No evidence for transvection *in vivo* by a superenhancer:promoter

pair integrated into identical open chromatin at the *Rosa26* locus

Keiji Tanimoto^{1, 2, *}, Hitomi Matsuzaki^{1, 2}, Eiichi Okamura³, Aki Ushiki², Akiyoshi

Fukamizu^{1, 2}, and James Douglas Engel⁴

¹ Faculty of Life and Environmental Sciences, Life Science Center for Survival Dynamics,
Tsukuba Advanced Research Alliance (TARA), University of Tsukuba, Tsukuba, Ibaraki
305-8577, Japan

² Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki
305-8577, Japan

³ Graduate School of Biomedical Sciences, Tokushima University, Tokushima 770-8503, Japan

⁴ Department of Cell and Developmental Biology, University of Michigan, USA

* Corresponding author: Faculty of Life and Environmental Sciences,
University of Tsukuba, Tennoudai 1-1-1
Tsukuba, Ibaraki 305-8577, Japan
Phone/Fax: (+81) 29-853-6070
E-mail: keiji@tara.tsukuba.ac.jp

23 **Abstract**

24 Long-range associations between enhancers and their target gene promoters have been shown
25 to play critical roles in executing genome function. Recent variations of chromosome capture
26 technology have revealed a comprehensive view of intra- and inter-chromosomal contacts
27 between specific genomic sites. The locus control region of the β -globin genes (β -LCR) is a
28 super-enhancer that is capable of activating all of the β -like globin genes within the locus in *cis*
29 through physical interaction by forming DNA loops. CTCF helps to mediate loop formation
30 between LCR-HS5 and 3'HS1 in the human β -globin locus, in this way thought to contribute to
31 the formation of a “chromatin hub”. The β -globin locus is also in close physical proximity to
32 other erythrocyte-specific genes located long distances away on the same chromosome. In
33 this case, erythrocyte-specific genes gather together at a shared “transcription factory” for
34 co-transcription. Theoretically, enhancers could also activate target gene promoters on
35 different chromosomes *in trans*, a phenomenon originally described as transvection in
36 *Drosophilla*. Although close physical proximity has been reported for the β -LCR and the
37 β -like globin genes when integrated at the mouse homologous loci in *trans*, their structural and
38 functional interactions were found to be rare, possibly because a lack of suitable regulatory

elements that might facilitate *trans* interactions. Therefore, we re-evaluated presumptive transvection-like enhancer-promoter communication by introducing CTCF binding sites and erythrocyte-specific transcription units into both LCR-enhancer and β -promoter alleles, each inserted into the mouse *ROSA26* locus on separate chromosomes. Following cross-mating of mice to place the two mutant loci at the identical chromosomal position and into active chromatin in *trans*, their transcriptional output was evaluated. The results demonstrate that there was no significant functional association between the LCR and the β -globin gene *in trans* even in this idealized experimental context.

Introduction

Gene expression is tightly regulated by DNA *cis* elements and their binding *trans*-factors, in which specific enhancer-promoter communications play a pivotal role. While genome-wide sequencing of the human and mouse genomes disclosed the number of genes to be more than 20,000, the number of enhancer elements is predicted to far exceed the number of genes [1]. Because accumulating evidence suggests that perturbation of enhancer function can be a major cause of pathogenesis in human diseases [2], it is of paramount importance to assign the activity of any individual enhancer to a specific target gene(s) in order to predict its function. Recent genome-wide interactome analyses revealed that enhancers can *physically* interact with genes over enormous distances, exceeding several hundreds of kilobase pairs in *cis*, or even with genes located on different chromosomes in *trans* [3], suggesting the presence of molecular mechanisms that allow specific enhancer-promoter interactions to take place over very long distances.

In the interphase nucleus, the genome adopts a higher-order chromatin architecture, in which transcription factors play important roles. Among those, CTCF, first identified as a transcriptional activator or repressor and subsequently, as an insulator, binds to two distinct

genome regions to bring those two sites into close spatial proximity [4]. Interactome analysis in ES cells revealed that the number of intra- or inter-chromosomal interactions mediated by CTCF was 1,480 and 336, respectively [5]. However, how frequently gene expression is reflected by changes in CTCF-mediated genome architecture is not well understood. On the other hand, it has been reported that genes with similar transcriptional specificity migrate into transcription factories in the nucleus that are rich in transcription factors engaged in the expression of those genes [6]. According to this mechanism, two distinct genome regions carrying genes with the same expression pattern should meet at the shared foci for co-transcription.

The human β -like globin genes are organized within a 70-kbp span on human chromosome 11, with the embryonic ϵ -globin gene located most 5', followed by the two fetal γ -globin genes ($G\gamma$ and $A\gamma$), while the adult δ - and β -globin genes are at the 3' end of the locus (Fig. 1A). Expression of all the β -like globin genes in primitive, as well as in definitive erythroid cells, depends on the activity of the locus control region (LCR; [7, 8]), a super-enhancer element located 48 kbp 5' to the transcription initiation site of the β -globin gene. The LCR consists of five DNaseI hypersensitive sites (HSSs), among which HS1 to 4

are constituent enhancers and rich in binding sites for transcription factors [9], while HS5 carries CTCF binding sites [10].

How the distal LCR enhancer activates β -globin gene expression has long been a subject of intense debate [11]. In 2002, RNA TRAP [12] and chromatin conformation capture (3-C; [13, 14]) assays elegantly revealed that the LCR and β -globin promoters were positioned in close proximity: these observations were consistent with a looping model, in which proteins bound to the LCR enhancer and to the gene promoters physically interact with the intervening DNA sequences looped out [15]. Erythroid specific transcription factors, such as GATA-1, NF-E2 and EKLF are essential for efficient globin genes transcription through binding to both the LCR and globin gene promoters. It is therefore presumed that they participate somehow in long-range enhancer-promoter interactions. In fact, both GATA-1 and NF-E2 are essential for LCR and β major-globin proximity in murine erythroid cells [16, 17], as well as for LCR and γ -globin proximity in human erythroid cells [18]. Similarly, EKLF is required for loop formation between the LCR and β -globin promoter sites, but not for LCR-HS5 and 3'HS1 sites [19]).

Interestingly, CTCF binding was found around the HSSs at both ends of the locus,

i.e. LCR-HS5 and the 3'HS1 regions (Fig. 1A; [13]). Although 3C assays revealed proximal positioning of these sites in the nucleus, it was not confined to erythroid cells. In globin expressing cells, the LCR and 3'HS1 regions are further located proximally to the actively expressed β -globin genes, which structure has been termed an active chromatin hub (Fig. 1A; [13, 20]). Therefore, transcriptional activation of the β -like globin genes is predicted to be a multi-step process, in which HS5-3'HS1 interaction may help to bring LCR enhancer sequences within close proximity of the β -globin promoter, thus facilitating their productive interaction (Fig. 1A).

It has been reported that non-genic LCR sequences are transcribed in erythroid cells [21, 22]. Therefore, LCR and β -globin gene may be co-transcribed in the same RNA polymerase II (PolII) factory, which then aids their physical association and transcriptional activation of the β -globin gene by the LCR enhancer. In accord with this notion, the β -globin gene locus on mouse chromosome 7 was found to colocalize with erythroid specific genes located 20 Mb away on the same chromosome in erythroid cell nuclei [6]. Furthermore, the murine β -globin gene locus colocalized with the *Slc4a1* (Chr. 11) and *Cd47* (Chr. 16) genes at a shared polII factory [23].

112 Because the LCR makes direct contact with its targets through loop formation, it can
113 theoretically touch and activate such targets on separate chromosomes. Such *trans*
114 chromosomal interaction for functional enhancer-promoter communication was dubbed
115 transvection in *Drosophilla* [24, 25]. As mentioned above, recent development of
116 biochemical technologies disclosed that interchromosomal *physical* interaction was common
117 throughout the genome, not just in *Drosophilla* but also in mammals [26]. To test whether
118 inter-chromosomal *functional* as well as physical association between the LCR enhancers and
119 the β -like globin gene promoters take place, Noordermeer *et al.* knocked-in each sequence into
120 the mouse genome separately on homologous chromosomes [27]. Because they found
121 upregulation of endogenous murine β -like globin genes (~2-fold) on another chromosome,
122 they concluded that the LCR must have some affinity for the β -globin promoter even in *trans*.
123 However, neither frequent interchromosomal homologous chromosome interactions nor
124 transvection-like activation of reporter genes in *trans* was observed,
125 In their experimental design, though, CTCF-assisted or co-transcriptionally
126 mediated mechanisms were not considered. We therefore decided to re-evaluate transvection
127 in mammals by incorporating well-characterized β -globin *cis* elements at the *ROSA26* locus.

128 Firstly, LCR-HS5 and 3'HS1 sequences were introduced into enhancer and promoter alleles,
 129 respectively, in expectation that CTCF factors bound at these sites might promote the
 130 formation of an interchromosomal bridge, which would in turn facilitate functional interactions
 131 between the LCR enhancer elements (HS4~1) and the β -globin promoter. In addition,
 132 β -globin transcription units were included in both alleles, anticipating that two alleles bearing
 133 the same promoter would likely migrate into same PolII factory in the nucleus. Even in this
 134 idealized experimental design, however, functional association between two alleles on separate
 135 chromosomes was not observed.

136

Results

Generation of enhancer and promoter knock-in alleles at the mouse *Rosa26* locus

A targeting vector for the enhancer allele (Fig. 1B, top) carried the LCR (HSs 1~5) and the β -globin gene sequences. The one for the promoter allele (Fig. 1C, top) carried the β -globin gene and the 3'HS1 sequences. To distinguish human β -globin gene transcripts expressed from each allele by PCR, portions of the β -globin gene were replaced either with corresponding segments of the ϵ - and γ -globin genes in the enhancer and promoter targeting vectors, respectively. In this experimental design, expression of both chimeric genes is under the control of a common human β -globin proximal promoter element (1.6 kb).

In the absence of a *cis*-linked LCR, the human β -globin gene locus becomes heterochromatinized (e.g. in the Hispanic thalassemia patient; [28, 29]) and the expression of all the β -like globin genes is reduced in transgenic mice [30, 31]. Therefore, a β -globin transgene without an LCR enhancer in *cis* on the promoter allele is expected to be heterochromatinized and to not be efficiently activated by the LCR on the enhancer allele in *trans*. We therefore

152 chose the mouse *ROSA26* locus (on chromosome 6) for testing the transvection phenomenon,
 153 because this locus has a stable open chromatin structure in virtually all tissues [32].

154 To test for possible involvement, if any, of co-transcription of the β -globin genes in
 155 the enhancer and promoter alleles for transvection analysis, an ϵ -marked β -globin [$\beta(\epsilon)$ -globin]
 156 gene sequence was surrounded by loxP sites (floxed) in the enhancer allele so that it could be
 157 removed by conditional *in vivo* cre-loxP-mediated homologous recombination (Fig. 1B). To
 158 test for involvement of allelic proximity mediated by CTCF factors in the transvection
 159 experiment, the 3'HS1 sequence in the promoter allele was also floxed (Fig. 1C). Following
 160 homologous recombination with these targeting vectors in R1-ES cells, genomic DNA was
 161 prepared and correct recombination events were confirmed by Southern blot analyses using
 162 several combinations of restriction enzymes and specific probes (Fig. 1D and 1E).

163

164 ***In vivo* Cre-loxP recombination to derive daughter sublines**

165 Following establishment of germ line modified mouse lines from the mutant ES cells by
 166 co-culture aggregation, the one carrying the enhancer knock-in allele ($LCR+\beta(\epsilon)+Neo^r$; Fig. 2A,
 167 top) was mated with *cre*-expressing TgM to remove either the “Neo^r” or “Neo^r + $\beta(\epsilon)$ -globin

gene” sequences, thereby generating either “*LCR*+ $\beta(\epsilon)$ ” or “*LCR*” alleles, respectively (Fig. 2A). Similarly, “ $\beta(\gamma)$ +3’*HS1*” and “ $\beta(\gamma)$ ” alleles were derived from lines carrying the promoter knock-in allele ($\beta(\gamma)$ +3’*HS1*+*Neo*^r) by removing “*Neo*^r” or “*Neo*^r+3’*HS1*” sequences, respectively (Fig. 2B). Correct cre-loxP recombination events were confirmed by Southern blotting (Fig. 2C), as well as PCR analyses of tail genomic DNAs (Fig. 2D).

Evaluation of enhancer activity *in vivo*

To analyze chimeric β -globin gene expression in the knock-in mutant alleles, adult mice were made anemic, nucleated erythroid cells were collected from their spleens and recovered RNAs were reverse-transcribed. Expression of hybrid $\beta(\gamma)$ - and $\beta(\epsilon)$ -globin genes was analyzed using a primer set common to both chimeric gene sequences. Addition of 3’*HS1* to the $\beta(\gamma)$ -globin gene increased its expression level by only 1.6-fold *in vivo* (Fig. 3A), which was less prominent when compared with its effect in YAC TgM [33]. In human β -globin YAC TgM, deletion of 3’*HS1* sequences attenuated adult β -globin gene expression by more than 10-fold, possibly because this sequence plays additional roles in higher order chromatin organization at the human β -globin locus in the 150-kb YAC TgM [34].

184 The expression level of the $\beta(\epsilon)$ -globin gene linked to the LCR in *cis* was 16-fold
185 higher than that of the $\beta(\gamma)$ -globin gene in isolation (Fig. 3A). This magnitude seemed much
186 less significant than when LCR is deleted from the whole locus in endogenous [29] or
187 transgenic environments [35]. It is generally accepted that the LCR not only potentiates
188 promoter activity as an activator but also opens chromatin [29]. While this latter activity is a
189 part of the “enhancer” function in the context of the native β -globin locus, the *Rosa26* locus is
190 in an open chromatin configuration by its nature and therefore, the observed 16-fold activation
191 may represent a promoter potentiation function of the LCR alone.

192

193 **Evaluation of read-through transcription from the *Rosa26* promoter**

194 Because the enhancer and promoter constructs were integrated at the ubiquitously expressed
195 *Rosa26* gene locus, some portion of the chimeric β -globin gene transcription could be driven
196 by the *ROSA26* gene promoter. We therefore quantified how much read-through transcription
197 from the *Rosa26* gene promoter might contribute to expression of the β -globin gene sequence
198 (Fig. 3B). Total RNA was extracted from the spleen, liver and kidney of anemic adult mice
199 and the expression levels of $\beta(\gamma)$ -globin+3'HS1 gene and the mouse α -globin gene, relative to

the GAPDH gene expression, were determined by qRT-PCR. The mouse α -globin gene was preferentially expressed only in the spleen, as expected (Fig. 3B, left). In contrast, while $\beta(\gamma)$ -globin+3'HS1 was highly expressed in the spleen, its low level expression was also observed in the liver and kidney. Since α -globin gene expression was barely detected in these non-hematopoietic tissues (in adults), we concluded that contamination of erythroid cells in these tissues was negligible. Therefore, low level $\beta(\gamma)$ -globin+3'HS1 gene expression in the liver and kidney (and probably in the spleen) was under the control of the *Rosa26* promoter. In other words, it appears that at least 90% of $\beta(\gamma)$ -globin+3'HS1 gene transcription in the spleen initiates from the β -globin gene promoter.

Confirmation of CTCF binding to the LCR-HS5 and 3'HS1 regions

To test for CTCF binding to well-established CTCF binding sites in LCR-HS5 and 3'HS1 of the human β -globin locus in mutant animals, ChIP analyses were conducted using chromatin prepared from anemic spleen (erythroid) cells (Fig. 3C). PCR primers for *H19* ICR and *necdin* [4] loci were included as positive and negative controls for CTCF binding, respectively.

215 As expected, CTCF enrichment was observed at the LCR-HS5 and 3'HS1 regions, confirming
216 its binding to these sites *in vivo*.

217

218 **Cross-mating to derive animals carrying distinct pairs of enhancer and promoter alleles**

219 To generate four distinct combinations of enhancer- and promoter-knock-in alleles (Fig. 4A),
220 animals carrying heterozygous enhancer alleles ($LCR^{+/-}$ or $LCR+\beta(\epsilon)\text{-globin}^{+/-}$) and those
221 carrying homozygous promoter alleles ($\beta(\gamma)\text{-globin}^{+/+}$ or $\beta(\gamma)\text{-globin}+3'HSI^{+/+}$) were mated.
222 Theoretically, in the next generation, half of the litters carry both enhancer and promoter
223 alleles in *trans* and the other half carry promoter allele only (Fig. 4A). Derivation of animals
224 bearing the expected genotypes was confirmed by allele-specific PCR (Fig. 4B; results for
225 animals with promoter allele alone is not shown) and Southern blot analyses (data not shown).

226

227 **The test for functional interchromosomal enhancer-promoter interactions**

228 The expression of the hybrid $\beta(\gamma)$ -globin gene was compared in two animal groups carrying
229 either the promoter allele alone or both enhancer and promoter alleles in *trans* (Fig. 4C-F and

Supplemental Figure 1). Anemic spleens were collected from ~1 month old animals and the expression levels of human β -globin and mouse α -globin genes were determined by qRT-PCR. Human $\beta(\gamma)$ -globin gene expression normalized to that of the mouse α -globin gene in the two groups (with or without the LCR *in trans*) did not differ significantly, providing no evidence for *trans*-activation (Fig. 4C and Supple. Fig. 1A). Even when compared within single litters, no significant difference was observed between two groups (data not shown). This result was consistent with a report by Noordermeer *et al* [27], in which the LCR and γ -globin gene were individually integrated at a gene-dense site on mouse chromosome 8 and tested for inter-chromosomal interaction.

Next, the $\beta(\gamma)$ -globin gene with an attached 3'HS1 sequence was used as a reporter and the experiment was repeated (Fig. 4D and Supple. Fig. 1B). Because the LCR-HS5 and 3'HS1 sequences were bound by CTCF (Fig. 3C), it was possible that the LCR and 3'HS1 come into close proximity, which then facilitates trans-activation of $\beta(\gamma)$ -globin gene by the LCR. Even in this idealized experimental setting, however, no significant activation was observed.

245 Then, the combination of LCR+ $\beta(\epsilon)$ -globin and $\beta(\gamma)$ -globin genes as enhancer and
246 promoter alleles, respectively, was tested (Fig. 4E and Supple. Fig. 1C). Because
247 transcription units in these two alleles share the same transcriptional regulatory sequences (*i.e.*
248 the β -globin proximal promoter), it was possible that two alleles would migrate into a shared
249 transcription factory, which would then lead to $\beta(\gamma)$ -promoter activation by the LCR enhancer
250 in *trans*, caused by the close proximity of the two alleles. As shown in Fig. 4E, however, no
251 sign of *trans*-activation was detected.

252 Finally, the combination of the LCR+ $\beta(\epsilon)$ -globin and $\beta(\gamma)$ -globin+3'HS1 alleles
253 was investigated (Fig. 4F and Supple. Fig. 1D). Although in some of the litters, statistically
254 significant difference in the expression levels in between promoter alone and
255 enhancer+promoter alleles was observed (Supple. Fig. 1D), this significant but subtle
256 difference dissapeared when the sample number increased.

257 Discussion

258

259 Long-range chromatin interactions can occur over long distance between regions on the same
 260 chromosome in *cis* or on distinct chromosomes in *trans*. Because chromatin clustering events
 261 are dynamic, consequential formation of higher-order genome organization is believed to play
 262 an important role in regulating gene expression [36, 37]. In *Drosophila*, a pair of homologous
 263 chromosomes can physically interact to allow productive enhancer-promoter communication in
 264 *trans*, a phenomenon referred to as transvection [24, 25]. In the case of the *yellow* locus, for
 265 example, an enhancer of one copy of a gene (that lacks promoter activity) regulates the
 266 expression of the paired copy of the gene (lacking enhancer activity) in *trans*. More recently,
 267 many inter-chromosomal interactions were identified using 3C-based biochemical strategies in
 268 mammalian cells [38], and growing evidence suggests that transvection-like phenomena may be
 269 generally possible for eukaryotic enhancer-promoter communication.

270 Several mechanisms have been proposed for establishing preferential interactions
 271 between genomic loci, including enhancer-promoter communication via specific sets of
 272 transcription factors, co-migration of gene loci with similar transcriptional profiles into shared

273 PolII foci, or reflection of structural constraints mediated by architectural proteins such as
 274 CTCF/cohesin. These mechanisms may also account for inter-chromosomal chromatin
 275 interactions, such as those observed in the interferon- γ gene [39], the mouse *Hoxb1* gene [40],
 276 and the housekeeping *Rad23a* gene [41]. Among such examples, the *H19* imprinting control
 277 region (*H19* ICR) on mouse chromosome 7 has been shown to interact with regions on several
 278 different chromosomes in *trans*, including the maternally inherited insulin-like growth factor 2
 279 receptor (*Igf2r*) on chromosome 17, to suppress its transcription [42]. Because deletion of the
 280 *H19* ICR affected the expression of interacting genes on different chromosomes,
 281 interchromosomal interaction was suggested play a functional role in their transcriptional
 282 regulation.

283 To gain insight into possible inter-chromosomal gene regulatory mechanisms,
 284 Noordermeer *et al.* employed an LCR super-enhancer and human β -like globin promoter as
 285 regulatory elements to test for functional consequences (*i.e.* gene transcription) of placing them
 286 separately at corresponding *cis* locations on homologous chromosomes in mice [27] since these
 287 regulatory elements represent one of the most robust and most thoroughly examined
 288 enhancer-promoter pairs capable of interacting with each other over extremely long distances

289 [6]. Despite their clear affinity in the native chromatin configuration, they did not show
 290 significant inter-chromosomal interaction when integrated into the *Rad23a* gene locus [41].
 291 Our observations reported here are consistent with their results (*LCR* x $\beta(\gamma)$) in Fig. 4C).
 292
 293 Because enhancer-promoter looping in the β -globin gene locus is dependent on
 294 erythroid-specific transcription factors, such as EKLF and GATA-1, it has been anticipated that
 295 long-range chromatin loops were formed only in the cells where target genes are active [13].
 296 In fact, in the EKLF knock-out mice, significantly reduced chromatin loop formation between
 297 the LCR and β -globin gene was accompanying diminished β -globin gene transcription [19].
 298 Recent observations in mice, however, suggest that chromatin connectivity is not necessarily
 299 coupled to transcriptional events. For example, sonic hedgehog (*Shh*) gene transcription is
 300 regulated by a distal enhancer and they form a chromatin loop. This loop, however seems to
 301 be preset and detectable even when *Shh* is not transcribed [43]. In addition, among several
 302 interactions between regulatory sequences within the *HoxD* locus, some are present even in the
 303 absence of target gene activation [44, 45]. In the β -globin gene locus, formation of an active
 304 chromatin hub leading to erythroid-specific gene activation turned out to be a multistep process;

305 *i.e.* the chromatin loop between the LCR-HS5 and 3'HS1 was preformed in erythroid progenitor
 306 cells in a CTCF-dependent fashion prior to globin gene transcription [13]. Importantly, this
 307 developmentally early structure was not affected by EKLF ablation in the previously cited
 308 knock-out experiment [19]. It can therefore be assumed that, upon erythroid cell maturation,
 309 this prestructure (outer loop) may facilitate subsequent gene activation that accompanies inner
 310 loop formation between an enhancer and promoter, only when cell type-specific transcription
 311 factors are present.

312 By using a cell-based system, in which nuclear levels of GATA-1 can be
 313 experimentally manipulated, it was shown that β -globin gene activity, as well as the contact
 314 between the LCR and the gene required nuclear GATA-1 [16]. The knockdown of LDB1, a
 315 cofactor of GATA-1, also impaired both β -globin gene transcriptional activity and
 316 LCR- β -globin loop formation, presumably by affecting GATA-1 function [46]. Interestingly,
 317 when LDB1 was tethered to the β -globin promoter region via an artificial zinc-finger motif,
 318 chromatin looping and β -globin gene activity was partially restored, even in the absence of
 319 GATA1 [47]. Furthermore, forced loop formation between the LCR and the γ -globin promoter
 320 in adult erythroid cells caused ectopic upregulation of γ -globin transcription [48]. These

321 instances demonstrate that engineered formation of a chromatin loop alone between an enhancer
322 and promoter was sufficient to activate gene expression. In other words, certain forms of
323 chromatin interactions may play an instructive role in gene expression and set the permissive
324 condition for actual gene activation.

325

326 It has been reported that some fraction of CTCF binding sites were located at enhancer and
327 promoter regions and involved in their physical associations [49]. To facilitate
328 enhancer-promoter interactions in *trans*, we therefore placed binding sites for CTCF in both
329 enhancer and promoter alleles in our experiment. As mentioned above, it has been reported
330 that the LCR-HS5 and the 3'HS1 are in close proximity in both globin-expressing and
331 non-expressing erythroid cells, suggesting that they have physical affinity independent of
332 β -globin gene expression. However, we failed to observe upregulation of $\beta(\gamma)$ -globin gene
333 expression even in this idealized condition (Fig. 4D & F). While 3C analysis of our human
334 β -globin YAC TgM clearly identified a chromatin loop between the LCR and the β -globin
335 promoter in *cis*, any sign of *trans*-interaction was not validated by our current work (data not
336 shown). Recently, Hi-C analysis of the human genome demonstrated that chromatin loops

337 were often established between two opposed CTCF sites in opposite motif directions [50].

338 Furthermore, as for the establishment of CTCF/cohesin-mediated loops, an “extrusion model”

339 has been proposed [51], in which a loop-extruding molecule (*i.e.* cohesin) forms progressively

340 larger loops but are stalled at CTCF-bound sites (*i.e.* boundary elements). If this is the singular

341 mechanism that allows CTCF-mediated interaction between distinct genomic sites,

342 inter-chromosomal association via CTCF/cohesin machinery cannot be possible. However,

343 because a massive number of CTCF-associated inter-chromosomal interactions have been

344 reported [5], intervening chromatin should not absolutely be required for the formation of

345 chromatin interactions. Nevertheless, addition of CTCF sites on both enhancer and promoter

346 knock-in alleles did not augment their functional *trans* association in our experiments.

347

348 The majority of protein coding genes are transcribed by RNA polymerase II (PolII) and this

349 process is also engaged in long-range associations between gene loci [52]. It is conceivable

350 that coregulated genes (and enhancers) with common temporal and spatial specificities may

351 migrate to preassembled PolII factories for transcription and therefore their loci can be in close

352 proximity, even when they are separated by long distance in *cis* on the same chromosome or in

353 *trans* on chromosome pairs. Such a situation has been reported for genes on mouse
354 chromosome 7, which are involved in globin synthesis and regulation. They are
355 transcriptionally active in erythroid cells and their loci co-localized at shared PolIII foci with the
356 active β -globin genes. [6, 23]. Therefore, preferential long-range interactions between distinct
357 gene loci can be generated by genes being transcribed. In other words, two genomic elements
358 with spatial proximity in the nuclear space would have a chance to interact with each other.
359 Although it has been reported that the LCR is transcribed in erythroid cells [21], the enhancer
360 knock-in allele employed by Noordermeer *et al.* was unlikely to be transcribed efficiently
361 because of a lack of the ERV-9 LTR sequence [22]. In addition, although the *Rad23a* is a
362 housekeeping gene, it is not known if the knock-in alleles in that study were efficiently
363 transcribed. We therefore introduced β -globin transcriptional regulatory units both in enhancer
364 and promoter knock-in alleles at the *ROSA26* locus ($LCR+\beta(\epsilon) \times \beta(\gamma)$ or $\beta(\gamma)+3'$ HS1 in Fig. 4A).
365 Despite their significant expression in erythroid cells (Fig. 3A & B), however, we did not
366 observe increased reporter gene expression when compared to that in the absence of a paired
367 enhancer allele in *trans* (Fig. 4E & F).

368

369 In the examples of the *Shh* and the *HoxD* loci, higher order pre-structures have been proposed to
370 facilitate future promoter–enhancer contacts over very long distances [43, 44]. *Cis*-interaction
371 between the LCR-HS5 and 3'HS1 found at the endogenous β -globin locus (outer loop) may not
372 be sufficiently stable to facilitate LCR- β promoter interaction (inner loop) when they are
373 situated in *trans*. It is also possible that the function of the outer loop may simply be to isolate
374 a given locus from its surrounding chromatin environment [4, 13, 53]). Further functional
375 evaluation of the higher order chromatin structure is apparently required to ask if “form always
376 follows function”.

377

Materials and Methods

Targeting vectors

Floxed neomycin resistance gene (flNeo^r) cassette was released from pMC1neopA_5'/3'-loxP [54] by *XbaI*/*NheI* digestion and inserted into *XbaI* site of pROSA26-1 (generous gift of Dr. Philippe Soriano; nucleotide position at 180,029 in AC155722; RP24-204J8) to generate pROSA26/MC1neopA_5'/3'-loxP(-).

5'-upstream portion of human β -globin gene (nucleotides 60,577-60,882 in HUMHBB; U01317.1; GenBank) was PCR-amplified by using a set of oligonucleotides: ICI-02-5S; 5'-GGGGTACC TCTAGATCTCTATTTATTTAGCA-3' (artificial *KpnI* and *XbaI*, and endogenous *Bgl*III [at 60,557] sites are underlined) and ICI-02-3A; 5'-GGTCAGCGTAGGGTCTCAGT-3'. Following *KpnI* and *ApaI* (at 60,882) digestion, this fragment and 3'-downstream portion of the gene (*ApaI*-*XbaI* fragment; nucleotides 60,882-65,426) were linked and cloned into *KpnI*/*XbaI* sites of pBluescriptII KS(+). *Bam*HI site (at 60,676) of this plasmid was then eliminated to generate p β -globin_K-X- Δ B for facilitating cloning procedure.

394 Portions of ϵ - and A γ -globin gene sequences were PCR-amplified by using
395 following two sets of oligonucleotides: ICI-04-5S;
396 5'-GGCACCATGGTGCATTTTACTGCT-3' (artificial NcoI site underlined) and ICI-03-3A;
397 5'-TCAGGATCCACATGCAGCTT-3' (BamHI) or ICI-03-5S;
398 5'-CACACACTCGCTTCTGGAAC-3' and ICI-03-3A, respectively. Following *NcoI* (partial)
399 and *BamHI* digestion, ϵ (nucleotides 19,539-19,959) and A γ (nucleotides 39,465–39,885) gene
400 sequences were replaced with corresponding portion of β -globin gene (nucleotides
401 62,185-62,613) in p β -globin_K-X- Δ B to generate pm β / ϵ and pm β / γ , respectively.
402
403 [Promoter targeting vector]
404 Following two oligonucleotides were annealed, phosphorylated and inserted into *NdeI* site (at
405 65,287 in HUMHBB) of the β -globin gene in pm β / γ to generate pm β / γ _loxP(-): BTLX-5S;
406 5'-TATCGGATCCTATAACTTCGTATAATGTATGCTATACGAAGTTATAGA-3' and
407 BTLX-3A; 5'-TATCTATAACTTCGTATAGCATACATTATACGAAGTTATAGGATCCGA-3'.
408 In each oligo, loxP sequences are italicized and *BamHI* sites underlined. The *XbaI* fragment,
409 carrying β -globin promoter, γ sequence-marked β -globin coding region, and a loxP site, was

410 released from pm β / γ _loxP(-) and introduced into *Xba*I site of

411 pROSA26/MC1neopA_5'/3'-loxP(-) to generate pR26/loxP-Neo/m β / γ .

412 Human β -globin 3'HS1 sequence (4,194 bp, *Sma*I-*Hind*III) was subcloned from

413 BAC clone, RP11-1205H24 (nucleotides 35,934-40,127 in AC129505; GenBank). Upon

414 conversion of *Hind*III site to *Sma*I site, the fragment was introduced into *Sma*I site of

415 pR26/loxP-Neo/m β / γ to generate pR26/loxP-Neo/m β / γ /3'HS1 (Promoter targeting vector).

416

417 [Enhancer targeting vector]

418 Following two oligonucleotides were annealed, phosphorylated and inserted into *Bgl*II site (at

419 60,557 in HUMHBB) of the β -globin gene in pm β / ϵ to generate pm β / ϵ _loxP(-): ICI-08-5S;

420 5'-GATCGGCGCGCCATAACTTCGTATAATGTATGCTATACGAAGTTAT-3' and ICI-08-3A;

421 5'-GATCATAACTTCGTATAGCATACATTATACGAAGTTATGGCGCGCC-3'. In each oligo,

422 loxP sequences are italicized and *Asc*I sites underlined. The *Xba*I fragment, carrying a loxP

423 site, β -globin promoter, and ϵ sequence-marked β -globin coding region, was released from

424 pm β / ϵ _loxP(-) and introduced into *Xba*I site of pROSA26/MC1neopA_5'/3'-loxP(-) to generate

425 pR26/loxP-Neo/m β / ϵ .

426 For facilitating cloning procedure, following two oligonucleotides were annealed,
427 phosphorylated and inserted into *KpnI/SacI* site of pBluescriptII KS(+): ICI-01-5S; 5'-
428 GGCGCGCCGGTACCTATGCGGCCGCGGCGCGCCAGCT-3' and ICI-01-3A;
429 5'-GGCGCGCCGCGGCCGCATAGGTACCGGCGCGCCGTAC-3', which resulted in
430 *AscI-KpnI-NotI*-[*SacII*]-*AscI* sites formation. Human β -globin LCR sequences (17,590 bp,
431 *EcoRI-XbaI*) were recovered from pRS/LCR [55] as *KpnI-NotI* (both in the multicloning sites)
432 fragment and inserted into *KpnI-NotI* site of above plasmid to generate pBS/LCR. Upon
433 eliminating *SacII* site (parenthesized site in the above double-stranded oligo) from pBS/LCR, 3'
434 downstream region of HS1 was accidentally deleted and the final LCR size cloned was (17,198
435 bp, nucleotides 95,131-77,934 in AC104389; Ensemble). Finally, LCR sequences were
436 recovered from pBS/LCR as *AscI* fragment and introduced into *AscI* site of
437 pR26/loxP-Neo/m β /ε to generate pR26/LCR/loxP-Neo/m β /ε (Enhancer targeting vector).

438

439 **Gene targeting in ES cells and generation of mutant mice**

440 Target vectors were linearized by *SacII* digestion. R1-ES cells were grown on embryonic
441 fibroblast feeder cells. Following electroporation (Bio-Rad GenePulser Xcell [0.4 mm gap] at

442 setting of 250 V and 500 microfarads) of cells (1.0×10^7 cells) with a linearized targeting vector
 443 (20 μ g), cells were selected in 0.4 mg/ml G418. Homologous recombination in ES cells was
 444 first screened by PCR and then confirmed by Southern blotting with several combinations of
 445 restriction enzymes and probes shown below.

446 ROSA26-5' (*EcoRV-SalI*) probe: nucleotides 181,927 - 182,316 (in AC155722)

447 ROSA26-3' (*XbaI-XbaI*) probe: nucleotides 173,746 - 174,359 (in AC155722)

448 3'-neo (*PstI-BamHI*, 621 bp) probe: 918 - 1,543 (in U43611; Genbank)

449

450 Chimeric mice were generated by a coculture method using eight-cell embryos from CD1 mice
 451 (ICR, Charles River Laboratories), bred with CD1 females, and germ line transmission of the
 452 mutant allele was determined by PCR and Southern blot analyses.

453 TgM ubiquitously expressing cre recombinase were mated with knock-in mice to
 454 partially or completely execute Cre-loxP recombination, which was confirmed by PCR and
 455 Southern blot analyses of tail DNA of offsprings.

456 [Primers for enhancer allele]

457 LCR: ROSA5FL-5S2: 5'-CCCTCGTGATCTGCAACTCC-3'

458 ROSA-LCR2: 5'-TCACTTTTGGAGGTCAGGAA-3'

459 ε - β : BT-1S: 5'-AACTGTGTTCACTAGCAACCTCAA-3'

460 EP-1A: 5'-GGGCTTGAGGTTGTCCATGTTT-3'

461 Neo^r: Neo-S: 5'-AGAGGCTATTCGGCTATGAC-3'

462 Neo-AS: 5'-CACCATGATATTCGGCAAGC-3'

463 [Primers for promoter allele]

464 γ - β : BT-1S & GM-1A: 5'-CCTTGAGATCATCCAGGTGCTTT-3'

465 3'HS1: h3'HS1-5S: 5'-AGAAAGTTTGATGAACTACTTCTGACCC-3'

466 h3'HS1-3A: 5'-GACACCCACACATGTCCTGCC-3'

467 Neo^r: Neo-down-5S2: 5'-GACAGAATAAAACGCACGGGT-3'

468 ROSA-3A2: 5'-TGGGGCTAAAATGAGTGTTTC-3'

469 [Southern blot probe]

470 ROSA-3'-383 probe (*ScaI*-*HindIII* 390 bp): 179,269 -179,658 (in AC155722)

471

472 **Animal procedures**

473 Mice were housed in a pathogen-free barrier facility in a 12-hour light/12-hour dark cycle, and

fed standard rodent chow. Adult mice were sacrificed by cervical dislocation and the organs were immediately removed and flash-frozen in liquid nitrogen.

Animal experiments were performed in a humane manner under approval from the Institutional Animal Experiment Committee of the University of Tsukuba. Experiments were performed in accordance with the Regulation of Animal Experiments of the University of Tsukuba and the Fundamental Guidelines for Proper Conduct of Animal Experiments and Related Activities in Academic Research Institutions under the jurisdiction of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Expression analysis

Total RNA was extracted from phenyl hydrazine-induced anemic adult spleens (1 to 2 months old) by ISOGEN (Nippon Gene) and converted to cDNA using ReverTra Ace qPCR RT Master Mix with gDNA Remover (TOYOBO). One-fortieth of the reaction mixture was subjected to quantitative PCR amplification using the KOD SYBR qPCR Mix (Toyobo) and thermal Cycler Dice (TaKaRa Bio) with the following parameters: 95°C for 5s and 60°C for 30s, 40 cycles. The PCR primer sets used for human $\beta(\gamma)$ - or $\beta(\epsilon)$ -globin genes amplification were GM-1S2 and

490 BT-3A3 (126-bp amplicon) or BT-1S3 and EP-3A (152-bp), respectively. Primer set common
491 to both $\beta(\gamma)$ - and $\beta(\epsilon)$ -globin genes amplification in Fig. 3A was BT-4S1 and BT-4A1. The
492 primer sets used for internal control gene expression analyses were MAgI-I and II (mouse
493 α -globin gene) or mGAPDH-5S and 3A (mouse *gapdh* gene).

494 GM-1S2: 5'-GCCATAAAGCACCTGGATGAT-3' (39,811-39,831 in HUMHBB)

495 BT-3A3: 5'-GGCCAGCACACAGACCAGCACG-3' (63,493-63,514)

496 BT-1S3: 5'-CAACTGTGTTCCTACTAGCAACCT-3' (62,153-62,174)

497 EP-3A: 5'-GGGTCCAGGGGTAAACAACG-3' (19,761-19,780)

498 BT-4S1: 5'-GTGGATCCTGAGAACTTCAG-3' (55,212-55,231)

499 BT-4A1: 5'-GATAGGCAGCCTGCACTGGT-3' (63,541-63,560)

500 MAgI-I: 5'-TGAGGTCAATGAAGGGGTCGT-3'

501 MAgI-II: 5'- CCTTTCCAGGGCTTCAGCTCCATAT-3'

502 mGAPDH-5S: 5'-AAAATGGTGAAGGTCGGTGTG-3'

503 mGAPDH-3A: 5'-TGAGGTCAATGAAGGGGTCGT-3'

504

505 **ChIP analysis**

506 The animals (2 to 4 months old) bearing both the *LCR*+ $\beta(\epsilon)$ and $\beta(\gamma)$ +3'*HS1* knock-in alleles in
 507 *trans* were made anemic and nucleated erythroid cells were collected from their spleens.
 508 Following fixation with 1% formaldehyde for 10 min at room temperature. Nuclei (2×10^7
 509 cells) were digested with 12.5 units/ml of micrococcal nuclease at 37°C for 20 min. The
 510 chromatin was incubated with anti-CTCF antibody (D31H2; Cell Signaling Technology) or
 511 purified rabbit IgG (Invitrogen) overnight at 4°C and was precipitated with preblocked
 512 Dynabeads protein G magnetic beads (Life Technologies, Carlsbad, CA). Immunoprecipitated
 513 materials were then washed and reverse cross-linked. DNA was purified with the QIAquick
 514 PCR purification kit (Qiagen, Venlo, The Netherlands) and subjected to qPCR analysis. The
 515 endogenous *H19* ICR and *Necdin* sequences were analyzed as positive and negative controls,
 516 respectively [56]. *LCR*-HS5 and 3'*HS1* primer sets are as follows:

517 HS5-CTCF-5S2: 5'-GGTCACAGAATAACCTGAGT-3'

518 HS5-CTCF-3A: 5'-CAAAAGGGCTCCTTAACAAC-3'

519 3'*HS1*-CTCF-5S2: 5'-TCACTGAAGTAGGGAGGGAAGAA-3'

520 3'*HS1*-CTCF-3A2: 5'-AAGGTCATTCCTTTAATGGTCTTTTC-3'

521

References

1. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi: 10.1038/nature11247. PubMed PMID: 22955616; PubMed Central PMCID: PMC3439153.
2. Smith E, Shilatifard A. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol*. 2014;21(3):210-9. doi: 10.1038/nsmb.2784. PubMed PMID: 24599251.
3. Chakalova L, Debrand E, Mitchell JA, Osborne CS, Fraser P. Replication and transcription: shaping the landscape of the genome. *Nat Rev Genet*. 2005;6(9):669-77. Epub 2005/08/12. doi: 10.1038/nrg1673. PubMed PMID: 16094312.
4. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. 2006;20(17):2349-54. Epub 2006/09/05. doi: 10.1101/gad.399506. PubMed PMID: 16951251; PubMed Central PMCID: PMC1560409.
5. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet*. 2011;43(7):630-8. Epub 2011/06/21. doi: 10.1038/ng.857. PubMed PMID: 21685913; PubMed Central PMCID: PMC3436933.
6. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*. 2004;36(10):1065-71. Epub 2004/09/14. doi: 10.1038/ng1423. PubMed PMID: 15361872.
7. Epner E, Reik A, Cimborra D, Telling A, Bender MA, Fiering S, et al. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell*. 1998;2(4):447-55. Epub 1998/11/11. PubMed PMID: 9809066.
8. Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. Locus control regions. *Blood*. 2002;100(9):3077-86. doi: 10.1182/blood-2002-04-1104. PubMed PMID: 12384402; PubMed Central PMCID: PMC3436933.
9. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A Phase Separation Model for Transcriptional Control. *Cell*. 2017;169(1):13-23. doi: 10.1016/j.cell.2017.02.007. PubMed PMID: 28340338; PubMed Central PMCID: PMC5432200.
10. Tanimoto K, Sugiura A, Omori A, Felsenfeld G, Engel JD, Fukamizu A. Human beta-globin locus control region HS5 contains CTCF- and developmental stage-dependent

- enhancer-blocking activity in erythroid cells. *Mol Cell Biol.* 2003;23(24):8946-52.
PubMed PMID: 14645507; PubMed Central PMCID: PMC309639.
11. Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene
activation. *Genes Dev.* 1999;13(19):2465-77. PubMed PMID: 10521391.
12. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. Long-range chromatin regulatory
interactions in vivo. *Nat Genet.* 2002;32(4):623-6. Epub 2002/11/12. doi: 10.1038/ng1051.
PubMed PMID: 12426570.
13. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The beta-globin
nuclear compartment in development and erythroid differentiation. *Nat Genet.*
2003;35(2):190-4. Epub 2003/10/01. doi: 10.1038/ng1244. PubMed PMID: 14517543.
14. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between
hypersensitive sites in the active beta-globin locus. *Mol Cell.* 2002;10(6):1453-65. Epub
2002/12/31. PubMed PMID: 12504019.
15. Choi OR, Engel JD. Developmental regulation of beta-globin gene switching. *Cell.*
1988;55(1):17-26. PubMed PMID: 3167976.
16. Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, et al. Proximity
among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1.
Mol Cell. 2005;17(3):453-62. Epub 2005/02/08. doi: 10.1016/j.molcel.2004.12.028.
PubMed PMID: 15694345.
17. Du MJ, Lv X, Hao DL, Zhao GW, Wu XS, Wu F, et al. MafK/NF-E2 p18 is required for
beta-globin genes activation by mediating the proximity of LCR and active beta-globin
genes in MEL cell line. *Int J Biochem Cell Biol.* 2008;40(8):1481-93. Epub 2008/03/01.
doi: 10.1016/j.biocel.2007.11.004. PubMed PMID: 18308612.
18. Woon Kim Y, Kim S, Geun Kim C, Kim A. The distinctive roles of erythroid specific
activator GATA-1 and NF-E2 in transcription of the human fetal gamma-globin genes.
Nucleic Acids Res. 2011;39(16):6944-55. Epub 2011/05/26. doi: 10.1093/nar/gkr253.
PubMed PMID: 21609963; PubMed Central PMCID: PMC3167640.
19. Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, et al. The active
spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes
Dev.* 2004;18(20):2485-90. Epub 2004/10/19. doi: 10.1101/gad.317004. PubMed PMID:
15489291; PubMed Central PMCID: PMC3529536.
20. Patrinos GP, de Krom M, de Boer E, Langeveld A, Imam AM, Strouboulis J, et al.
Multiple interactions between regulatory regions are required to stabilize an active

- 588 chromatin hub. *Genes Dev.* 2004;18(12):1495-509. Epub 2004/06/17. doi:
589 10.1101/gad.289704. PubMed PMID: 15198986; PubMed Central PMCID:
590 PMCPMC423198.
- 591 21. Routledge SJ, Proudfoot NJ. Definition of transcriptional promoters in the human beta
592 globin locus control region. *J Mol Biol.* 2002;323(4):601-11. PubMed PMID: 12419253.
- 593 22. Long Q, Bengra C, Li C, Kutlar F, Tuan D. A long terminal repeat of the human
594 endogenous retrovirus ERV-9 is located in the 5' boundary area of the human beta-globin
595 locus control region. *Genomics.* 1998;54(3):542-55. doi: 10.1006/geno.1998.5608.
596 PubMed PMID: 9878258.
- 597 23. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, et al. Preferential
598 associations between co-regulated genes reveal a transcriptional interactome in erythroid
599 cells. *Nat Genet.* 2010;42(1):53-61. Epub 2009/12/17. doi: 10.1038/ng.496. PubMed
600 PMID: 20010836; PubMed Central PMCID: PMCPMC3237402.
- 601 24. Tartof KD, Henikoff S. Trans-sensing effects from *Drosophila* to humans. *Cell.*
602 1991;65(2):201-3. PubMed PMID: 2015622.
- 603 25. Duncan IW. Transvection effects in *Drosophila*. *Annu Rev Genet.* 2002;36:521-56. doi:
604 10.1146/annurev.genet.36.060402.100441. PubMed PMID: 12429702.
- 605 26. Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a
606 comprehensive technique to capture the conformation of genomes. *Methods.*
607 2012;58(3):268-76. doi: 10.1016/j.ymeth.2012.05.001. PubMed PMID: 22652625;
608 PubMed Central PMCID: PMCPMC3874846.
- 609 27. Noordermeer D, de Wit E, Klous P, van de Werken H, Simonis M, Lopez-Jones M, et al.
610 Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat Cell*
611 *Biol.* 2011;13(8):944-51. Epub 2011/06/28. doi: 10.1038/ncb2278. PubMed PMID:
612 21706023; PubMed Central PMCID: PMCPMC3151580.
- 613 28. Driscoll MC, Dobkin CS, Alter BP. Gamma delta beta-thalassemia due to a de novo
614 mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proc Natl*
615 *Acad Sci U S A.* 1989;86(19):7470-4. PubMed PMID: 2798417; PubMed Central PMCID:
616 PMCPMC298086.
- 617 29. Forrester WC, Epner E, Driscoll MC, Enver T, Brice M, Papayannopoulou T, et al. A
618 deletion of the human beta-globin locus activation region causes a major alteration in
619 chromatin structure and replication across the entire beta-globin locus. *Genes Dev.*
620 1990;4(10):1637-49. PubMed PMID: 2249769.

- 621 30. Magram J, Chada K, Costantini F. Developmental regulation of a cloned adult beta-globin
622 gene in transgenic mice. *Nature*. 1985;315(6017):338-40. PubMed PMID: 3858676.
- 623 31. Townes TM, Lingrel JB, Chen HY, Brinster RL, Palmiter RD. Erythroid-specific
624 expression of human beta-globin genes in transgenic mice. *EMBO J*. 1985;4(7):1715-23.
625 PubMed PMID: 2992937; PubMed Central PMCID: PMC54408.
- 626 32. Soriano P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet*.
627 1999;21(1):70-1. doi: 10.1038/5007. PubMed PMID: 9916792.
- 628 33. Liu Q, Bungert J, Engel JD. Mutation of gene-proximal regulatory elements disrupts
629 human epsilon-, gamma-, and beta-globin expression in yeast artificial chromosome
630 transgenic mice. *Proc Natl Acad Sci U S A*. 1997;94(1):169-74. PubMed PMID: 8990180;
631 PubMed Central PMCID: PMC19271.
- 632 34. Fang X, Xiang P, Yin W, Stamatoyannopoulos G, Li Q. Cooperativeness of the higher
633 chromatin structure of the beta-globin locus revealed by the deletion mutations of DNase I
634 hypersensitive site 3 of the LCR. *J Mol Biol*. 2007;365(1):31-7. doi:
635 10.1016/j.jmb.2006.09.072. PubMed PMID: 17056066; PubMed Central PMCID:
636 PMC2826273.
- 637 35. Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent, high-level
638 expression of the human beta-globin gene in transgenic mice. *Cell*. 1987;51(6):975-85.
639 PubMed PMID: 3690667.
- 640 36. Dekker J. Gene regulation in the third dimension. *Science*. 2008;319(5871):1793-4. doi:
641 10.1126/science.1152850. PubMed PMID: 18369139; PubMed Central PMCID:
642 PMC2666883.
- 643 37. Bonev B, Cavalli G. Organization and function of the 3D genome. *Nat Rev Genet*.
644 2016;17(12):772. doi: 10.1038/nrg.2016.147. PubMed PMID: 28704353.
- 645 38. Dekker J, Misteli T. Long-Range Chromatin Interactions. *Cold Spring Harb Perspect Biol*.
646 2015;7(10):a019356. doi: 10.1101/cshperspect.a019356. PubMed PMID: 26430217;
647 PubMed Central PMCID: PMC4588061.
- 648 39. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations
649 between alternatively expressed loci. *Nature*. 2005;435(7042):637-45. doi:
650 10.1038/nature03574. PubMed PMID: 15880101.
- 651 40. Wurtele H, Chartrand P. Genome-wide scanning of HoxB1-associated loci in mouse ES
652 cells using an open-ended Chromosome Conformation Capture methodology.
653 *Chromosome Res*. 2006;14(5):477-95. doi: 10.1007/s10577-006-1075-0. PubMed PMID:

654 16823611.

655 41. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear
656 organization of active and inactive chromatin domains uncovered by chromosome
657 conformation capture-on-chip (4C). *Nat Genet.* 2006;38(11):1348-54. doi:
658 10.1038/ng1896. PubMed PMID: 17033623.

659 42. Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, et al. Circular
660 chromosome conformation capture (4C) uncovers extensive networks of epigenetically
661 regulated intra- and interchromosomal interactions. *Nat Genet.* 2006;38(11):1341-7. doi:
662 10.1038/ng1891. PubMed PMID: 17033624.

663 43. Amano T, Sagai T, Tanabe H, Mizushina Y, Nakazawa H, Shiroishi T. Chromosomal
664 dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and
665 active transcription. *Dev Cell.* 2009;16(1):47-57. doi: 10.1016/j.devcel.2008.11.011.
666 PubMed PMID: 19097946.

667 44. Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, et al. A regulatory
668 archipelago controls *Hox* genes transcription in digits. *Cell.* 2011;147(5):1132-45. doi:
669 10.1016/j.cell.2011.10.023. PubMed PMID: 22118467.

670 45. Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, et al. A switch
671 between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science.*
672 2013;340(6137):1234167. doi: 10.1126/science.1234167. PubMed PMID: 23744951.

673 46. Song SH, Hou C, Dean A. A positive role for *NLI/Ldb1* in long-range beta-globin locus
674 control region function. *Mol Cell.* 2007;28(5):810-22. doi: 10.1016/j.molcel.2007.09.025.
675 PubMed PMID: 18082606; PubMed Central PMCID: PMCPMC2195932.

676 47. Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, et al. Controlling long-range
677 genomic interactions at a native locus by targeted tethering of a looping factor. *Cell.*
678 2012;149(6):1233-44. doi: 10.1016/j.cell.2012.03.051. PubMed PMID: 22682246;
679 PubMed Central PMCID: PMCPMC3372860.

680 48. Deng W, Rupon JW, Krivega I, Breda L, Motta I, Jahn KS, et al. Reactivation of
681 developmentally silenced globin genes by forced chromatin looping. *Cell.*
682 2014;158(4):849-60. doi: 10.1016/j.cell.2014.05.050. PubMed PMID: 25126789; PubMed
683 Central PMCID: PMCPMC4134511.

684 49. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory
685 sequences in the mouse genome. *Nature.* 2012;488(7409):116-20. doi:
686 10.1038/nature11243. PubMed PMID: 22763441; PubMed Central PMCID:

687 PMCPMC4041622.

688 50. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D
689 map of the human genome at kilobase resolution reveals principles of chromatin looping.
690 Cell. 2014;159(7):1665-80. Epub 2014/12/17. doi: 10.1016/j.cell.2014.11.021. PubMed
691 PMID: 25497547; PubMed Central PMCID: PMCPMC5635824.

692 51. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of
693 Chromosomal Domains by Loop Extrusion. Cell Rep. 2016;15(9):2038-49. doi:
694 10.1016/j.celrep.2016.04.085. PubMed PMID: 27210764; PubMed Central PMCID:
695 PMCPMC4889513.

696 52. Mitchell JA, Fraser P. Transcription factories are nuclear subcompartments that remain in
697 the absence of transcription. Genes Dev. 2008;22(1):20-5. doi: 10.1101/gad.454008.
698 PubMed PMID: 18172162; PubMed Central PMCID: PMCPMC2151011.

699 53. Oti M, Falck J, Huynen MA, Zhou H. CTCF-mediated chromatin loops enclose inducible
700 gene regulatory domains. BMC Genomics. 2016;17:252. Epub 2016/03/24. doi:
701 10.1186/s12864-016-2516-6. PubMed PMID: 27004515; PubMed Central PMCID:
702 PMCPMC4804521.

703 54. Matsuzaki H, Okamura E, Takahashi T, Ushiki A, Nakamura T, Nakano T, et al. De novo
704 DNA methylation through the 5'-segment of the H19 ICR maintains its imprint during
705 early embryogenesis. Development. 2015;142(22):3833-44. doi: 10.1242/dev.126003.
706 PubMed PMID: 26417043.

707 55. Leach KM, Nightingale K, Igarashi K, Levings PP, Engel JD, Becker PB, et al.
708 Reconstitution of human beta-globin locus control region hypersensitive sites in the
709 absence of chromatin assembly. Mol Cell Biol. 2001;21(8):2629-40. doi:
710 10.1128/MCB.21.8.2629-2640.2001. PubMed PMID: 11283243; PubMed Central PMCID:
711 PMCPMC86894.

712 56. Okamura E, Matsuzaki H, Sakaguchi R, Takahashi T, Fukamizu A, Tanimoto K. The H19
713 imprinting control region mediates preimplantation imprinted methylation of nearby
714 sequences in yeast artificial chromosome transgenic mice. Mol Cell Biol.
715 2013;33(4):858-71. doi: 10.1128/MCB.01003-12. PubMed PMID: 23230275; PubMed
716 Central PMCID: PMCPMC3571351.

717

718 **Figure legends**

719

720 **Figure 1. Generation of enhancer and promoter knock-in alleles in mice**

721 **(A)** Structure of the human β -globin gene locus shown in 1D (left) and 3D (right) views. **(B)**

722 The enhancer targeting vector carrying the human β -globin LCR and β -globin gene that is

723 marked by an ϵ -globin sequence, wild-type *ROSA26* locus, and the correctly targeted enhancer

724 knock-in locus are shown. In the targeting vector, neomycin resistance (Neo^r) and diphtheria

725 toxin (DT)-A genes are shown as striped and solid boxes, respectively. The solid triangles

726 indicate the loxP sequences. Probes used for Southern blot analyses in (D) are shown as filled

727 rectangles. Expected restriction fragments with their sizes are shown beneath the partial

728 restriction enzyme maps. **(C)** The promoter targeting vector carrying the human β -globin gene

729 (marked by an γ -globin sequence) and 3'HS1, wild-type *ROSA26* locus, and the correctly

730 targeted promoter knock-in locus are shown. Probes used for Southern blot analyses in (E) are

731 shown as filled rectangles. **(D and E)** Genomic DNA from ES clones was digested with

732 restriction enzymes, separated on agarose gels, and Southern blots were hybridized to the

733 probes. Asterisks denote nonspecific bands.

Figure 2. Derivation of enhancer/promoter-allele variants by *in vivo* Cre-loxP

recombination

(A) Enhancer knock-in mouse bearing the $LCR+\beta(\epsilon)+Neo^r$ locus was mated with Cre-TgM to induce *in utero*, partial cre-loxP recombination, which resulted in selective excision of either Neo^r or $Neo^r+\beta(\epsilon)$ -globin sequences to generate $LCR+\beta(\epsilon)$ or LCR alleles, respectively. A, *AseI*; B, *BamHI*; 36, *Bsu36I*. (B) Similarly, $\beta(\gamma)+3'HSI$ or $\beta(\gamma)$ alleles were derived from the promoter knock-in mouse bearing the $\beta(\gamma)+3'HSI+Neo^r$ locus by deletion of the Neo^r or $Neo^r+3'HSI$ sequences, respectively. (C) Successful cre-loxP recombination was confirmed by Southern blot analysis. Tail genomic DNA of mutant mice were digested with *AseI* (enhancer knock-in series) or *Bsu36I* (promoter knock-in series), separated on agarose gels, and Southern blots were hybridized to the ROSA-3'-383 probe shown in panels A and B. (D) Each allele was discriminated by multiplex PCR analyses of tail genomic DNA from mutant mice. The LCR, ϵ - β and Neo^r sequences in the enhancer knock-in alleles were amplified by PCR primers shown by paired open arrowheads in panel A. The γ - β , 3'HSI and Neo^r sequences in the promoter knock-in alleles were amplified by PCR primers shown by paired open arrowheads in panel B.

750 **Figure 3. Expression of human β -globin genes in knock-in mice**

751 **(A)** Total RNA was prepared from spleens of 1-month-old anemic mice (N=4 for each
752 genotype). Expression levels of the human $\beta(\gamma)$ - or $\beta(\epsilon)$ -globin genes (analyzed by common
753 primer set targeted at β -globin sequence) and endogenous mouse α ($m\alpha$)-globin gene were
754 analyzed by semi-quantitative RT-PCR. The ratio of h β / $m\alpha$ -globin genes was calculated (the
755 expression value of the $\beta(\gamma)$ -globin was set at 1). **(B)** Total RNA was prepared from spleens,
756 livers and kidneys of 1-month-old anemic mice (N=4 for each tissues). Expression of
757 $m\alpha$ -globin, $\beta(\gamma)$ -globin and endogenous mouse (m)GAPDH genes was analyzed by
758 semi-quantitative RT-PCR. The expression levels of $m\alpha$ - (left panel) or $\beta(\gamma)$ -globin (right)
759 genes, both compared to that of mGAPDH gene were calculated (the expression values in
760 spleen samples were set at 100). **(C and D)** ChIP was conducted for CTCF in the spleen cells
761 of anemic animals bearing both $LCR+\beta(\epsilon)$ and $\beta(\gamma)+3'HSI$ alleles. The *Necdin* gene and the
762 *H19* ICR sequences were analyzed as negative and positive controls, respectively.
763 Quantitative PCR was repeated at least three times for each sample. Fold enrichment of CTCF
764 relative to IgG control (average values with S.D.) was calculated and graphically depicted
765 (average value of negative controls was set at 1.0).

Figure 4. Expression of $\beta(\gamma)$ -globin genes in knock-in mice

(A) Schematic representation of four different combinations of enhancer and promoter knock-in alleles to test for enhancer-promoter interaction *in trans*. (B) Mouse genotypes shown in (A) were confirmed by multiplex PCR analyses of tail genomic DNA of mutant mice. The ϵ - β , LCR, γ - β and 3'HS1 sequences were amplified by PCR primers shown by paired open arrowheads in panel A. (C-F) Total RNA was prepared from spleens of 1-month-old anemic mice. Numbers analyzed for each genotype are shown in the Supplemental Figure 1. Expression of $\beta(\gamma)$ -globin and endogenous α -globin genes was analyzed by semi-quantitative RT-PCR. The ratio of h $\beta(\gamma)$ -globin / α -globin genes was calculated and average value with S.D. was graphically depicted for each genotype group (Although values are arbitrary, they can be quantitatively compared between the panels).

Figure 1

bioRxiv preprint doi: <https://doi.org/10.1101/393363>; this version posted August 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.