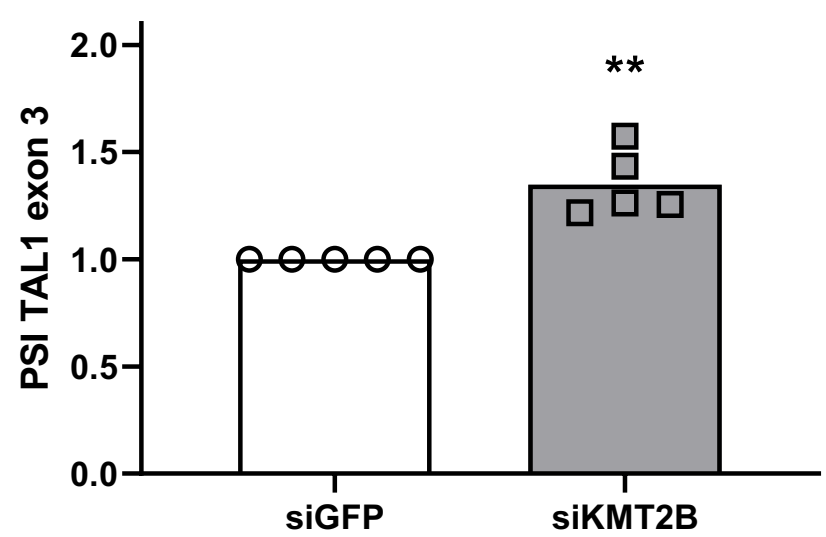


Fig. 3

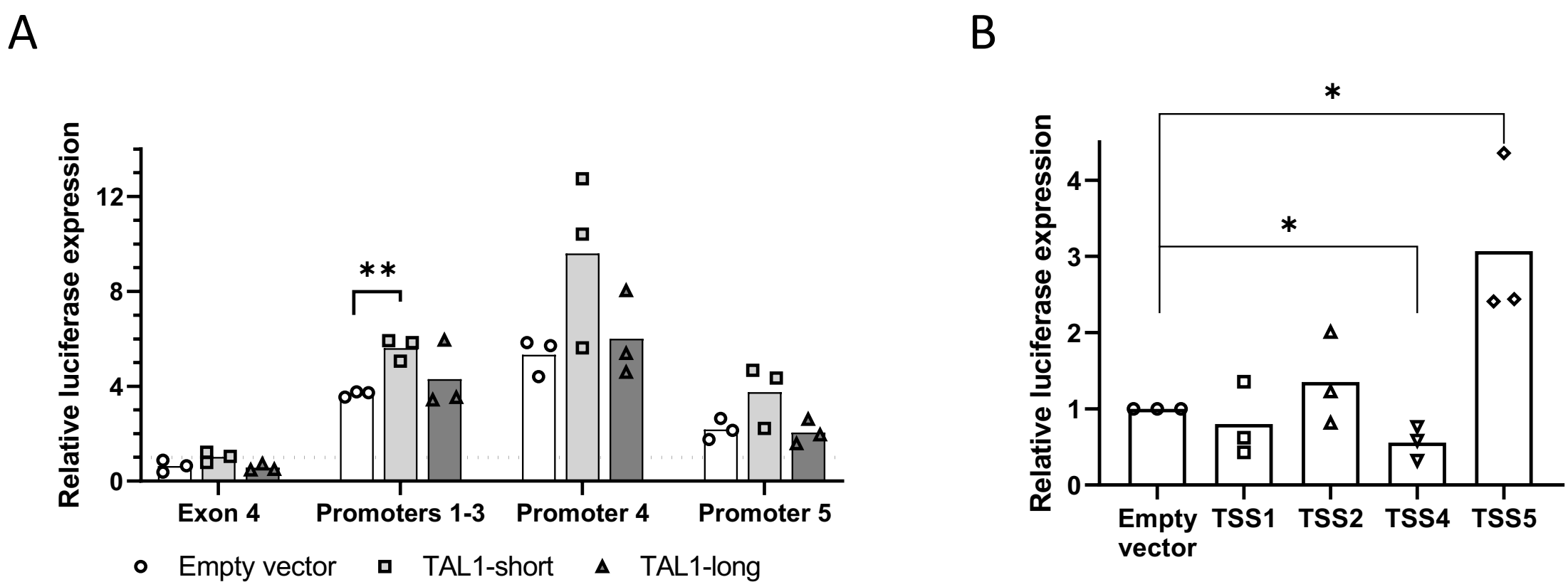

A
Relative luciferase expression
12
8
4
0
**
Exon 4
Promoters 1-3
Promoter 4
Promoter 5
Empty vector
TAL1-short
TAL1-long
B
Relative luciferase expression
4
3
2
1
0
*
*
Empty vector
TSS1
TSS2
TSS4
TSS5

Fig. 4

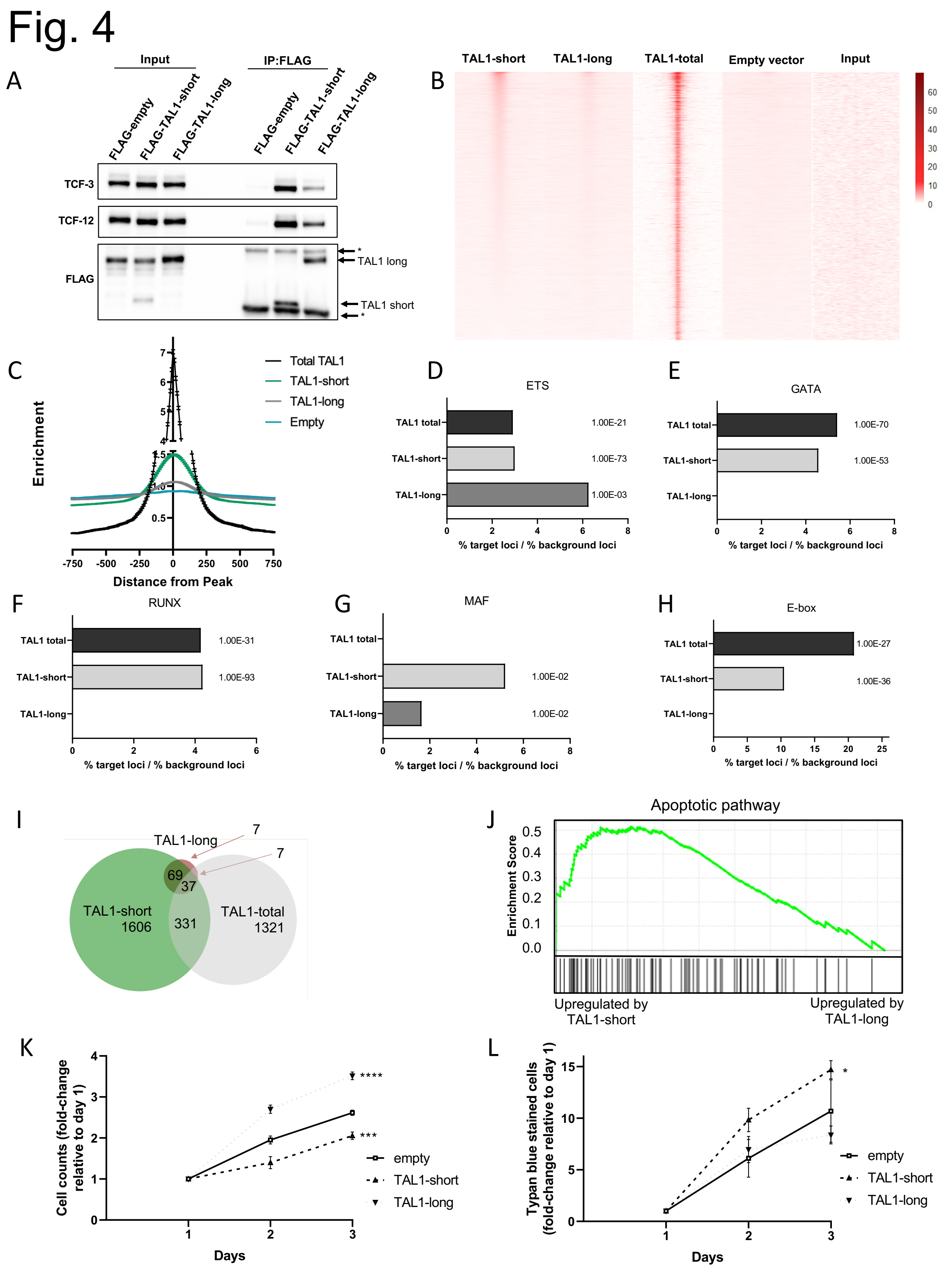
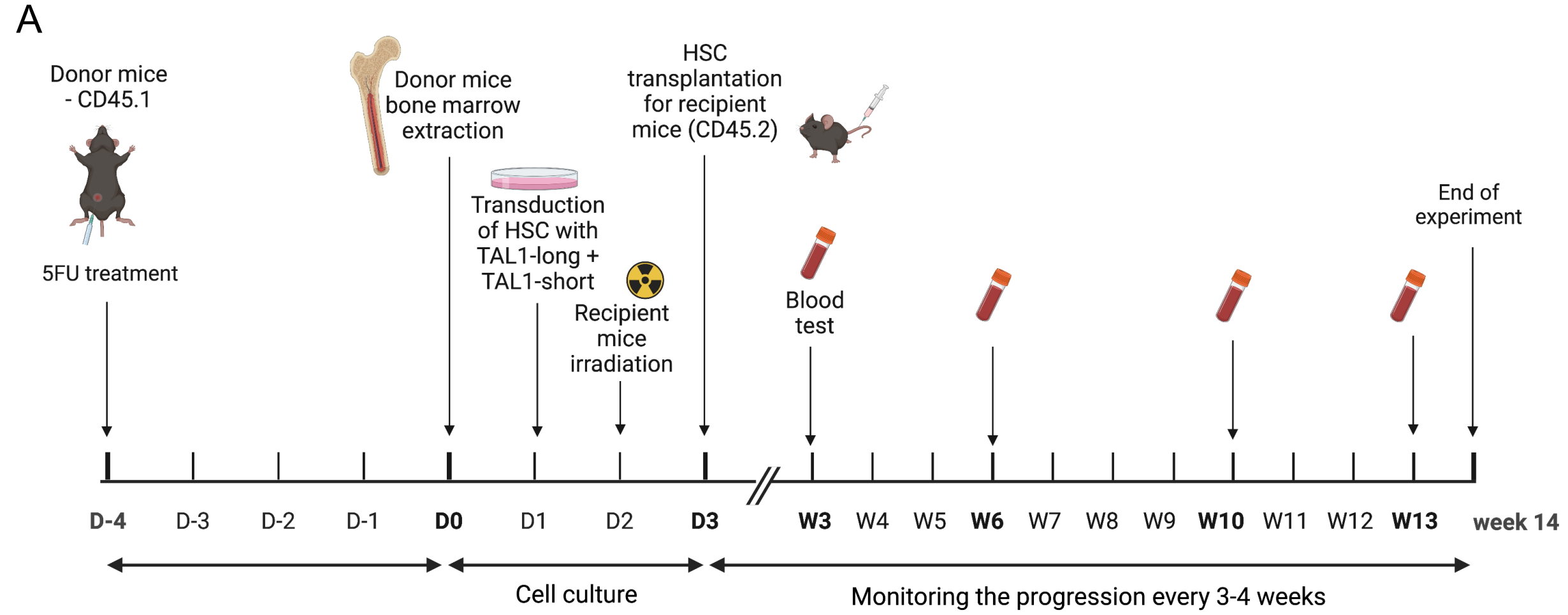
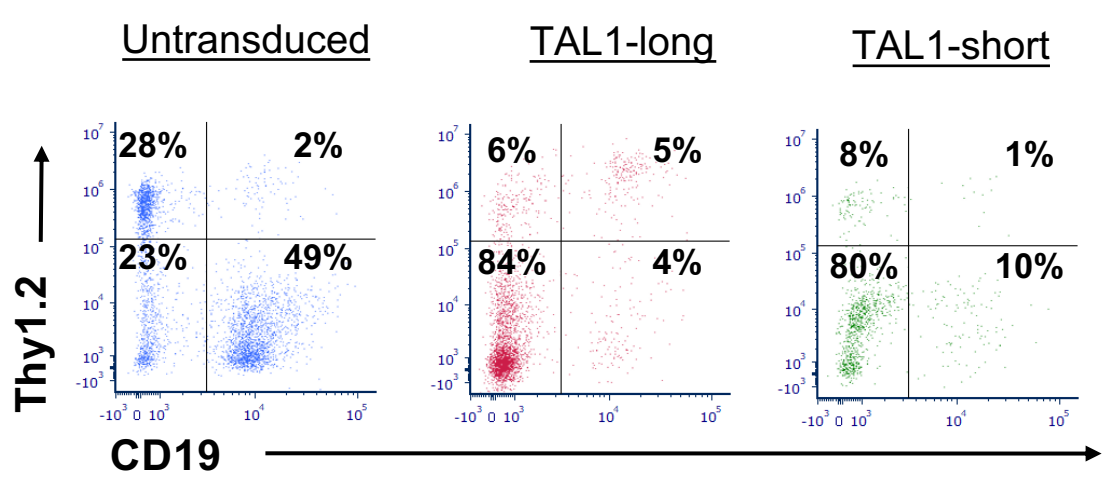

A
Input
IP:FLAG
FLAG-empty
FLAG-TAL1-short
FLAG-TAL1-long
TCF-3
TCF-12
FLAG
*
TAL1 long
TAL1 short
*
B
TAL1-short
TAL1-long
TAL1-total
Empty vector
Input
60
50
40
30
20
10
0
C
Total TAL1
TAL1-short
TAL1-long
Empty
Enrichment
Distance from Peak
D
ETS
TAL1 total
TAL1-short
TAL1-long
1.00E-21
1.00E-73
1.00E-03
% target loci / % background loci
E
GATA
1.00E-70
1.00E-53
F
RUNX
1.00E-31
1.00E-93
G
MAF
1.00E-02
1.00E-02
H
E-box
1.00E-27
1.00E-36
I
TAL1-long
7
7
69
37
TAL1-short
1606
331
TAL1-total
1321
J
Apoptotic pathway
Enrichment Score
Upregulated by TAL1-short
Upregulated by TAL1-long
K
Cell counts (fold-change relative to day 1)
****
***
empty
TAL1-short
TAL1-long
Days
L
Typan blue stained cells (fold-change relative to day 1)
*
empty
TAL1-short
TAL1-long
Days

Fig. 5

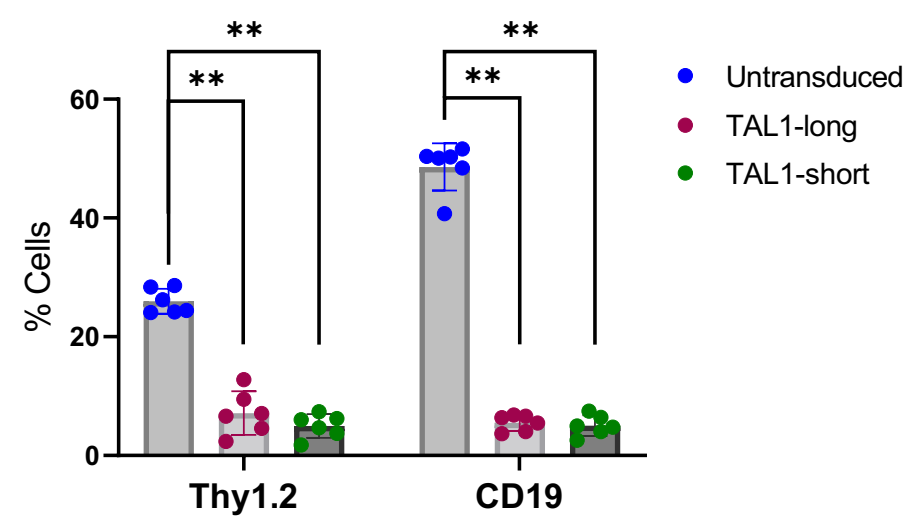
A



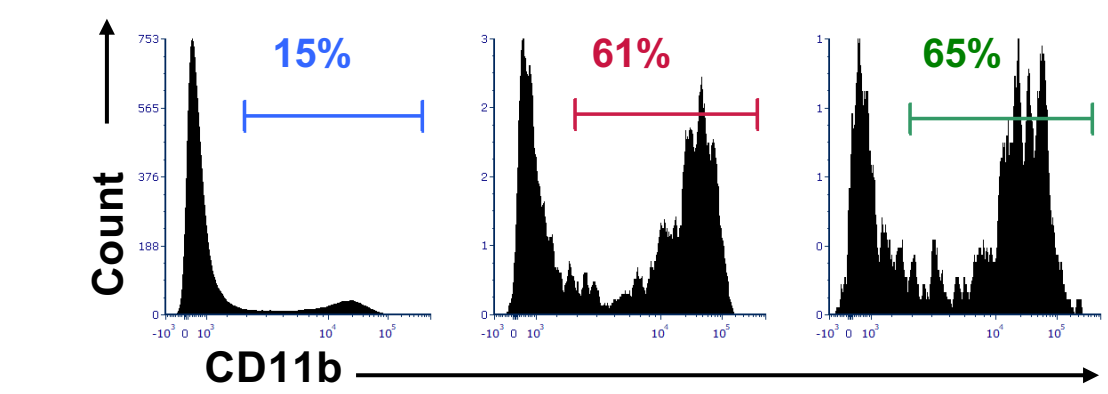
B



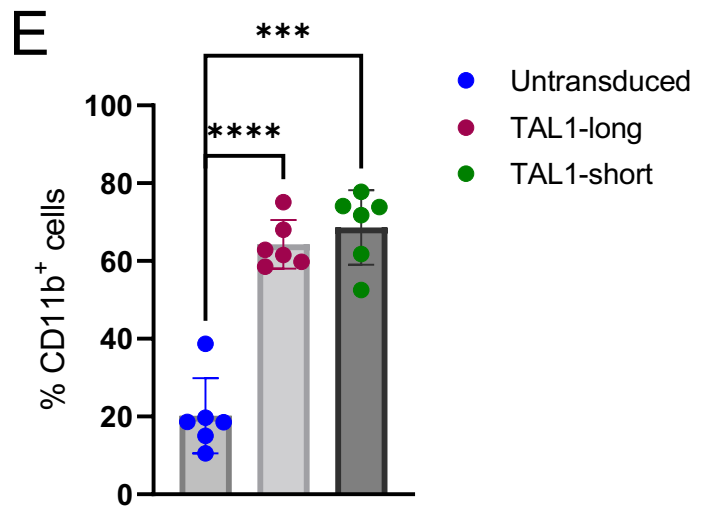
C



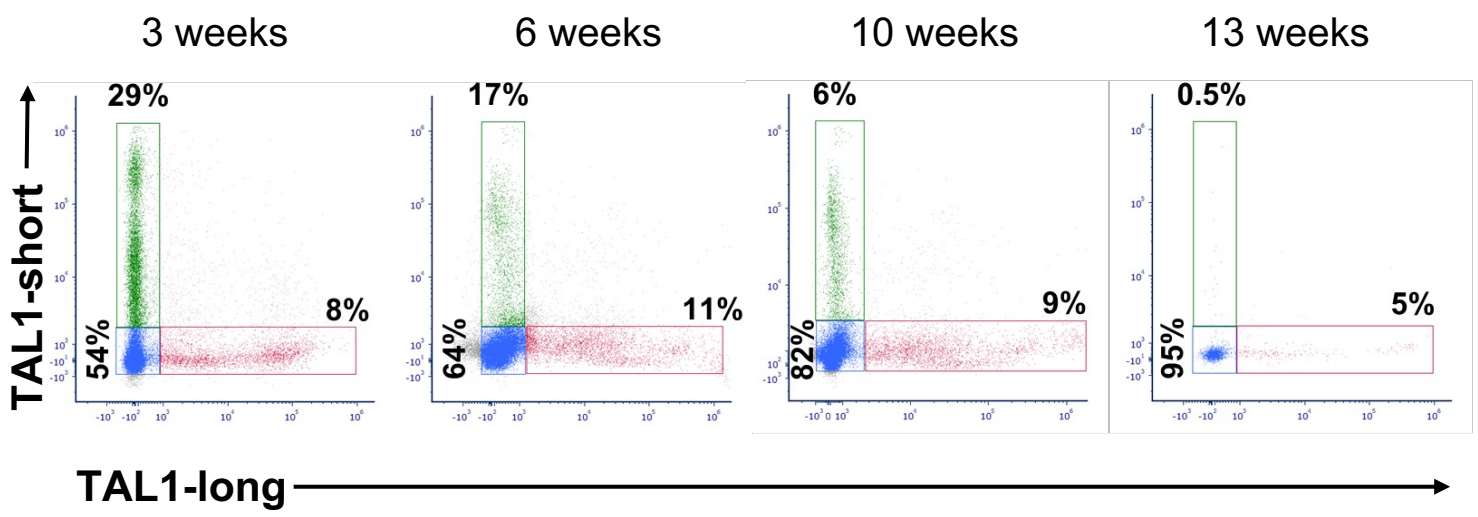
D



E



F



G

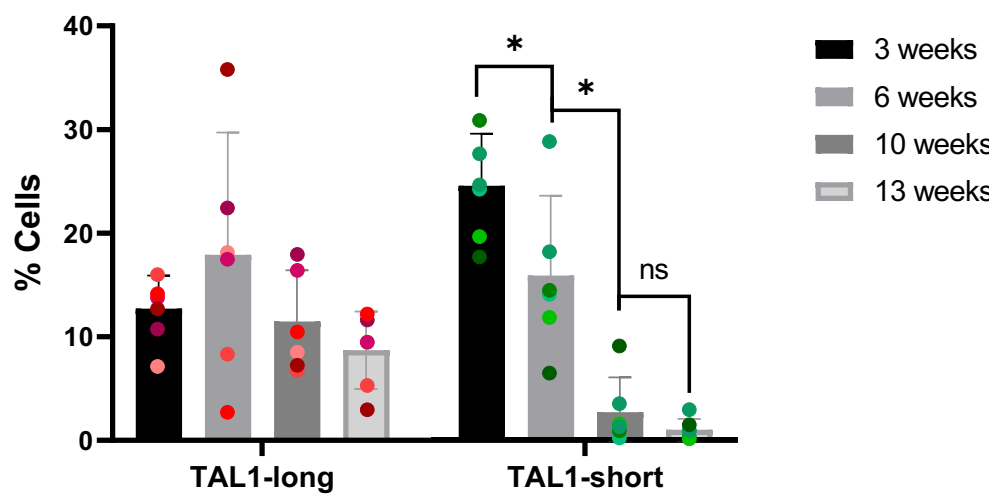
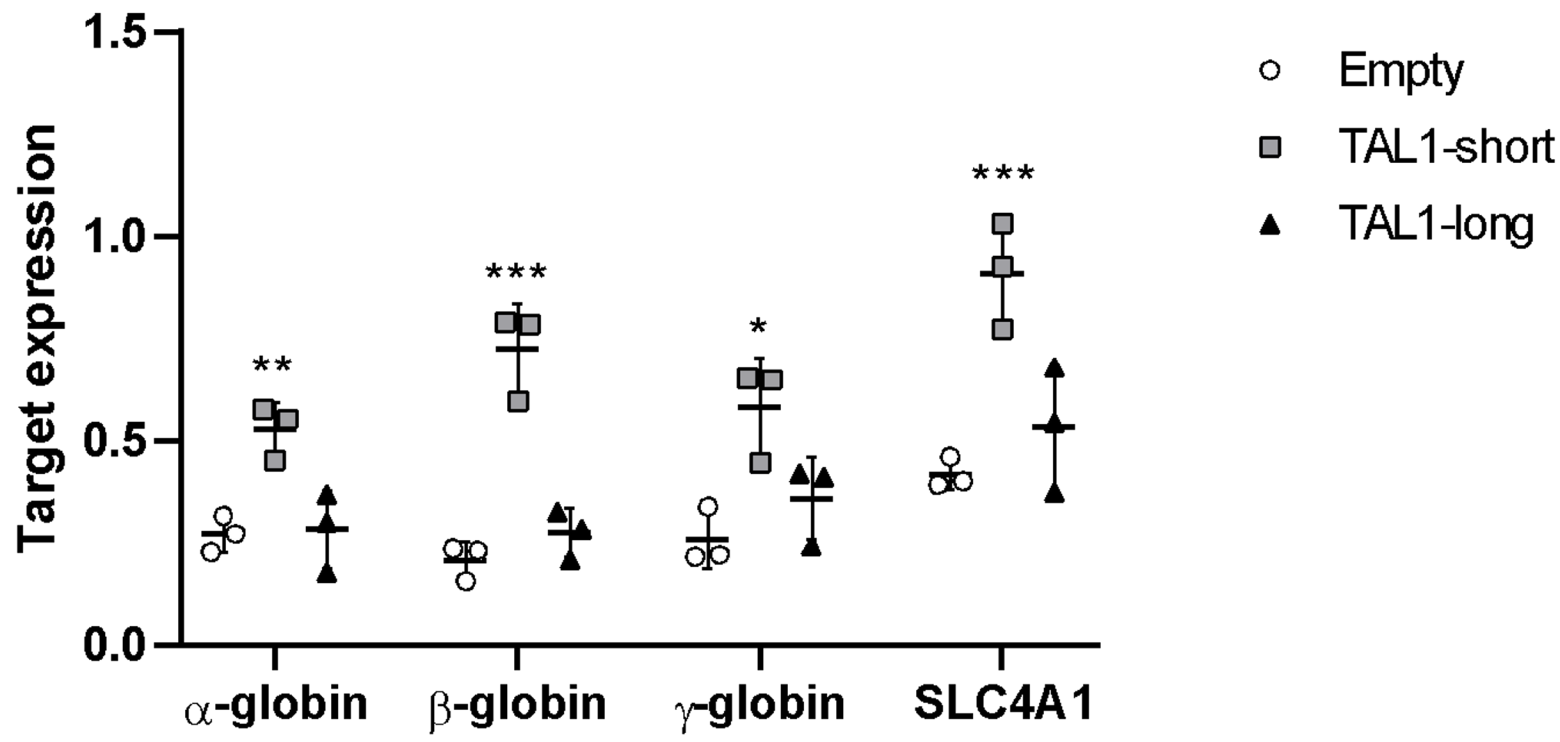
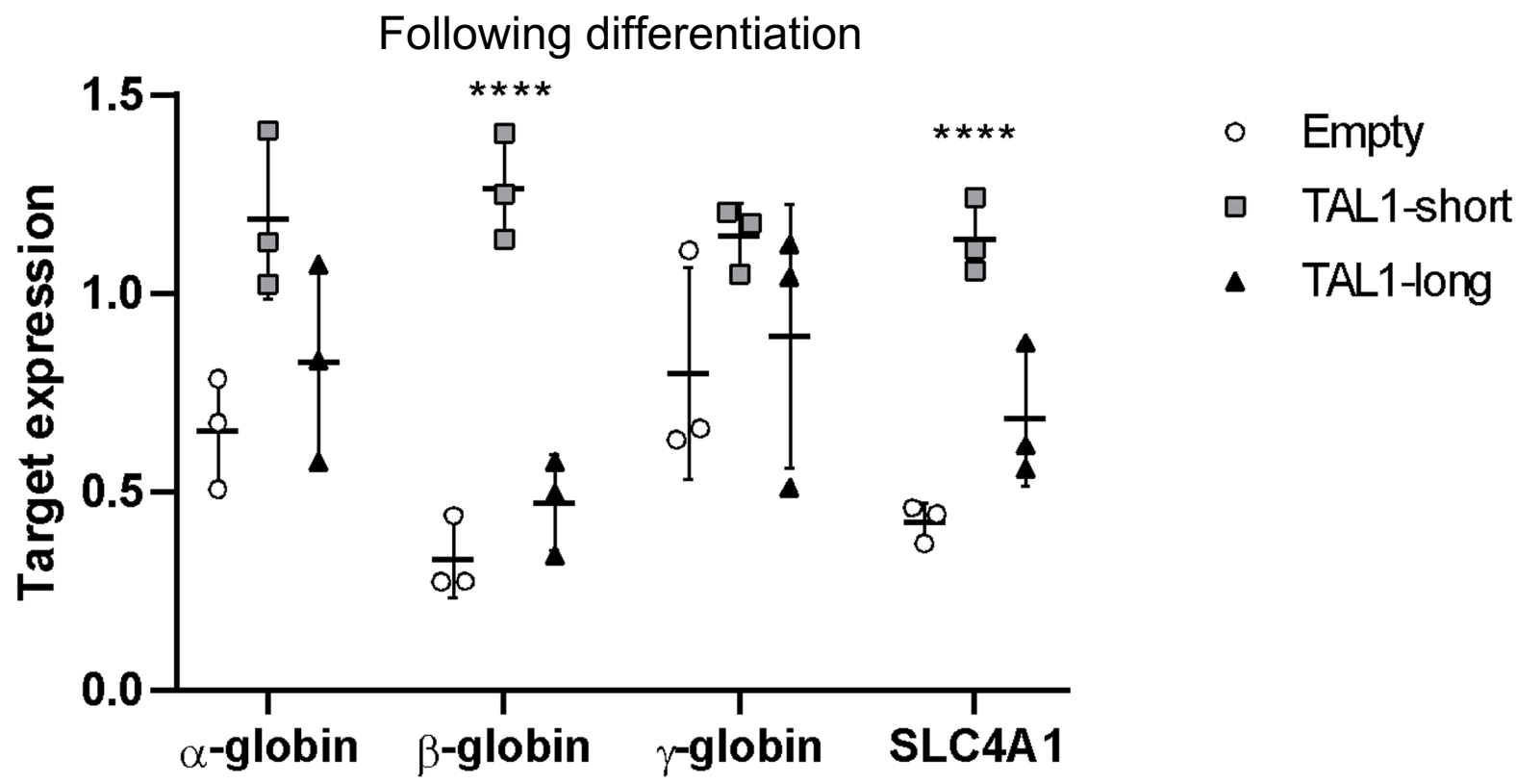


Fig. 6

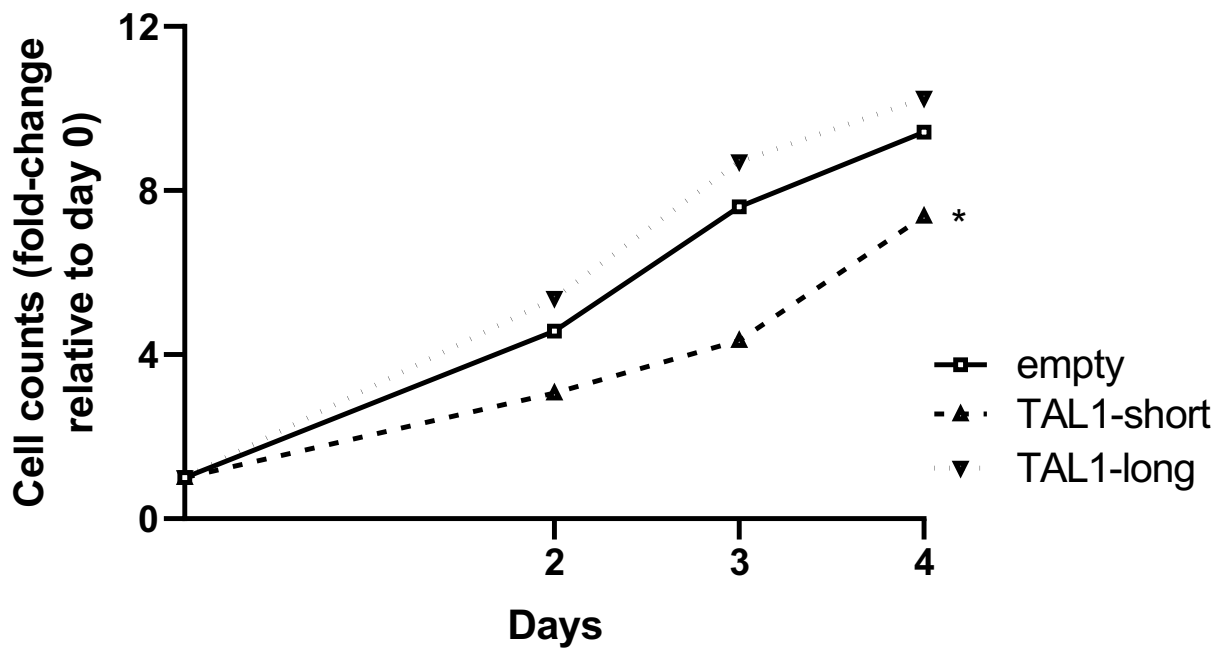
A


Target expression
1.5
1.0
0.5
0.0
**
***
*
***
α-globin
β-globin
γ-globin
SLC4A1
Empty
TAL1-short
TAL1-long

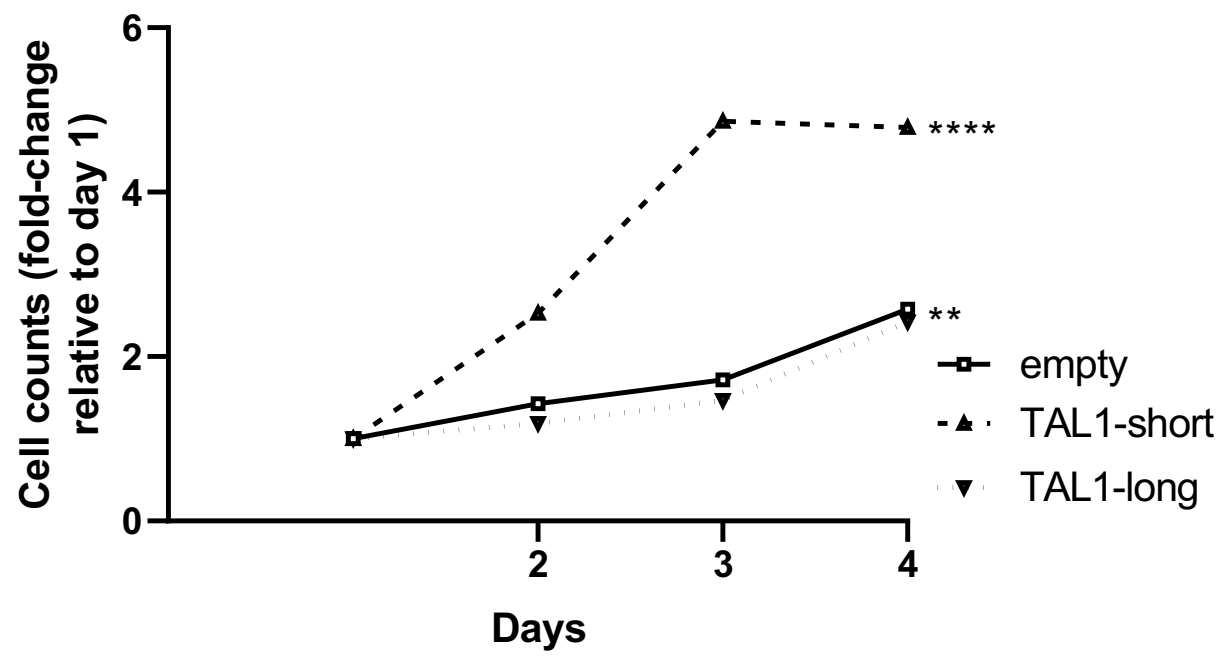
B


Following differentiation
Target expression
1.5
1.0
0.5
0.0
****
****
α-globin
β-globin
γ-globin
SLC4A1
Empty
TAL1-short
TAL1-long

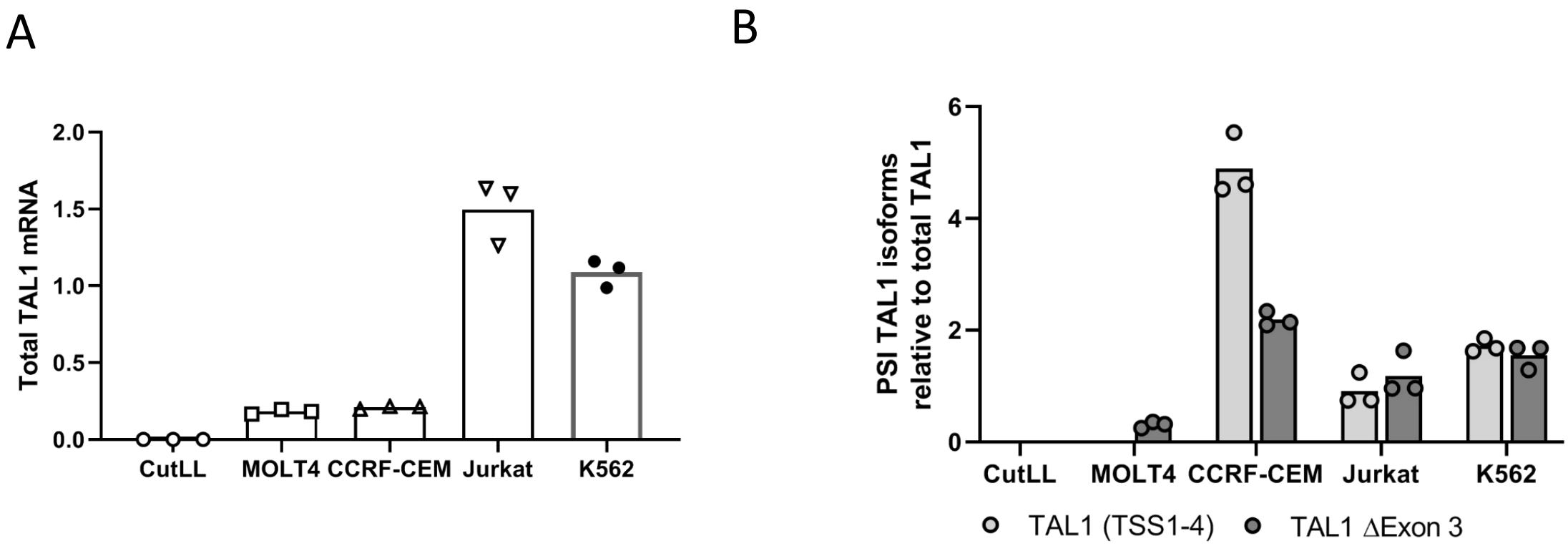
C

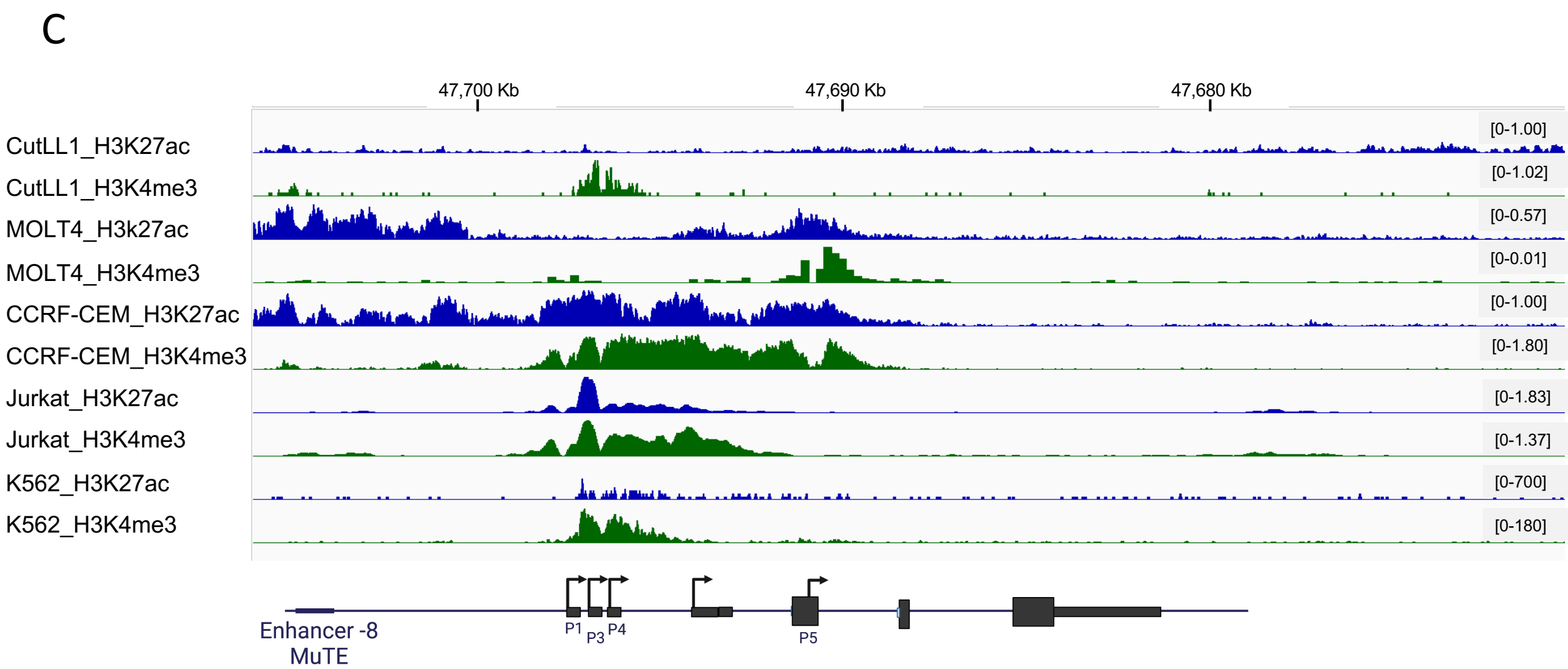

Cell counts (fold-change relative to day 0)
12
8
4
0
2
3
4
Days
*
empty
TAL1-short
TAL1-long

D


Cell counts (fold-change relative to day 1)
6
4
2
0
2
3
4
Days
****
**
empty
TAL1-short
TAL1-long

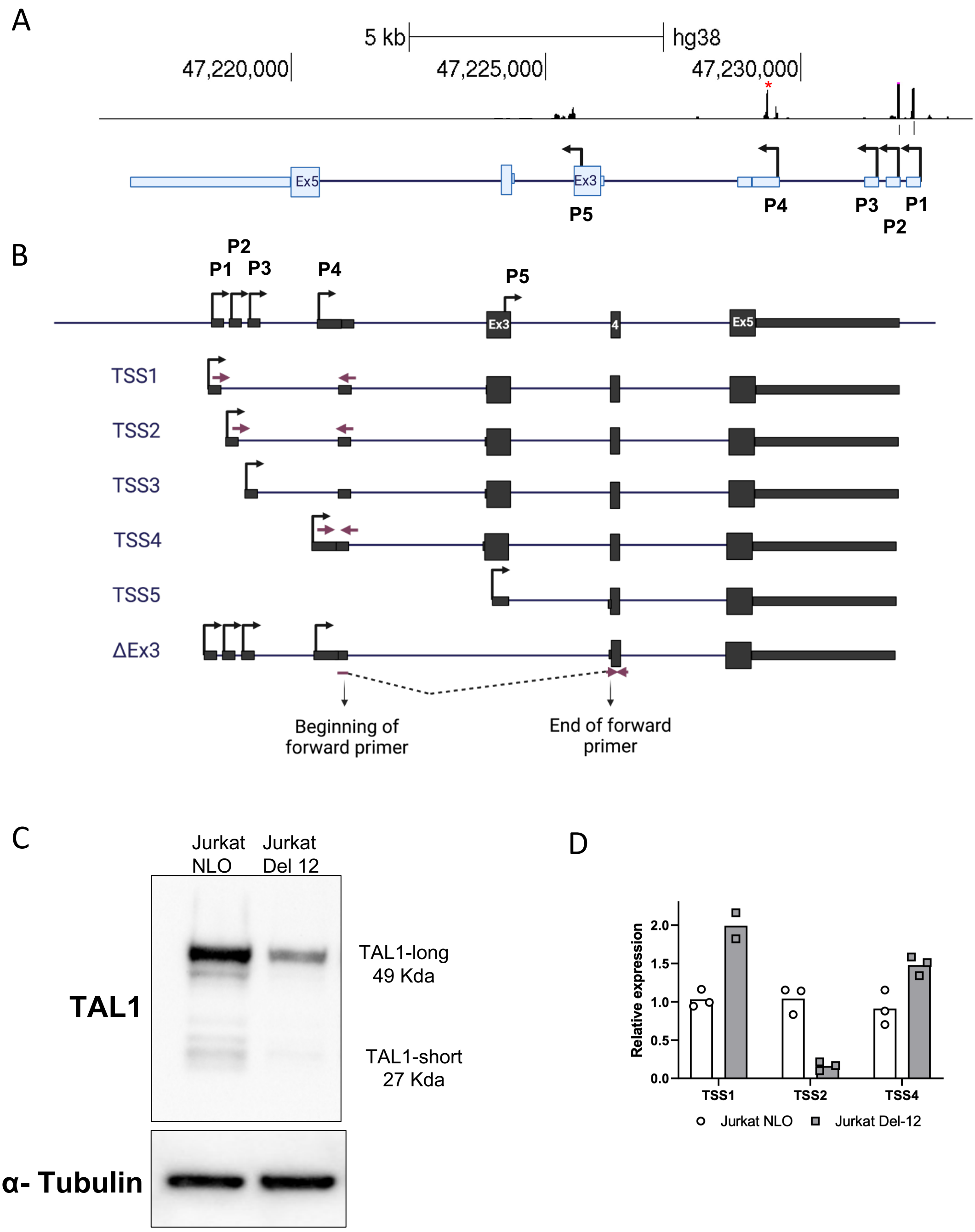
Supplementary Figure 1





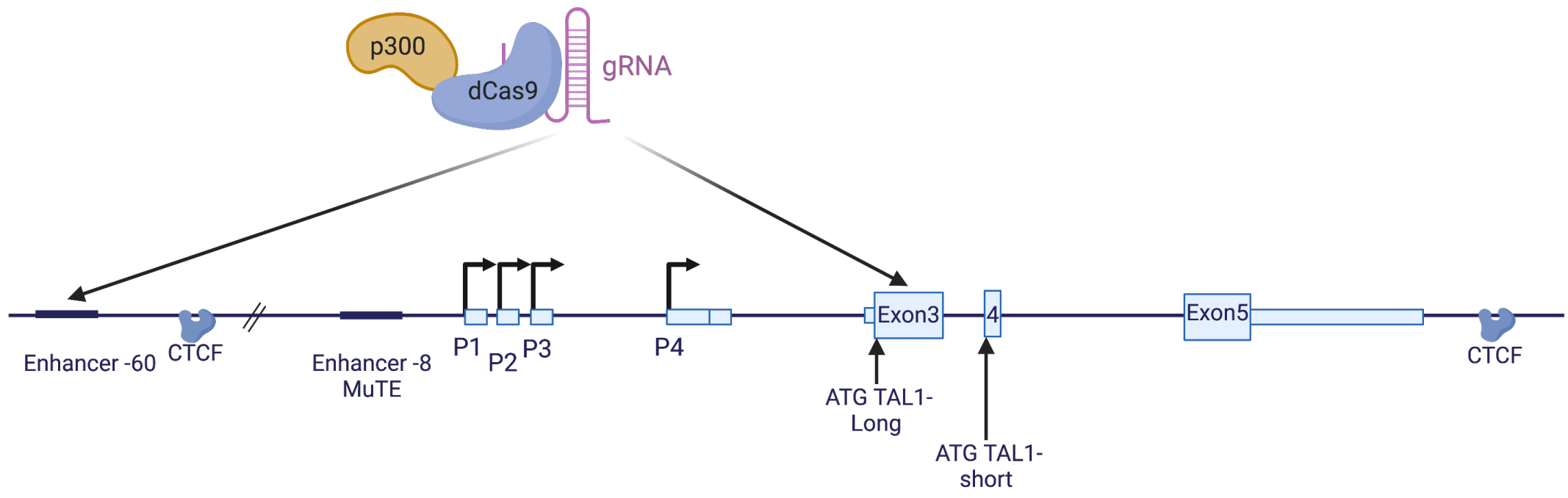
Supplementary Figure 1: A&B. RNA was extracted from the indicated cell lines and analyzed by real-time PCR for total mRNA amount of TAL1 relative to CycloA and hTBP reference genes (S1_Data) (A) and for ΔEx3 relative to endogenous TAL1 total mRNA amount. PSI was calculated by ΔEx3 relative to endogenous TAL1 total mRNA (S1_Data) (B). **C**. ChIP-seq tracks for H3K27ac and H3K4me3 at the TAL1 locus in the indicated cell-lines (genome build hg19).

Supplementary Figure 2

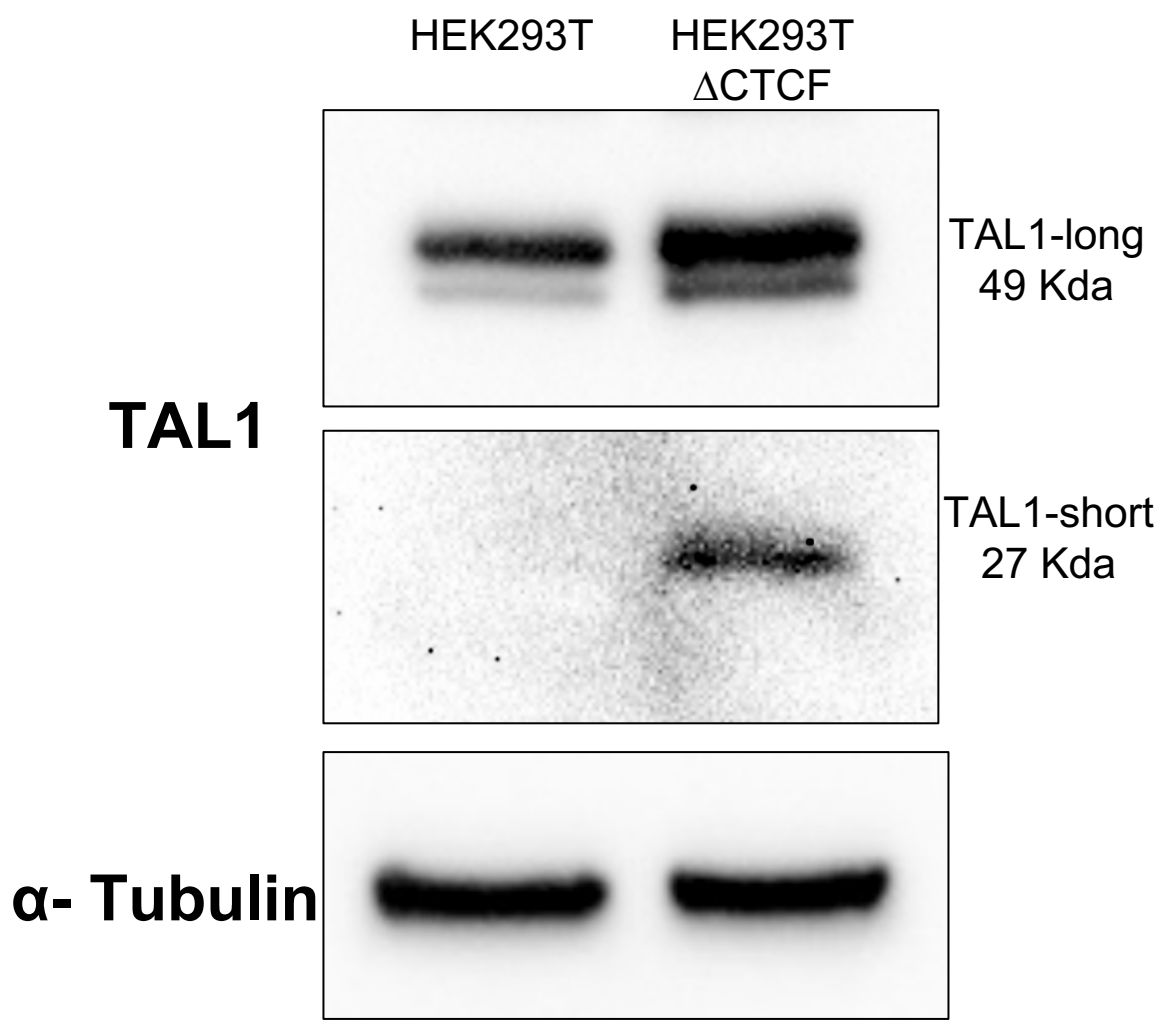


Supplementary Figure 2

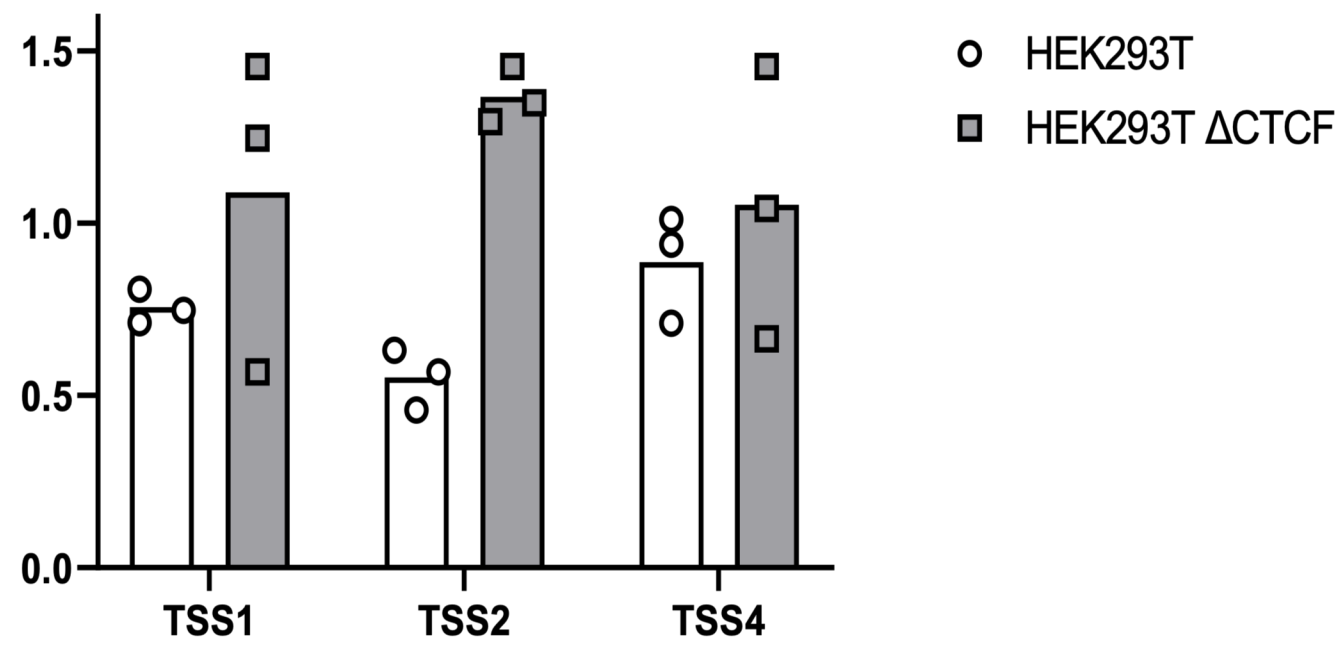
E



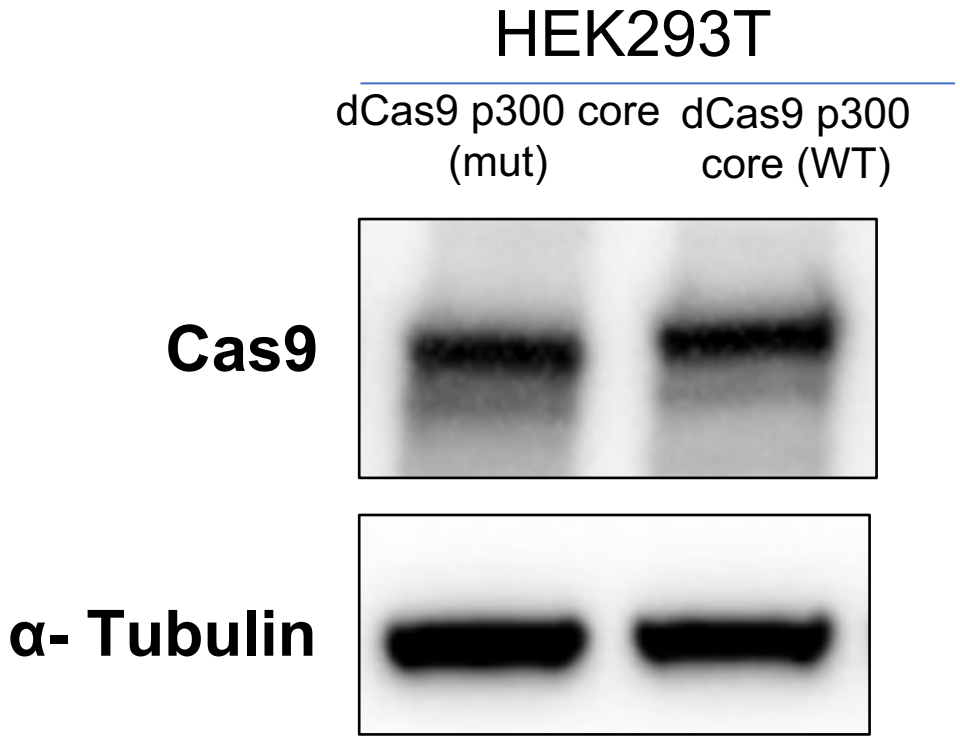
F



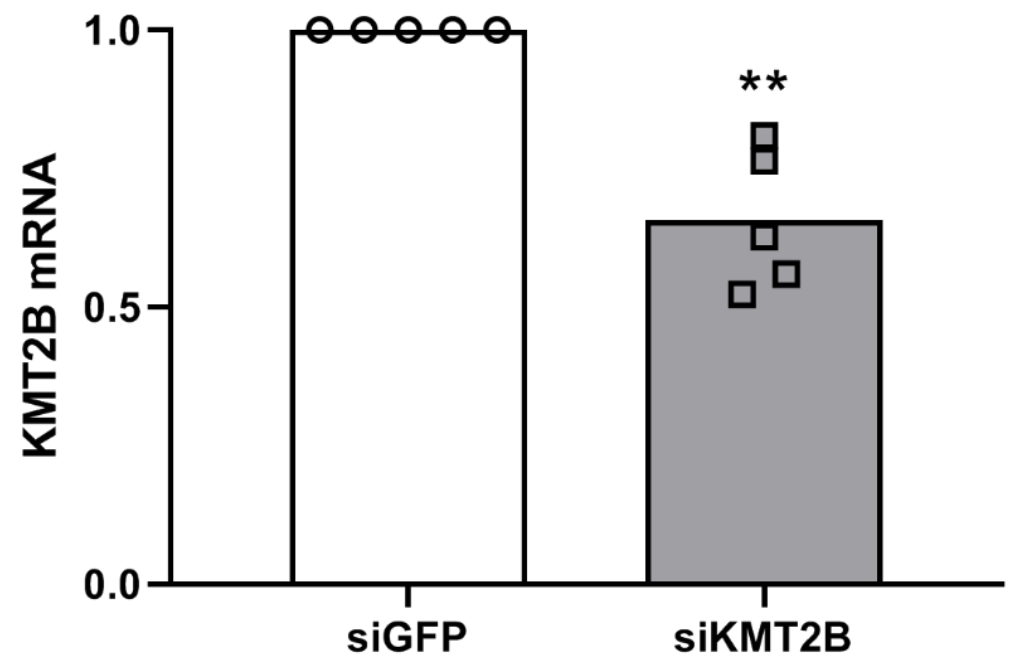
G

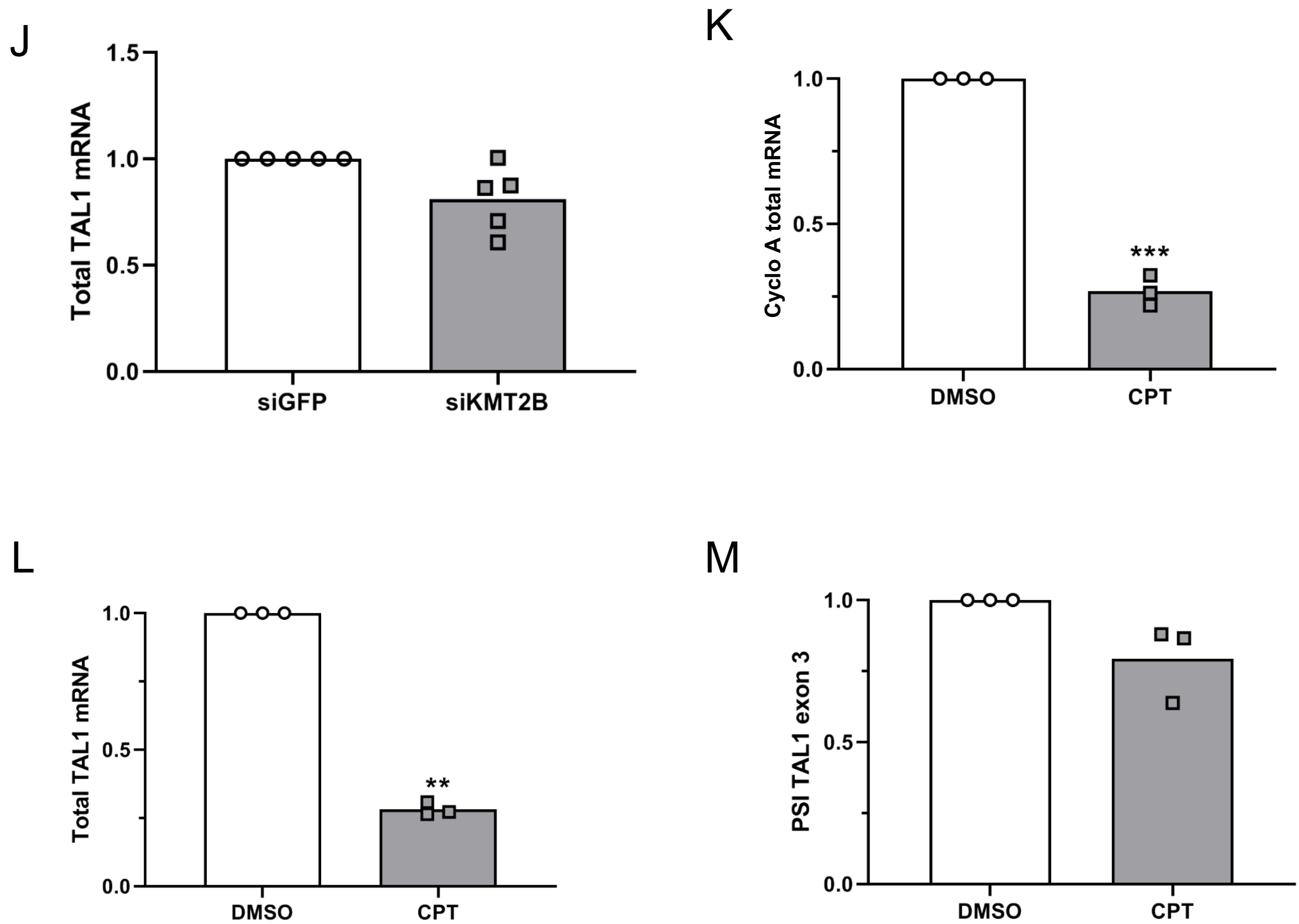


H



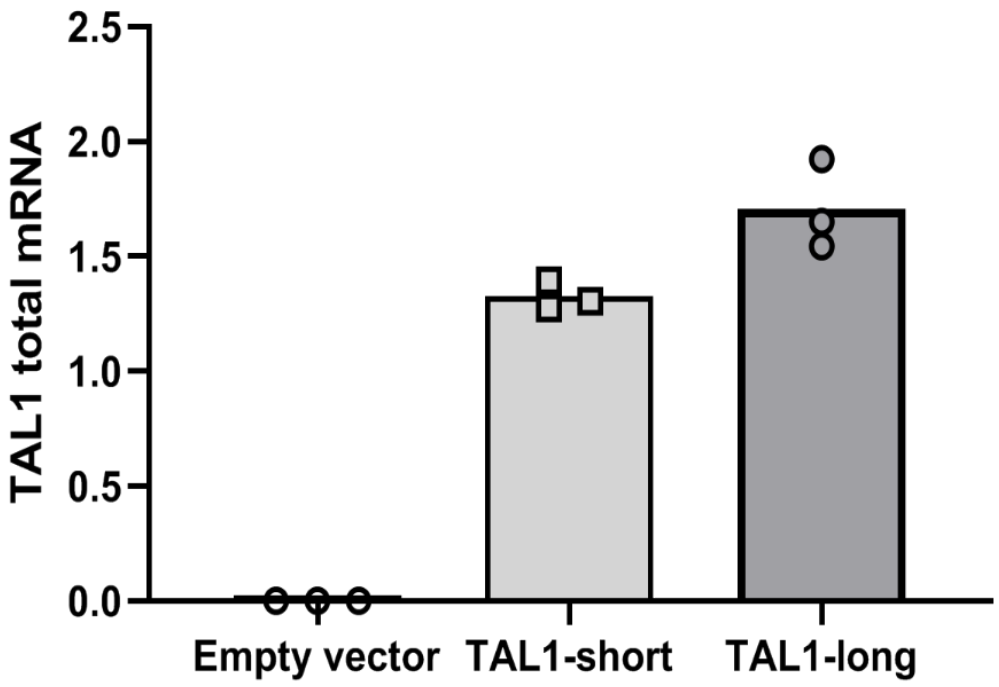
I

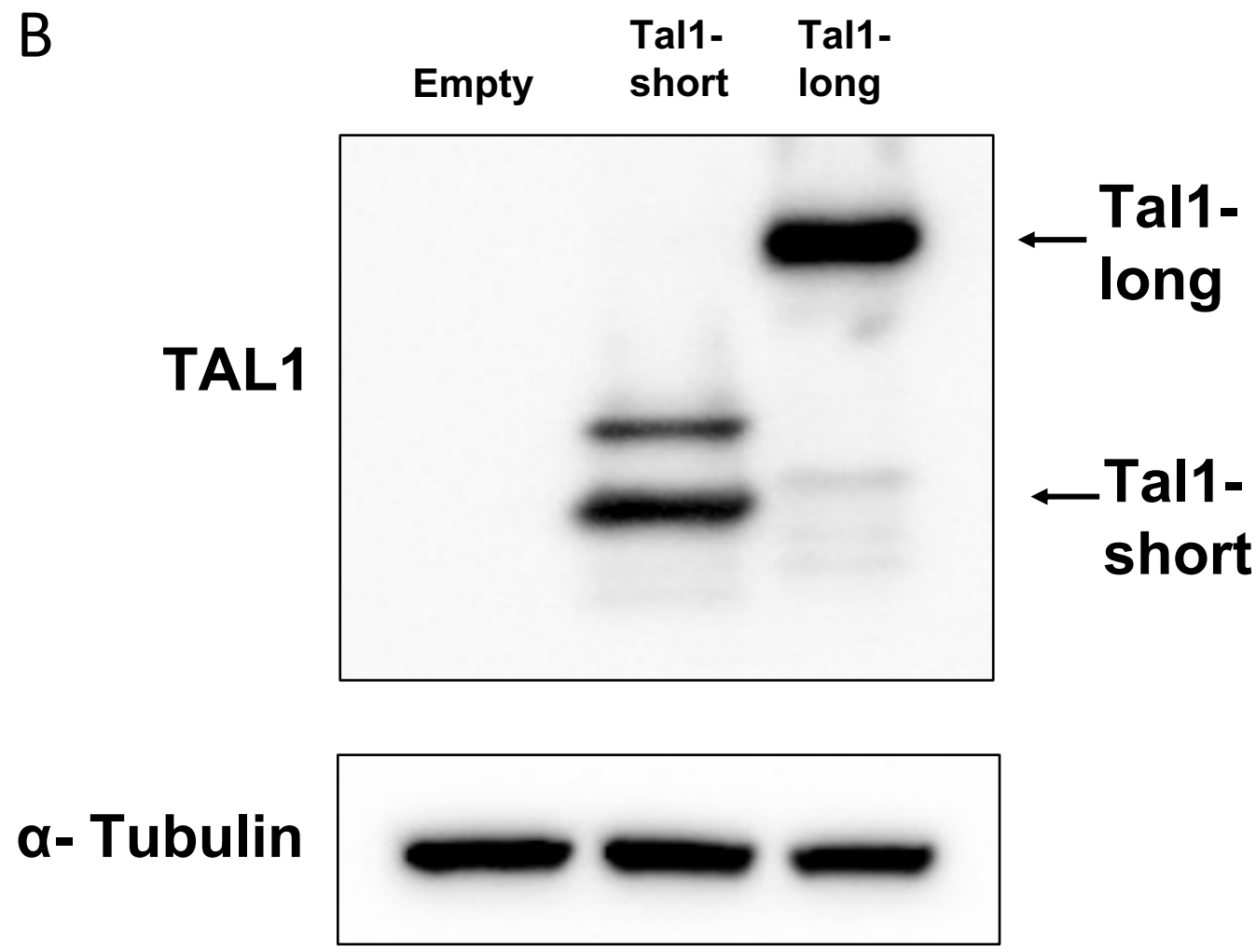


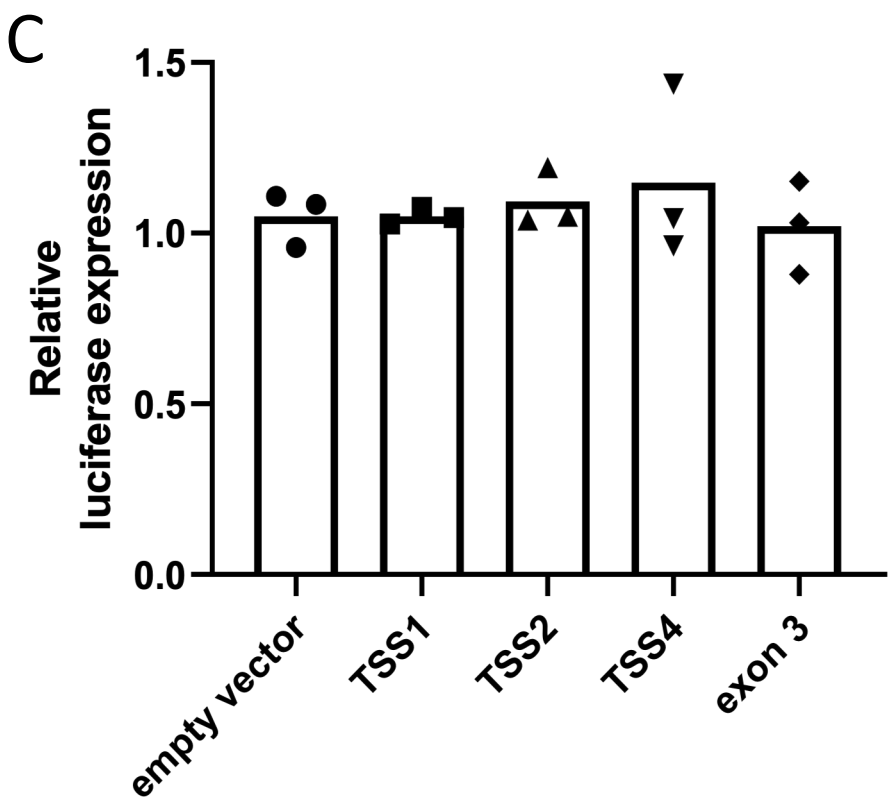


Supplementary Figure 2: A Sequencing of 5' RACE PCR in Jurkat cells aligned to TAL1 locus. Peaks mark the sequence of the 5' UTR. Red asterisk marks a peak that did not align to any known TSS. **B.** Schematic representation of TAL1 mRNA isoforms. Rectangles: exons, black lines: introns; arrow: transcription initiation site. **C&D**. whole cell lysate was extracted from Jurkat NLO cells, expressing TAL1 exogenously, and enhancer mutated cells Jurkat Del-12 , and subjected to western blot analysis using the indicated antibodies (S2_Data) (C). RNA was extracted and real-time PCR was conducted TAL1 promoter relative to to CycloA and hTBP reference genes (S1_Data) (D). **E.** Schematic representation of TAL1 genomic locus. Rectangles: exons, black lines: introns, arrow: transcription initiation site. **F-H**. Whole cell lysate was extracted from HEK293T cells and CTCF mutated cells, HEK293T ΔCTCF, and subjected to western blot analysis using the indicated antibodies (S2_Data) (F). RNA was extracted and analyzed by real-time PCR for each of TAL1's promoters relative to CycloA and hTBP reference genes (S1_Data) (G). **H.** HEK293T cells were transfected with either dCas9-p300 core (mut) or dCas9-p300 core (WT) with four gRNAs targeted to the TAL1 -60 enhancer for 30 h. Whole cell lysate was extracted and subjected to Western blot analysis using the indicated antibodies (S2_Data) . **I&J.** Jurkat cells were transfected with siKMT2B or a negative control siRNA (siGFP), and RNA was extracted 72 hours post-transfection. Real-time PCR was performed to KMT2B total mRNA amount relative to CycloA and hTBP reference genes (S1_Data) (I) and total mRNA amount of TAL1 relative to CycloA and hTBP reference genes (S1_Data) (J). The mean was calculated from three independent biological experiments, each performed with three technical replicates. **K-M.** Jurkat cells were treated with 6 μM CPT for 6 h and analyzed by real-time PCR for total mRNA amount of CycloA relative to 18S reference gene transcribed by Polymerase I (K) and TAL1 relative to 18S reference gene (L) and for ΔEx3 relative to TAL1 total mRNA amount. PSI was calculated by ΔEx3 relative to TAL1 total mRNA (S1_Data) (M).

Supplementary Figure 3

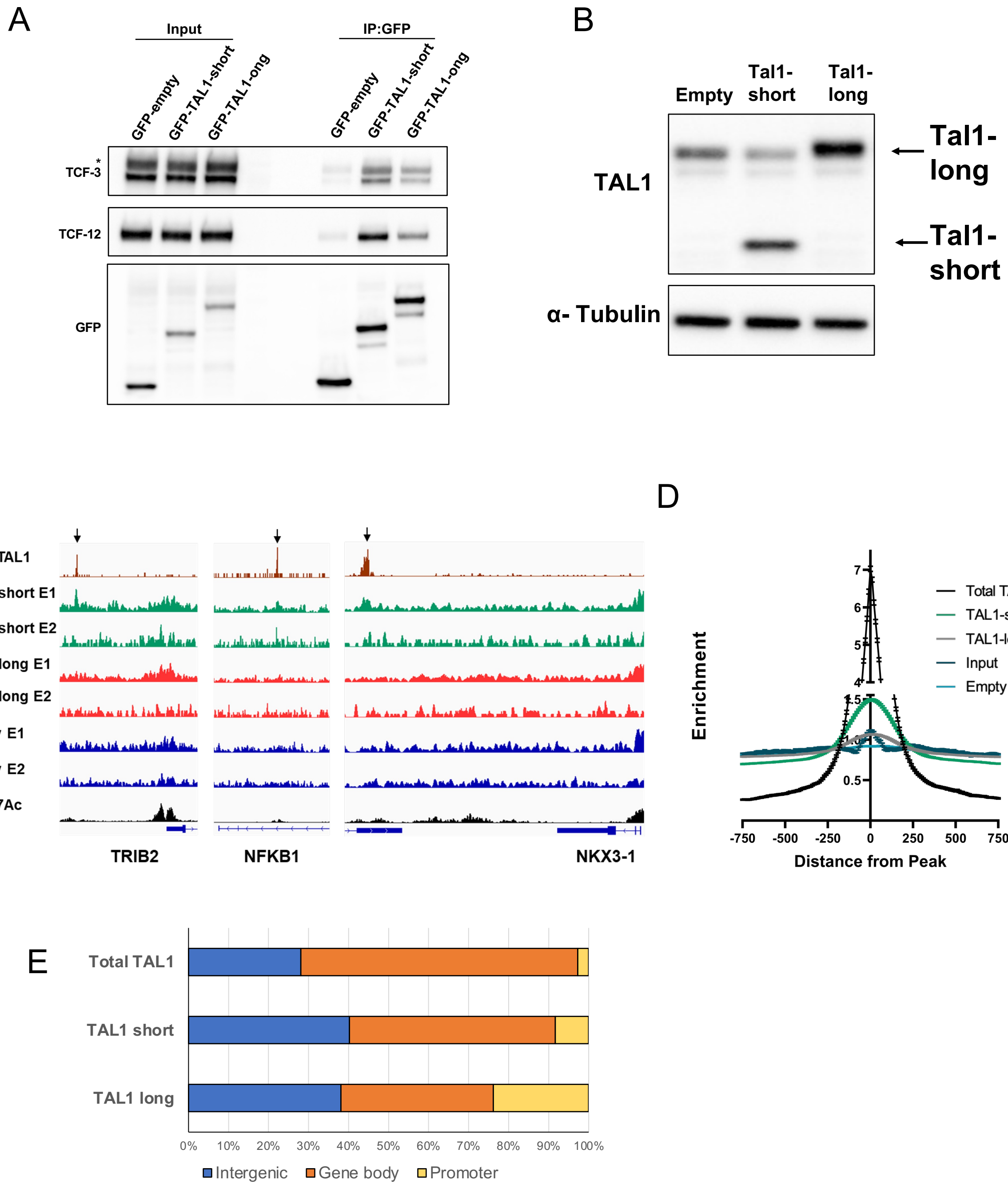






Supplementary Figure 3: A-C. HEK293T cells were co-transfected with TAL1 promoters: promoters 1-3, 4, 5 and exon 4 as a negative control. The second plasmid was an empty vector, TAL1-short or TAL1-long. After 30 h RNA was extracted and real-time PCR was performed to TAL1 total mRNA amount relative to CycloA and hTBP reference genes (S1_Data) (A) and whole cell lysate was extracted and subjected to Western blot analysis using the indicated antibodies (S2_Data) (B) and RNA was extracted and analyzed by real-time PCR for luciferase mRNA relative to renilla mRNA (S1_Data) (C).

Supplementary Figure 4



F. ETS

	Motif	P value	% Target Sequences with Motif	% Background Sequences with Motif
Total TAL1		1e-21	37.6	12.9
TAL1-short		1e-72	47.0	15.7
TAL1-long		1e-3	29.4	4.7

G. GATA

	Motif	P value	% Target Sequences with Motif	% Background Sequences with Motif
Total TAL1		1e-70	58.5	10.8
TAL1-short		1e-53	24.2	5.3
TAL1-long	---			

H. RUNX

	Motif	P value	% Target Sequences with Motif	% Background Sequences with Motif
Total TAL1		1e-31	36.4	8.7
TAL1-short		1e-93	42.0	9.9
TAL1-long	---			

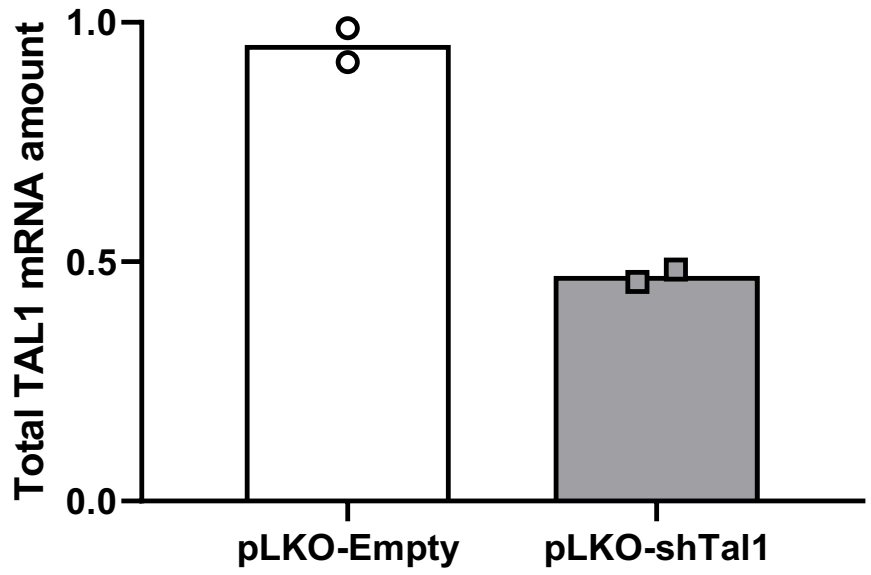
I. MAF

	Motif	P value	% Target Sequences with Motif	% Background Sequences with Motif
Total TAL1	---			
TAL1-short		1e-2	23.5	4.5
TAL1-long		1e-2	5.1	3.1

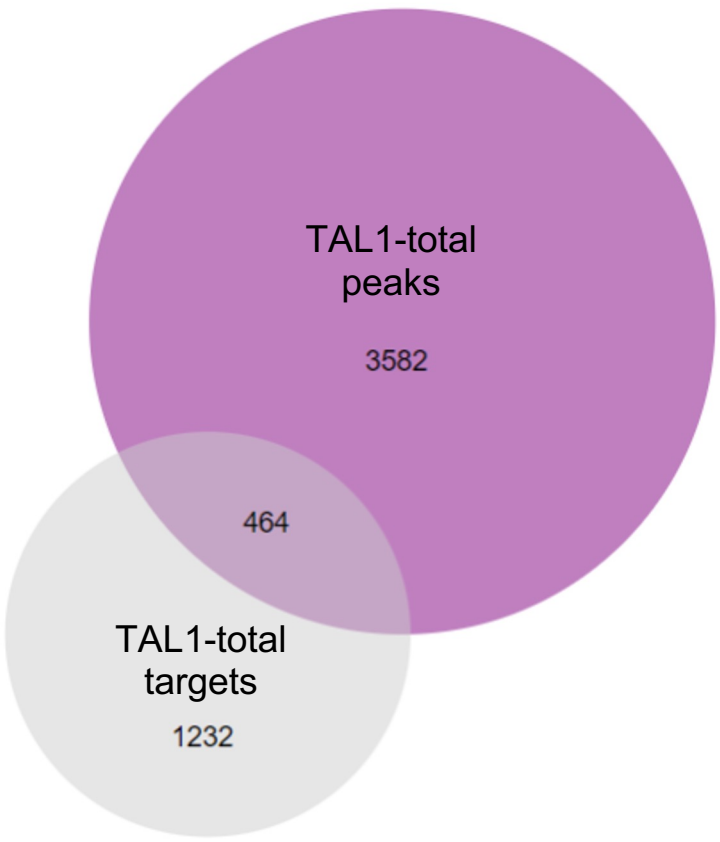
J. E-box

	Motif	P value	% Target Sequences with Motif	% Background Sequences with Motif
Total TAL1		1e-27	12.5	0.6
TAL1-short		1e-36	9.4	0.9
TAL1-long	---			

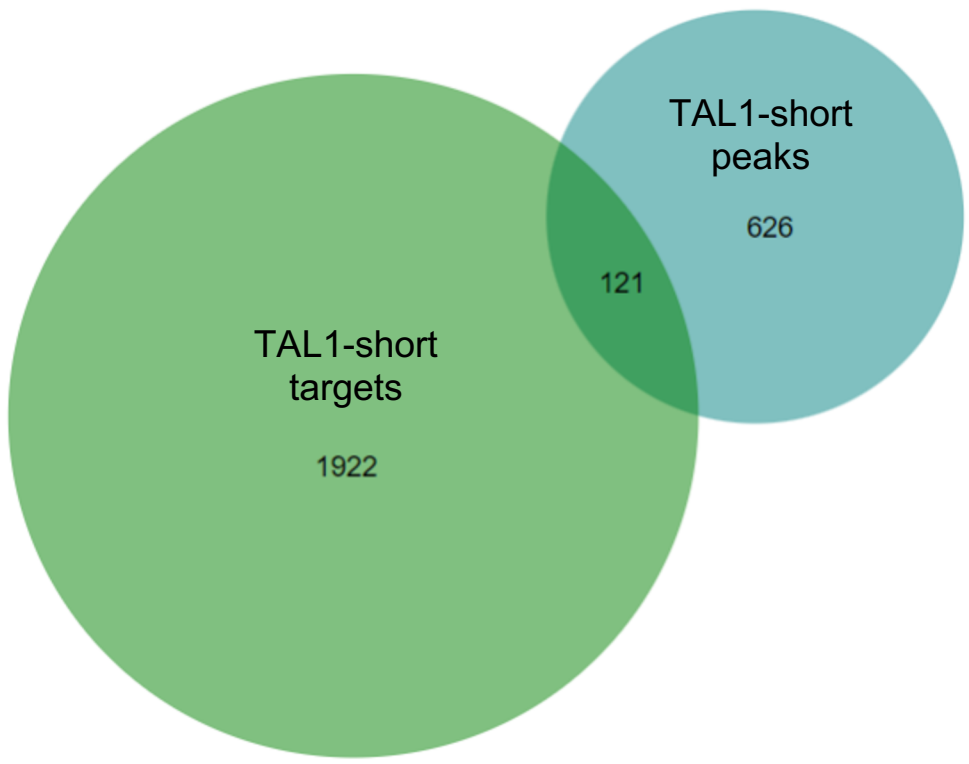
K



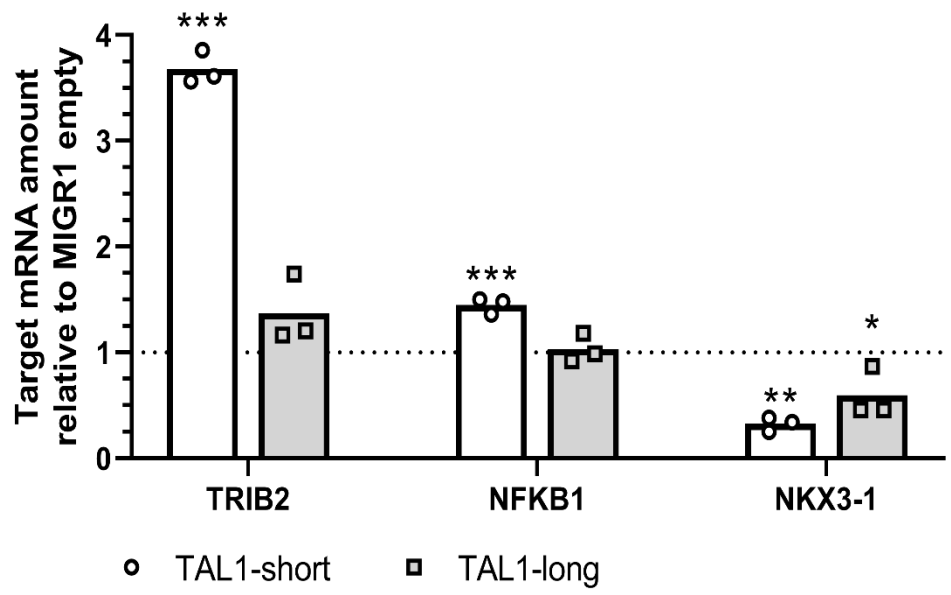
L



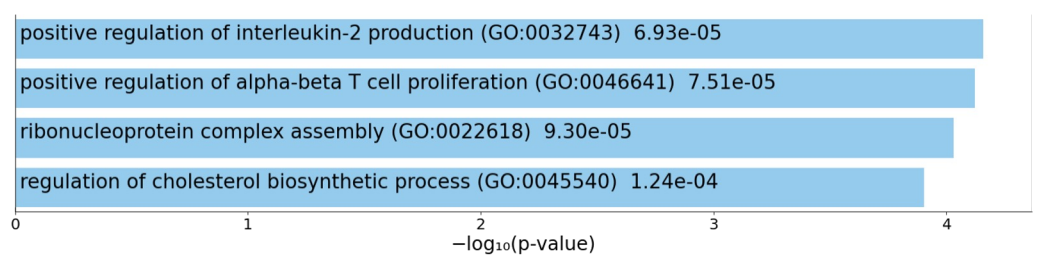
M



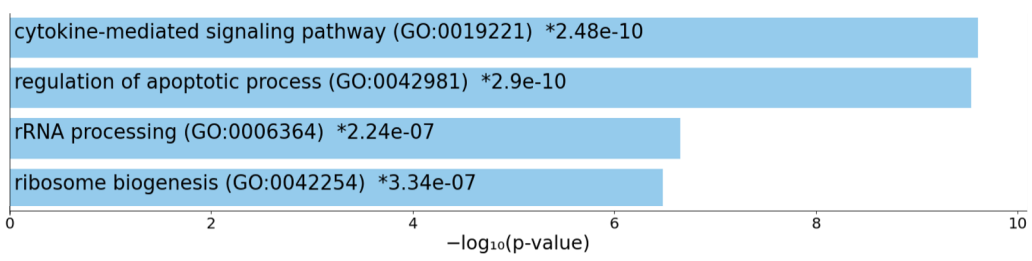
N



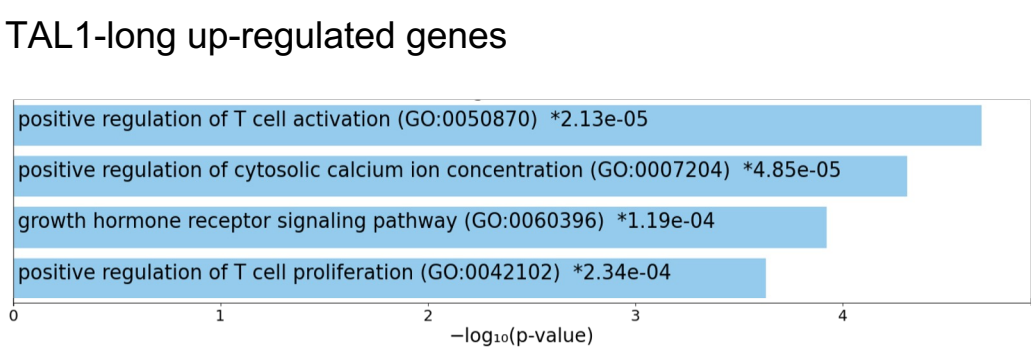
O



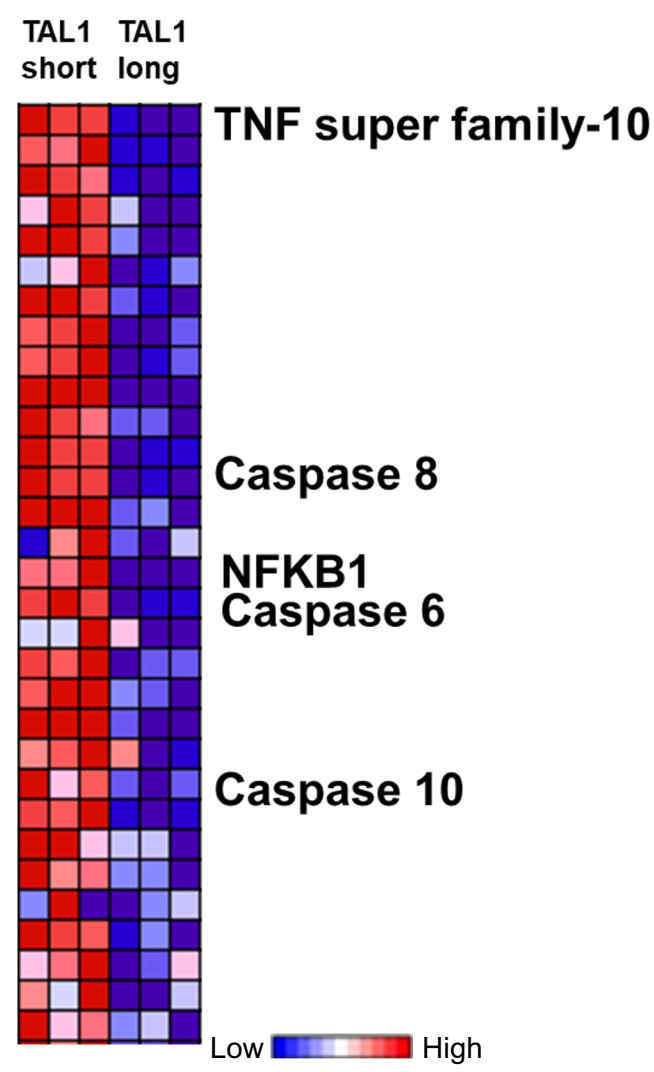
P



Q



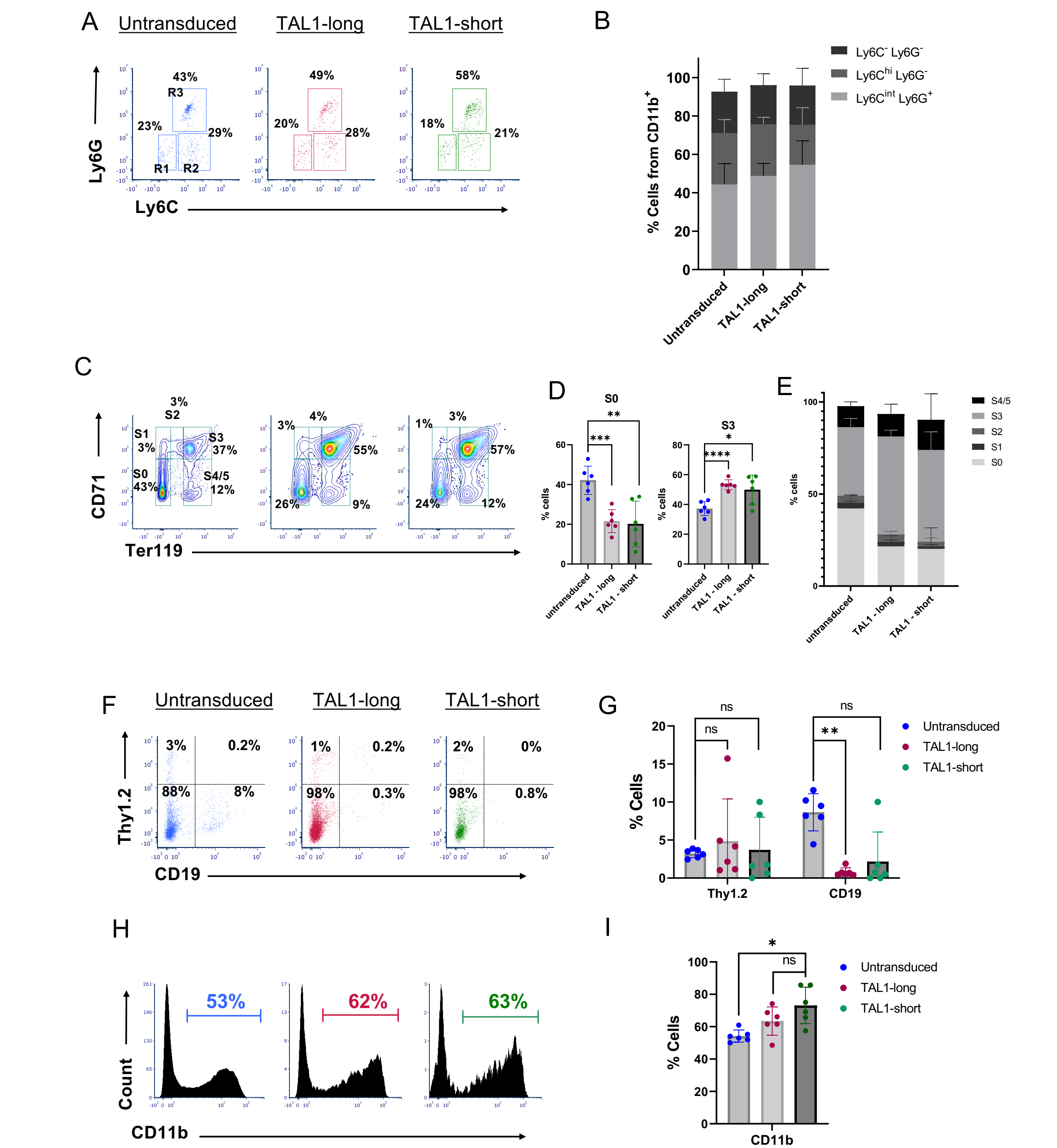
R

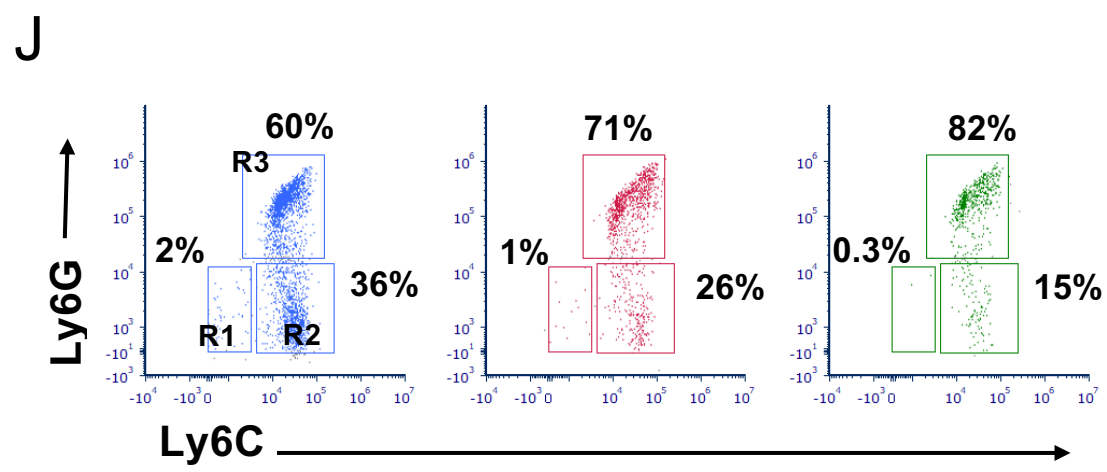


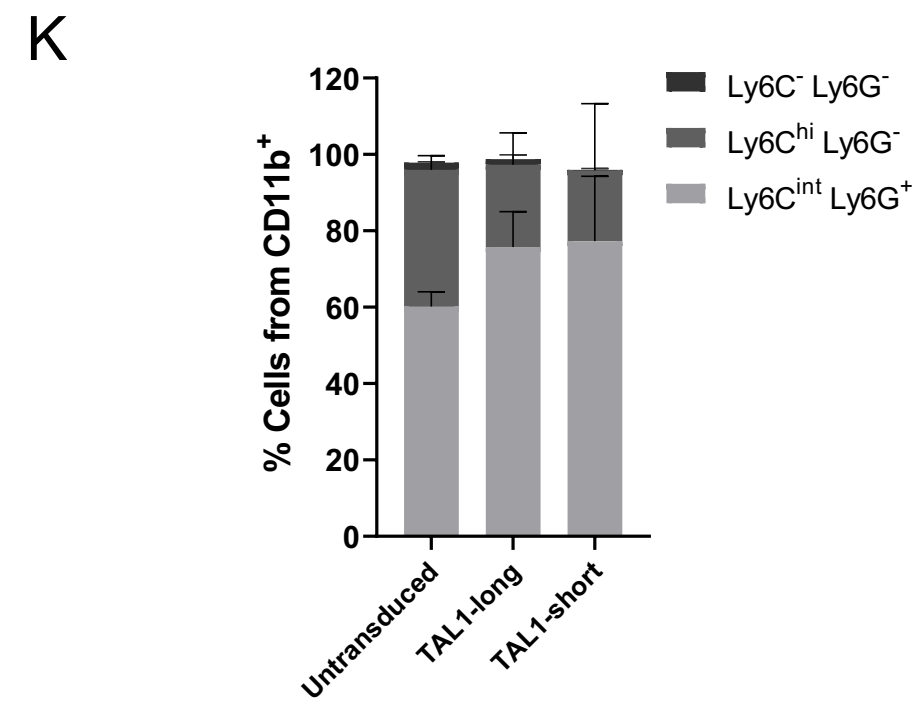
Stastistics for GSEA plot Apoptosis (TAL1-short vs TAL1-long)	
Enrichment Score (ES)	0.51318172
Normalized Enrichment Score (NES)	1.997536
Nominal p-value	0
FDR q-value	0.00697

Supplementary Figure 4: A. Jurkat cells were infected with either empty vector, GFP-TAL1-short or GFP-TAL1-long. GFP was immunoprecipitated and interacting proteins were detected with indicated antibodies (S2_Data). **B-J.** Jurkat cells were infected with either empty vector, FLAG-TAL1-short or FLAG-TAL1-long. ChIP-seq was performed with anti-FLAG magnetic beads. In addition we analyzed available data for TAL1-total (see methods). Whole cell lysate was extracted and subjected to Western blot analysis using the indicated antibodies (S2_Data) (B). ChIP-seq tracks are shown at the indicated locus for each biological experiment. Arrows indicate known TAL1-total binding site (C). ChIP-seq average signal for TAL1-total, TAL1-short and TAL1-long, empty vector and input as a function of distance from TAL1-total peaks (S1_Data) (D). ChIP-seq peaks across genomic regions (S1_Data) (E). Most abundant DNA sequence motifs identified in ChIP-seq peaks (F-J). **K-R.** Jurkat cells were infected with shRNA to TAL1 3' UTR and later with empty vector, FLAG-TAL1-short or FLAG-TAL1-long. Induction with tetracycline was conducted for 72hr and silencing was measured using real-time PCR for total mRNA amount of endogenous TAL1 relative to CycloA and hTBP reference genes (S1_Data) (K). RNA-seq was preformed on three biological replicates and analyzed for isoforms specific targets. In addition we analyzed available data for TAL1-total (see methods). Venn diagrams showing numbers of distinct and common RNA-seq targets and peak-associated genes of TAL1-total (L) and TAL1-short (M). Real-time PCR for total mRNA amount of three TAL1-short targets was performed relative to CycloA and hTBP reference genes (S1_Data) (N). Bar chart of top four enriched terms from the GO_Biological_Process_2018 gene set library for TAL1-total (O) and TAL1-short (P) and TAL1-long (Q). Heatmap and statistics for expression level of apoptosis pathway genes from GSEA plot (R).

Supplementary Fig. 5

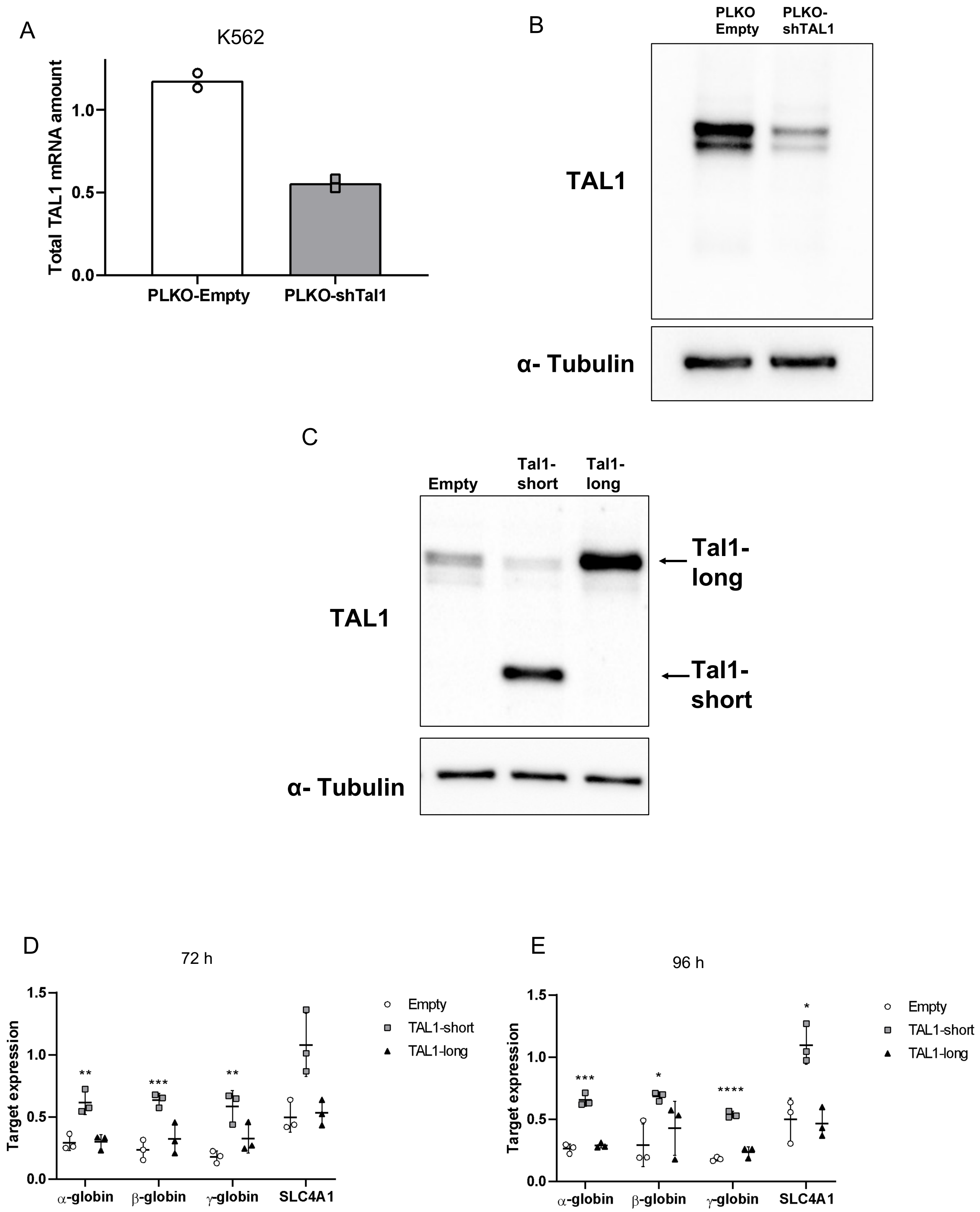


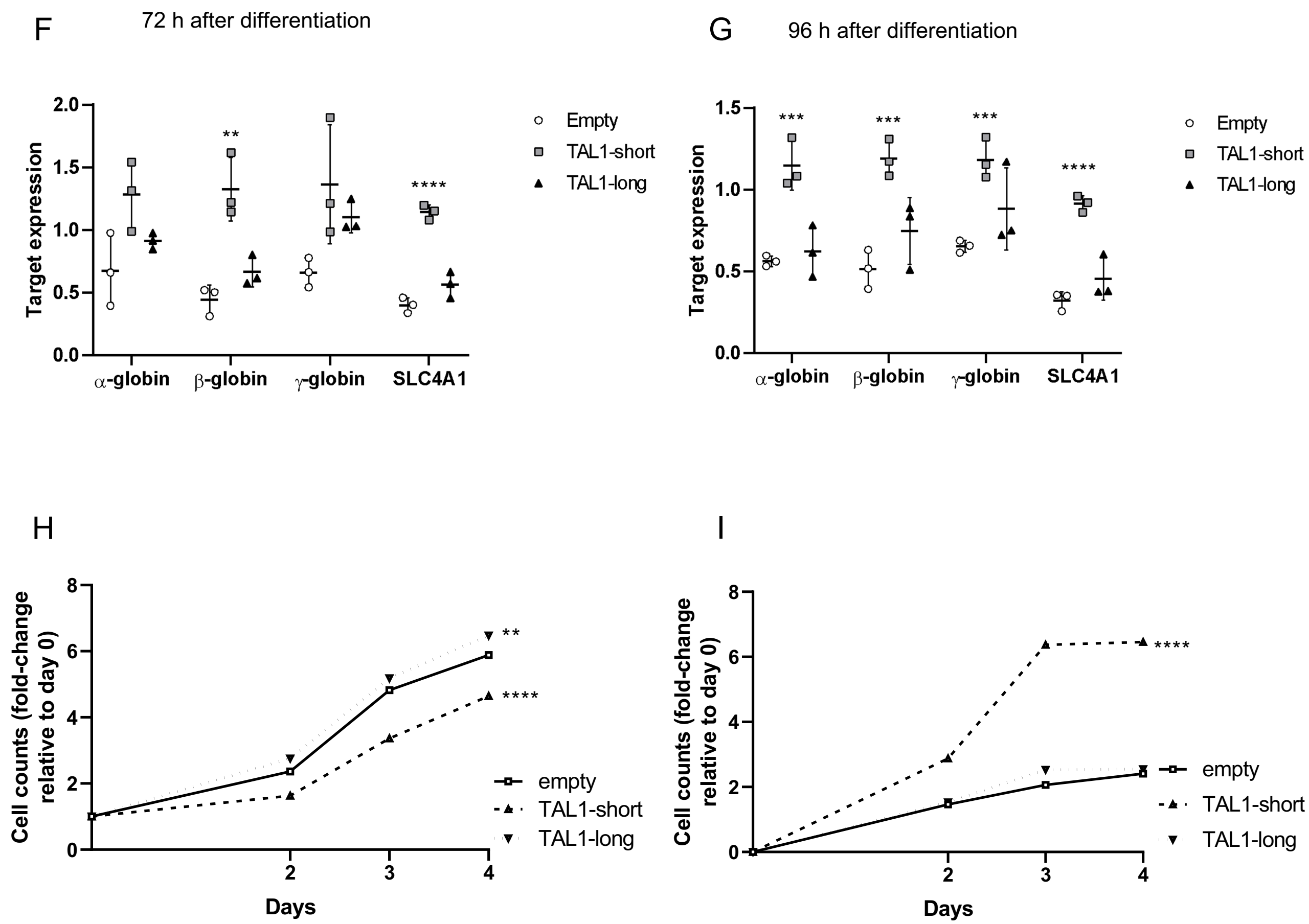




Supplementary Figure 5: A-K. Equal numbers of bone marrow cells from 5FU-treated CD45.1 wild-type mice were transduced with retroviruses expressing either TAL1-short-GFP or with TAL1-long-dtTomato. A mixture of 1:1 ratio of the transduced bone marrow cells was then transplanted into lethally-irradiated CD45.2 wild-type recipient mice. Mice were sacrificed fourteen weeks after BMT and the bone marrow cells were harvested and analyzed using flow cytometric analysis. **(A)** Representative flow cytometry dot plots of Ly6C vs Ly6G staining gated on CD11b positive cells that are GFP-/dtTomato- (untransduced, left panel), dtTomato+ (Tal1-long, middle panel) or GFP+ (Tal1-short, right panel) splenocytes. Regions represent the following populations: R1- CD11b$^+$, Ly6C$^-$, Ly6G$^-$(circulating monocytes), R2- CD11b$^+$, Ly6C^{hi}, Ly6G$^-$ (inflammatory monocytes), R3- CD11b$^+$, Ly6C^{int}, Ly6G$^+$ (granulocytes). **(B)** Bar graph summarizing results in c (S1_Data). **(C)** Representative flow cytometry dot plots analysis of CD71 and Ter119 staining gated on CD45-negative cells that are GFP-/dtTomato- (untransduced, left panel), dtTomato+ (Tal1-long, middle panel) or GFP+ (Tal1-short, right panel). Regions S0-S5 represent the stage in erythrocyte differentiation. **(D)** Bar graph summarizing results in e (S1_Data). Left panel- percentages of CD71$^-$ Ter119$^-$ cells (the S0 phase in erythrocytes differentiation) and Right panel- CD71$^+$ Ter119^{hi} (the S3 phase in erythrocytes differentiation. **(E)** Frequencies of the sub-population of Erythrocytes in the different differentiation stages based on the CD71 and Ter119 expression (S1_Data). **(F)** Representative flow cytometry dot plots of Thy1.2 vs CD19 staining gated on GFP-/dtTomato- (untransduced, left panel), dtTomato+ (Tal1-long, middle panel), or GFP+ (Tal1-short, right panel) bone marrow cells. **(G)** Bar graph summarizing results in a (S1_Data). **(H)** Representative flow cytometry histograms of CD11b staining gated on GFP-/dtTomato- (untransduced, left panel), dtTomato+ (Tal1-long, middle panel) or GFP+ (Tal1-short, right panel) bone marrow cells. **(I)** Bar graph summarizing results in c (S1_Data). **(j)** Representative flow cytometry dot plots of Ly6C vs Ly6G staining gated on CD11b positive cells that are GFP-/dtTomato- (untransduced, left panel), dtTomato+ (Tal1-long, middle panel) or GFP+ (Tal1-short, right panel) splenocytes. Regions represent the following populations: R1- CD11b$^+$, Ly6C$^-$, Ly6G$^-$(circulating monocytes), R2- CD11b$^+$, Ly6C^{hi}, Ly6G$^-$ (inflammatory monocytes), R3- CD11b$^+$, Ly6C^{int}, Ly6G$^+$ (granulocytes). **(K)** Bar graph summarizing results in e (S1_Data). Two-tailed paired t-test, *<0.05, **<0.01, ***<0.001, ****<0.0001, sd. *(n=6).*

Supplementary Figure 6





Supplementary Figure 6: A-I. K562 cells were infected with inducible shRNA against the 3' UTR of TAL1. In addition, the cells were infected with MIGR1 plasmid with GFP at the C-terminal followed by an IRES element. Infection was with either empty vector, TAL1-short or TAL1-long. Following induction with tetracycline for 72 h, cells were treated with 20 uM hemin to promote erythroid differentiation for the indicated time points. RNA was extracted silencing was measured by real-time PCR for total endogenous TAL1 relative to CycloA and hTBP reference genes (S1_Data) (A) and whole cell lysate was extracted and subjected to Western blot analysis using the indicated antibodies (S2_Data) (B&C). RNA was analyzed by real-time PCR for total mRNA amount of α β, γ globin and SLCA4 relative to CycloA and hTBP reference genes after 72h (S1_Data) (D) and 96h (S1_Data) (E) and following treatment with hemin (inducing differentiation) after 72h (S1_Data) (F) and 96h (S1_Data) (G). Cells were seeded and counted every day following labelling by trypan blue following treatment with hemin (inducing differentiation). Live cells are plotted in H and trypan blue labeled cells are plotted in I (S1_Data) (the corresponding plots for cells without hemin treatment can be seen in Fig. 6C and D). Plots represent the mean of three independent experiments ± SD (*<0.05, **P < 0.01, ***P<0.001 and ****P<0.0001 Student's t test).

Supplementary Table S1 Primers used in this study		
Primer name	**Description**	**Sequence**
F_CyclophA	Endogenouse control for qPCR	GTCAACCCCACCGTGTTCTT
R_CyclophA		CTGCTGTCTTTGGGACCTTGT
F_hTBP	Endogenouse control for qPCR	CGCCAGCTTCGGAGAGTTC
R_hTBP		ACAACCAAGATTCACTGTGGATACA
F_TAL1_total	TAL1-total mRNA	GGAGCCGGATGCCTTCCCTAT
R_TAL1 total		GCTCCCGGCTGTTGGTGAAG
F_TAL1_3'TUR	TAL1 3' UTR	CTAGTGGCTTGTCCTCACC
R_TAL1_3'TUR		TCACAGGATCTTCCAAACGT
F_TAL1_ΔEx3	TAL ΔExon 3 mRNA	CTAAATATGCCCCAGCGGGT
R_TAL1_ΔEx3		CATCAGTAATCTCCATCTC
F_TAL1_TSS1-4	TAL1-long mRNA	AAAAAGGGGGAAAGCAAAGA
R_TAL1_TSS1-4		CCTCTAGCTGGGGGTCACT
F_TAL1_Exo	Exogenous+Endogenous TAL1 with MIGR1 infection	TGGAGATTACTGATGGTCCCC
R_TAL1_Exo		TCATTCTTGCTGAGCTTCTTGT
F_TSS1	TAL1 TSS1	TCAGGACCGGGCCTCAAAATG
R_TSS1		GGTGTGGTCCTGGGCGATGTT
F_TSS2	TAL1 TSS2	CATGCGGTGGGATTGTGAGAATC
R_TSS2		TGAATAGGATCTCCACTCCGCC
F_TSS4	TAL1 TSS4	AGCGTAACTGCAGGCCTCT
R_TSS4		TGAATAGGATCTCCACTCCGCC
F_Luciferase	Luciferase mRNA	GGACTTGGACACCGGTAAGA
R_Luciferase		CTTGTCGATGAGAGCGTTTG
F_Renilla	Renilla mRNA	GGAATTATAATGCTTATCTACGTGC
R_Renilla		CTTGCGAAAAATGAAGACCTTTTAC
F_gRNA_-60_Enh_1	gRNA target sequences for -60 Enhancer	CACCGCGGAGTTTAGGGAACCAGA
R_gRNA_-60_Enh_1		AAACTCTGGTTCCCTAAACTCCGC
F_gRNA_-60_Enh_2		CACCCTGGTCGTGTTCGTCCTCGG
R_gRNA_-60_Enh_2		AAACCCGAGGACGAACACGACCAG
F_gRNA_-60_Enh_3		CACCGGGACTCGTAGTCTGCGGCT
R_gRNA_-60_Enh_3		AAACAGCCGCAGACTACGAGTCCC
F_gRNA_-60_Enh_4		CACCTTAGCGTGTCGCTGATCGTG
R_gRNA_-60_Enh_4		AAACCACGATCAGCGACACGCTAA
F_gRNA_Ex3_1		CACCGCGGCCCTTTAAGTCTCTCG

R_gRNA_Ex3_1	gRNA target sequences for TAL1 Exon 3	AAACCGAGAGACTTAAAGGGCCGC
F_gRNA_Ex3_2		CACCGCGCCCAGTTCGATGACTGG
R_gRNA_Ex3_2		AAACCCAGTCATCGAACTGGGCGC
F_gRNA_Ex3_3		CACCGAGGACTCAGCTGCACCATG
R_gRNA_Ex3_3		AAACCATGGTGCAGCTGAGTCCTC
F_gRNA_Ex3_4		CACCCGTCCCGTCCCTCTAGCTGG
R_gRNA_Ex3_4		AAACCCAGCTAGAGGGACGGGACG
F_-60_Enhancer_ChIP	TAL1 -60 Enhancer for ChIP PCR	GAGGTTAGCGTGTCGCTGAT
R_-60_Enhancer_ChIP		AGACACGTTCGGTGACAGTG
F_TAL1_Exon3_ChIP	TAL1 Exon 3 for ChIP PCR	AGGAGGACGTGGACACCTTC
R_TAL1_Exon3_ChIP		CCAGCTAGCGCTGAGTTTCC
F_ α_Globin	Erythroid Lineage markers	CTGGAGAGGATGTTCCTGTCCTTG
R_ α_Globin		CAGCTTAACGGTATTTGGAGGTCAT
F_ β_Globin		ACTTCAGGCTCCTGGGCAAC
R_ β_Globin		GGCCCTTCATAATATCCCCCAG
F_ γ_Globin		GGACAAGGCTACTATCACAA
R_ γ_Globin		CAGTGGTATCTGGAGGACAG
F_SLC4A1		TGACCGCATCTTGCTTCTGT
R_SLC4A1		GTCTTCACCCGCTTGACGTA
F_TRIB2	TAL1 known Target genes	TGGGGACATGCATTCCTTCG
R_TRIB2		GACCCGAGTCCTCTCTTCGT
F_NFKB1		CGCCGCTTAGGAGGGAGA
R_NFKB1		CCATTCTGAAGCCGGGTGG
F_NKX3-1		CCTCCCTGGTCTCCGTGTA
R_NKX3-1		TGTCACCTGAGCTGGCATTAC
F_MLL4	MLL4 (KMT2D) Lysine methyltransferase 2D	TGCTATCACAGAAGCCCATG
R_MLL4		GGATCAATGATATCTTCTGCAG
F_18SrRNA	Human 18S rRNA (Endogenous control for CPT)	GGCCCTGTAATTGGAATGAGTC
R_18SrRNA		CCAAGATCCAACTACGAGCTT