

# **An intergenic non-coding RNA promoter required for histone modifications in the human $\beta$ -globin chromatin domain**

Emmanuel Debrand<sup>1</sup>, Lyubomira Chakalova<sup>1</sup>, Joanne Miles<sup>1</sup>, Yan-Feng Dai<sup>1</sup>, Beatriz Goyenechea<sup>1</sup>, Sandra Dye<sup>1</sup>, Cameron S. Osborne<sup>1</sup>, Alice Horton<sup>1</sup>, Susanna Harju-Baker<sup>2</sup>, Ryan C. Pink<sup>3</sup>, Daniel Caley<sup>3</sup>, David R. F. Carter<sup>3</sup>, Kenneth R. Peterson<sup>2,5</sup>, and Peter Fraser<sup>1,4,5</sup>

1. Laboratory of Chromatin and Gene Expression, The Babraham Institute, Babraham Research Campus, Cambridge, CB2 4AT, UK.

2. Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, Kansas City, KS, 66160, USA.

3. Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, OX3 0BP, UK.

4. Department of Biological Science, Florida State University, Tallahassee, FL, 32306, USA

5. To whom correspondence should be addressed: [kpeterson@kumc.edu](mailto:kpeterson@kumc.edu), tel: (913) 588-6907; [pfraser@bio.fsu.edu](mailto:pfraser@bio.fsu.edu), tel: (850) 644-5037.

## Abstract

Transcriptome analyses show a surprisingly large proportion of the mammalian genome is transcribed; much more than can be accounted for by genes and introns alone. Most of this transcription is non-coding in nature and arises from intergenic regions, often overlapping known protein-coding genes in sense or antisense orientation. The functional relevance of this widespread transcription is unknown. Here we characterize a promoter responsible for initiation of an intergenic transcript located approximately 3.3 kb and 10.7 kb upstream of the adult-specific human  $\beta$ -globin genes. Mutational analyses in  $\beta$ -YAC transgenic mice show that alteration of intergenic promoter activity results in ablation of H3K4 di- and tri-methylation and H3 hyperacetylation extending over a 30 kb region immediately downstream of the initiation site, containing the adult  $\delta$ - and  $\beta$ -globin genes. This results in dramatically decreased expression of the adult genes through position effect variegation in which the vast majority of definitive erythroid cells harbor inactive adult globin genes. In contrast, expression of the neighboring  $\epsilon$ - and  $\gamma$ -globin genes is completely normal in embryonic erythroid cells, indicating a developmentally specific variegation of the adult domain. Our results demonstrate a role for intergenic non-coding RNA transcription in the propagation of histone modifications over chromatin domains and epigenetic control of  $\beta$ -like globin gene transcription during development.

## Introduction

A staggering proportion of the mammalian genome is transcribed. More than half of the transcribed genomic regions in mammalian cells are intergenic regions [1-3] producing vast amounts of non-coding RNAs [4]. Such intergenic transcription patterns can be complex, with multiple overlapping sense, and antisense transcripts detected for a single locus [5]. During the process of mammalian development, proper temporal and spatial expression of genetic information must be achieved. The multigene  $\beta$ -globin locus has been intensively studied as a model for these events. In humans, it spans over 70 kb on chromosome 11 and its five genes (*HBE*, *HBG1*, *HBG2*, *HBD* and *HBB*) are arranged in the order of their developmental expression in erythroid cells. The locus includes a locus control region (LCR) [6], located 6-22 kb upstream of the embryonic *HBE* gene, which acts as a long-range regulatory element that physically interacts with the active globin genes during transcription [7, 8].

Our previous studies [9] have shown that the locus is divided into at least three differentially activated and developmentally regulated chromatin subdomains of 20-30 kb each (the LCR,  $\epsilon\gamma$ , and  $\delta\beta$  domains). Each domain displays differential general sensitivity to DNase I and extensive developmentally regulated H3K4 di and tri-methylation (H3K4me2 and H3K4me3) and H3 hyperacetylation [10]. Each of the domains are also delineated by non-coding intergenic transcription [9], similar to the *Hox* clusters [11, 12], correlating with their differential nuclease sensitivity and active histone modifications. Intergenic transcripts initiating from multiple discrete sites can be found in or near the LCR [13-16] at all stages of development in erythroid cells [9, 17], consistent with a developmentally stable open chromatin conformation. Intergenic transcripts in the gene domains appear during development in coordination with active expression of the globin genes within the embryonic or adult domains. In the case of the adult-specific  $\delta\beta$  domain, 5' RACE

has shown that intergenic transcripts initiate near the 5' boundary of the DNase I sensitive domain [9]. The analysis of transgenic mice carrying a human  $\beta$ -globin locus yeast artificial chromosome ( $\beta$ -YAC) with a 2.5kb deletion [18] encompassing the intergenic transcript initiation site suggested that the region is necessary for the normal DNase I sensitive chromatin domain surrounding the adult genes and high-level, non-variegated expression of *HBD* and *HBB*. Interestingly, a long-range chromatin interaction between the LCR and this region was recently revealed by 5C analysis of the *HBB* locus, and analysis of the various naturally occurring deletion of this region, results in defective HBG silencing, supporting the suggestion that the intergenic promoter has regulatory potential. In patients carrying the Corfu deletion, *HBG* transcription persists in adult erythroid cells where it effectively competes with the adult *HBB* gene for the activating effects of the LCR [19]. Analysis of intergenic transcripts in Corfu erythroid cells revealed persistent activation of an intergenic promoter upstream of the *HBG* genes (Goyenechea, Chakalova and Fraser, unpublished data) in a region implicated in HbF control. This tight correlation between non-coding transcription domains and domains of increased chromatin accessibility suggested that transcription through chromatin might propagate remodelling or histone modifying activities over large regions. This view is supported by findings of histone acetyltransferases (HATs) and methyltransferases associating with the elongating form of RNA polymerase II and the increasing number of studies linking transcription with simultaneous histone modification. However in other instances the ncRNAs themselves have been implicated in recruiting chromatin modifying activities to chromatin to affect gene expression [12, 20-23]. We previously showed that intergenic transcription in the globin locus is cell cycle-dependent, occurring in S and G1 phases [9, 10], suggesting that it may be involved in the 're-establishment' and/or spreading of an active chromatin state after replication and mitosis.

We hypothesized that regulation of chromatin structural domains plays a key role in control of developmental globin gene expression and that intergenic transcription is required for domain remodeling and subsequent gene activation. To discriminate between cause and effect, as well as effects of other potential regulatory sequences in the larger 2.5kb deletion that was previously analyzed, we generated transgenic mice carrying a 213 kb  $\beta$ -YAC transgene with a 300 bp sequence, corresponding to the putative  $\delta\beta$  promoter and transcriptional start site of the adult intergenic transcript, flanked by *loxP* sites. We show that alteration of intergenic promoter activity results in an adult stage-specific reduction in intergenic transcription and a domain-wide, markedly diminished histone modification profile compared to the normal active locus in transgenic mice, leading to dramatically reduced expression of the adult  $\beta$ -like globin genes.

## **Results**

### **Promoter activity in the region upstream of the adult $\beta$ -globin domain**

We previously showed through primer extension and 5'RACE that intergenic transcripts from the 5' region of the adult-specific  $\beta$ -globin domain appeared to initiate from a discrete site approximately 3.3 kb upstream of the *HBD* gene in transgenic mice [9]. To determine whether sequences in this region have genuine promoter activity, and if so, to perform a preliminary characterization of the promoter elements, we cloned various sub-fragments from -685 to +40 relative to the 5' end of the transcript in front of a promoter-less EGFP gene (Figure 1A). The resultant plasmids were linearized and stably transfected into the human K562 cell line, which expresses both erythroid and myeloid markers. Populations of transfected cells were assessed for EGFP expression by flow cytometry. The results show that the region indeed has promoter activity (Figure 1B). Expression of EGFP was low but occurred in a significantly higher proportion of transfected cells compared to the promoter-less construct itself or non-transfected cells. The data

indicate that most of the promoter activity is contained within the 150-200 bp region immediately upstream of the start site of the intergenic transcript. Truncation down to -50 (Xba) completely abolished promoter activity. The promoter appears to lack a canonical TATAA-like sequence within 30 bp of the start site, although there is an overlapping GATA1 TATAA sequence at approximately -190 bp. However deletion of that sequence from the promoter results in only a slight decrease in EGFP expression (compare Apo to Bst11). In addition, the region from -150 to 0 bp contains putative transcription factor binding sites indicative of possible promoter function. We conclude that the region has promoter activity and that a minimal promoter containing the first few hundred base pairs upstream of the start site retains promoter activity in stably transfected cells.

### **Analysis of transcription upstream of the adult $\beta$ -globin domain**

To investigate the pattern of intergenic transcription throughout the globin locus in human cells we utilized an *ex vivo* model of erythroid development. Nucleated cells were isolated from human peripheral blood and induced to differentiate in a two-phase culture system. During the second phase of growth the induction of HBB expression was measured and changes were observed in cell morphology characteristic of erythroid differentiation (not shown). We and others have previously shown that this is a good model for studying the activation of the adult  $\beta$ -globin domain [19, 24]. Total RNA was extracted at different time points during this erythroid progression, reverse-transcribed, fluorescently labelled and then hybridized to a custom high-density tiling array. At an early time-point (day 4) we saw no significant peak of transcription over the putative  $\delta\beta$ -promoter (Figure 1A), consistent with the relatively lower levels of *HBB* expression and lack of hemoglobinization at this stage of differentiation. Over time we saw an increase in signal over the putative intergenic promoter which peaked at around day 7 of erythroid differentiation (Figure 2A).

This is consistent with an increase in *HBB* expression at these later stages (not shown). The lack of a peak in the day 4 sample suggests that the peak observed at later time-points is not caused by locally increased non-specific background binding to the probes in this region, but is due to an increase in transcripts in the area. However, to confirm the latter we performed quantitative RT-PCR using primers upstream and downstream of the putative signal. The results show an increase in the ratio of downstream to upstream signal as erythroid differentiation proceeds (Figure 2B), which is consistent with increased transcription from a putative promoter in this region. To determine the directionality of transcription we performed a strand-specific RT-PCR using either Downstream-F (the forward primer, which detects a sense transcript) or Downstream-R (the reverse primer, which detects an antisense transcript) in the cDNA-generation step. Only the Downstream-F primer was able to generate a cDNA that could be amplified in PCR (Figure 2C), suggesting that the transcript downstream of the putative  $\delta\beta$ -promoter is fired in the direction towards the adult  $\beta$ -globin sub-domain. Taken together these data suggest that the  $\delta\beta$ -promoter is capable of firing transcripts in erythroid cells from humans as well as transgenic mice and reinforces the validity of using the latter as a model to study the role of the  $\delta\beta$ -promoter in  $\beta$ -globin regulation.

### **Generation of YAC transgenic mice with deletion of the $\delta\beta$ intergenic promoter**

To investigate the contribution of the  $\delta\beta$  intergenic promoter to the regulation of gene activity and chromatin structure in the globin locus, we designed a 213 kb YAC transgene construct containing the entire human  $\beta$ -globin locus with *loxP* sites flanking the transcriptional initiation site, from -266 to +40 (Figure 3A). The unmodified YAC has been used extensively in transgenic mice [25-30] and shown to be sufficient for normal, copy number-dependent, position-independent expression of the human beta-globin genes throughout development. The modified YAC was used

to generate three transgenic mouse lines using standard microinjection procedures. The structural integrity of the transgene was analyzed by *SfiI* digestion of genomic DNA followed by pulsed field gel electrophoresis and Southern blot using multiple probes across the locus. As can be seen in Figure 3B each of the lines harbors a characteristic large *SfiI* fragment that hybridizes to all probes from the 5' end to the 3' end of the *SfiI* fragment. This indicates that each line contains at least one intact copy of the *HBB* locus. PCR cloning and sequence analysis showed that lines FX101 and FX115 contained the two flanking *loxP* (floxed) sites whereas line FX14 contained only a single *loxP* site located downstream of the  $\delta\beta$  promoter. Of the two 'floxed' lines we chose FX115 for further study, based on the fact that it contained a single major band containing regions complimentary to all the tested probes verifying the structural integrity of the transgene. This line was bred with a Cre recombinase expressing transgenic line, Deleter Cre [31] to successfully generate a deletion line with a single, intact 115 kb *SfiI* fragment (Figure 3C). Southern blot (Figure 3D) and PCR analysis (not shown) confirmed that the deletion line  $\Delta 115$ , retains a single copy of the human  $\beta$ -globin locus with the 306 bp floxed  $\delta\beta$  promoter region removed.

### **Embryonic globin gene expression is unaffected by the $\delta\beta$ promoter deletion**

To ascertain the effect of the intergenic promoter deletion on globin gene expression we used RNA FISH to study the transcriptional regulation of the globin genes throughout development. Two wild type YAC transgenic lines with equivalent copy numbers were used for comparison. Line FX14 contains 2 copies of the same human beta-globin YAC as the floxed and deleted lines but has only a single *loxP* site in the position downstream of the  $\delta\beta$  promoter. Line 264W contains a single intact copy of a shorter human beta-globin YAC (150 kb) which contains the entire globin locus. This line has previously been shown to express the *HBB* genes at full levels, equivalent to the endogenous mouse *Hbb* genes [32] and has normal patterns of intergenic transcription and



histone modifications [10]. Transgenic mice containing the entire human globin locus and intact LCR normally express *HBE* and *HBG* at high levels in E10.5 embryonic red cells. RNA FISH analysis shows that embryonic expression of *HBE* and *HBG* is normal for lines FX115 and  $\Delta$ 115 (Figure 4A) with approximately 90% of the embryonic cells being positive for *HBE* and/or *HBG* transcription signals (Figure 4B). This degree of transcriptional activity is equivalent to the two wild type lines 264W and FX14, and previously published full-locus transgenic lines [33, 34], as well as the endogenous mouse *Hbb* locus in embryonic blood (>90%) [35]. Line FX 101 showed a similar high-level of transcription signals (>90%) in embryonic blood cells (data not shown). These results demonstrate that transcription of the human *HBE* and *HBG* genes is normal and non-variegated in these lines and is not affected by the intergenic  $\delta\beta$  promoter deletion, implying position independent transgene expression provided by a fully functioning LCR.

### **Intergenic promoter deletion severely impacts adult globin gene transcription**

The results in adult erythroid cells were strikingly different from the embryonic globin gene transcription pattern. Unexpectedly, we found that transcription of the adult-specific *HBB* gene was reduced 2-3 fold in the FX101 and FX115 lines in which the  $\delta\beta$  intergenic promoter is still present with flanking *loxP* sites (Figure 4C and D). Approximately 35% (34.6 %; n=321) of the erythroid cells in FX115 were positive for a human  $\beta$ -globin transcription signal compared to >90% in the wild type lines (see below). Strikingly, we found that deletion of the intergenic promoter (line  $\Delta$ 115) further reduced transcription of *HBB* compared to FX115. Only 5.8% (n=431) of the erythroid cells were positive for an *HBB* transcription signal in  $\Delta$ 115, approximately 6-fold lower than FX115 and 15-20 fold lower than the wild type lines. *HBB* transcription in both wild type lines were similar to the endogenous *Hbb* locus, in which >90% of  $\beta$ -globin loci are positive for transcription signals [35]. The dramatic reduction in adult-specific transcription of *HBB* in the

intergenic promoter deletion line  $\Delta 115$  shows that the 300bp sequence contains an element essential for normal high-level expression of the adult  $\beta$ -globin genes. This was confirmed by RT-PCR analysis, which showed that expression of both adult-type beta-genes; *HBB* and *HBD* are dramatically reduced in the deletion line compared to the wild type and floxed lines (Figure 4E). These results suggest the minimal intergenic promoter is necessary for high-level expression of both *HBB* and *HBD*, suggesting a possible domain-wide effect.

### **Flanking *loxP* sites interfere with intergenic promoter activity**

An unexpected finding from the above experiments was the fact that both floxed lines FX101 and FX115 showed reduced transcription of the adult-specific *HBB* gene even prior to deletion of the intergenic promoter (Figure 4D). This raised the question of potential structural problems in our YAC transgene, which may have been introduced, for example, during the process of homologous recombination in yeast. Pulsed-field gel analysis indicated that each of the YAC lines in question contained an intact *SfiI* fragment that hybridized with several probes across the locus suggesting that the bulk of the  $\beta$ -globin locus was intact (Figure 3B and C). Southern blot analyses using probes around the  $\delta\beta$  promoter did not reveal any obvious rearrangements (data not shown and Figure 3D). Similarly, sequence analysis of a 3.3 kb PCR product amplified from genomic DNA from line FX115 showed that the genomic sequence spanning the entire targeted region was the same as wild type except for the insertion of the *loxP* sites, thus ruling out rearrangements in that area during homologous recombination. We then considered the possibility that the inserted *loxP* sites interfered with intergenic promoter activity. To test this we measured intergenic transcript levels downstream of the  $\delta\beta$  promoter in FX115 and  $\Delta 115$  via RT-PCR. We found that both lines had reduced intergenic transcript levels compared to the wild type transgenic

strains 264W and FX14 (Figure 5A), which have similar transgene copy numbers. These results suggested that the *loxP* sites may indeed be affecting intergenic promoter activity in FX115.

To further investigate this possibility we used genomic DNA from FX115 and  $\Delta$ 115 to amplify PCR fragments containing the intergenic promoter regions out to approximately -700 relative to the intergenic transcription start site. These fragments were cloned in front of a promoter-less EGFP reporter plasmid with a flanking neo resistance gene. The plasmid constructs were sequenced to verify that no mutations occurred during the amplification or cloning procedures, and were then stably transfected into mouse erythroleukemia (MEL) cells. EGFP expression was measured in neo-resistant, transfected populations by flow cytometry. The wild type  $\delta\beta$  intergenic promoter drove consistent, low-level expression of the EGFP reporter in a very high percentage (66%) of stably transfected cells, compared to the potent CMV promoter/enhancer combination that we used as a positive control (Figure 5B). The CMV promoter population had only 51% EGFP positive cells but the level of expression per cell varied much more widely and was in general much higher than the  $\delta\beta$  promoter. As expected, the  $\Delta$ 115 promoter region with the 300 bp deletion showed no EGFP expression being essentially identical to mock transfected cells. Interestingly, the  $\delta\beta$  intergenic promoter with flanking *loxP* sites from FX115 had dramatically reduced EGFP expression, only slightly higher than the promoter deletion construct. These results show that the presence of the *loxP* sites partially interferes with intergenic promoter activity by dramatically decreasing the percentage of cells that achieve expression. Deletion of the 300 bp region leads to more or less complete abolition of promoter activity. These results provide a potential explanation for the reduced percentage of cells with *HBB* transcription signals in FX115 and FX101. We conclude that  $\delta\beta$  promoter activity and intergenic transcription are essential for normal high-level activation of the adult  $\beta$ -globin genes.

### **Absence of intergenic transcription results in variegated transgene expression**

In order to determine whether the absence of transcription of *HBB* in most erythroid cells is due to an instability of transcriptional activity, i.e. oscillating between the ON and OFF states in every cell, or the result of variegation of expression in which only a subset of cells stably express the *HBB* transgene, we performed RNA FISH with exonic *HBB* probes which detect both primary transcripts in the nucleus and processed *HBB* mRNAs in the cytoplasm[34]. We observed that only the erythroid cells exhibiting a *HBB* primary transcript signal in the nucleus had spliced *HBB* mRNA transcripts in the cytoplasm and vice-versa (Figure 6, and Figure 4D). This demonstrates that disruption of  $\delta\beta$  promoter activity leads to developmental stage-specific variegation of the *HBB* gene in erythroid cells, in which the vast majority of erythroid cells (~90% in  $\Delta 115$ ) never switch on or activate the human beta-globin transgene locus [33]. These findings reinforce the concept that the  $\delta\beta$  intergenic promoter element is essential to confer the correct epigenetic state of the adult domain containing the *HBD* and *HBB* genes.

### **Absence of intergenic transcription results in altered histone modification profile**

We previously showed that YAC transgenes with a much larger (2.5 kb)  $\delta\beta$  promoter deletion resulted in a 2- to 3-fold reduction in general DNase I sensitivity over the entire adult chromatin domain indicating that the normal, stage-specific chromatin opening of the domain surrounding the *HBD* and *HBB* genes failed in the absence of intergenic transcription [9]. In most cases studied, a DNase I sensitive chromatin conformation correlates with the appearance of “active” epigenetic modifications to histone tails in the region. Using both human primary erythroid cells and erythroid cells from transgenic mice 264WT at different developmental stages, we identified developmentally specific sub-domains of histone H3 hyper-acetylation and H3K4 di- and tri-methylation that correlate with the occurrence of intergenic transcription and increased general

DNase I sensitivity in the human beta-globin locus [10]. Having shown that deletion of the 300 bp core intergenic promoter results in severe variegation of the *HBB* and *HBD* genes, we sought to investigate the effect on histone modifications over this domain. We performed chromatin immunoprecipitations (ChIP) with antibodies directed against histone tail modifications characteristic of active chromatin followed by real time PCR to assess the effect of the deletion on the histone modification profile of the transgene locus in adult erythroid cells.

The ChIP data shows that deletion of the  $\delta\beta$  promoter results in a dramatic reduction of H3K4 di- and tri-methylation and H3 hyper-acetylation across the entire adult domain (Figure 7), from approximately 3.5 kb upstream of the *HBD* gene, extending over 30 kb downstream to 3' HS1 which may correspond to the 3' boundary of the adult domain [36]. The decreases ranged from 2 to 10 fold compared to the normal *HBB* transgene locus (264W) [10]. Reduction in histone modifications were not restricted to the gene promoters but occurred over intergenic regions throughout the domain. These results show that the spreading of active histone modifications across the entire adult domain is not simply dependent on the presence of intact gene promoters and the LCR, but requires the  $\delta\beta$  intergenic promoter.

A surprising finding from these experiments was that the  $\delta\beta$  promoter deletion also resulted in a significant reduction in active histone modifications throughout the LCR domain. The LCR was initially described as a dominant control element capable of chromatin activation at any integration site in the genome. Those early studies used constructs with genes directly linked to the LCR rather than at long range and normal domain context. Our results show that the LCR, though fully active in embryonic cells, is not capable of maintaining an active profile in adult erythroid cells in the absence of an active gene domain within the locus. RNA FISH in  $\Delta 115$  embryonic and early fetal liver cells (Figure 4 A, B and not shown) shows normal expression of the embryonic

globin genes *HBE* and *HBG*, implying full LCR function in primitive and early definitive erythroid cells. Normal, high level expression of these genes was confirmed by RT-PCR analysis (not shown). These results suggest some form of activation cooperativity or co-dependency occurs between the LCR and active gene domains in the YAC transgene.

## Discussion

Intergenic transcription is a widespread phenomenon. Aside from the handful of individual gene loci in which intergenic transcripts have been recognized and studied, it is now apparent from various transcriptome studies that most of the sequence of a mammalian genome is transcribed at one time or another, and that the vast majority of these transcribed regions are non-coding [37]. Many of these non-coding transcripts overlap known coding transcript regions and can be sense or antisense. In many cases, the non-coding transcripts correlate with activity of nearby genes and appear to be regulated in a similar fashion. In other cases, especially where antisense transcripts are involved it has been proposed that these transcripts may play a repressive role. Here we present the characterization of the  $\delta\beta$  intergenic promoter, which initiates sense intergenic transcription toward the downstream *HBD* and *HBB* genes. We show that deletion of a 300 bp sequence that contains the  $\delta\beta$  promoter results in dramatic alterations to the normal active histone modification profile over a 30 kb downstream region in YAC transgenic mice. As a result, expression of the adult-specific *HBD* and *HBB* genes is severely compromised in erythroid tissues. Our data show that intergenic promoter activity is knocked down by insertion of *loxP* sites flanking the  $\delta\beta$  promoter, with a corresponding knock down of adult gene expression. It is not clear why the inserted *loxP* sites cause this knockdown of promoter activity. One possibility is that the upstream *loxP* site disrupts a critical *cis* element needed for full intergenic promoter activity. A search for potential transcription factor binding sites indicates that the upstream *loxP* sequence disrupts a

potential E-box binding site, and could potentially bind TAL1 and/or E47. Transgenic line FX14, which has only one *loxP* site in the downstream position, does not suffer from a decrease in intergenic promoter activity, further implicating the upstream site. However, deletion of the region containing this putative binding site results in a doubling of the number of transfected cells that express the EGFP reporter gene in stably transfected K562 cells (compare *BstXI* and *Apo* in Figure 1B). This is seemingly at odds with the transgenic data but is important to point out that K562 cells express embryonic and fetal globins, HBE and HBG, and do not appear to initiate sense transcripts from the  $\delta\beta$  promoter [38]. In combination, these data suggest that this sequence element may be of importance in regulating  $\delta\beta$  intergenic promoter activity by repressing transcription in primitive erythroid cells and promoting it in definitive cells.

Our results demonstrate that transcription through the region downstream of the  $\delta\beta$  promoter, or the non-coding RNAs produced, or both, play a role in the activation and/or maintenance of the active chromatin structure of the domain. We cannot rule out the possibility that the region also contains an enhancer needed for gene activation, though this seems less likely as it would not account for the wider changes in chromatin structure. The process of transcription itself could potentially affect chromatin structure in a variety of ways. The simplest explanation is that passage of the polymerase results in a transient unfolding of the higher-order chromatin fiber and partial disassembly of nucleosomes. This may create an accessibility window in which, for example, sequence specific binding proteins jump onto their cognate sites [39]. Similarly, disruption may also increase access for chromatin remodelling complexes or histone modification enzymes. Transcription coupled disassembly may also provide the opportunity to insert variant histones such as H3.3 into the transcribed region via the replication independent chromatin assembly pathway [40]. The finding that histone acetyltransferases and methyltransferases are

associated with the elongating form of the polymerase, also suggest a possible role for transcription in propagation of modified chromatin domains [41]. Finally, transcription coupled disruption could alter the binding of repressive protein complexes such as the Polycomb group proteins which are thought to be displaced from PREs by intergenic transcription, thus acting as an epigenetic switch [42].

The enormous complexity of the emerging RNA world suggests that the non-coding RNAs themselves also have a role to play in gene control. Some of the most well characterized functional non-coding RNAs, such as *Xist* and *Air* are repressive in nature; however more recent findings have indicated enhancer roles for non-coding RNAs [43]. Non-coding RNAs in the *Drosophila* dosage compensation complex are involved in 2-fold up-regulation of genes on the male X chromosome [44]. The production of non-coding RNAs in the *Hox* clusters have also been tightly correlated with modified chromatin domains [11, 12]. It is possible that epigenetic control of the human  $\beta$ -globin locus involves both transcriptional and non-coding RNA mediated mechanisms since neither is mutually exclusive.

Intergenic transcription could also potentially play a role in nuclear positioning of the beta-globin locus. Studies on the mouse  $\beta$ -globin locus have shown that activation of the locus during erythroid differentiation is LCR independent and involves re-positioning of the locus away from centromeric heterochromatin in correlation with histone hyperacetylation [45, 46]. The fact that histone modification of the human locus is dependent on intergenic transcription, suggests that intergenic transcription may also correlate with nuclear re-positioning. For example, intergenic transcription through the endogenous  $\beta$ -globin locus, which in some cases initiates hundreds of kb upstream [10, 47] could keep the locus tethered to a transcription factory which may facilitate near-continuous, high-level transcription of the active globin genes [48, 49], while at the same time



maintaining a heritable epigenetic memory of the active state. A role for the LCR in contributing to promotion or maintenance an active chromatin profile in the locus may also be proposed. Dostie et al., [50] demonstrated a long-range interaction with the  $\delta\beta$  promoter region in human erythroid cell line, suggesting that the LCR may stimulate intergenic transcription that, in turn, may promote an active state. Other sequences in the  $\gamma\delta$  region between the  $^A\gamma$ -globin gene (*HBG2*) and the  $\delta$ -globin gene (*HBD*), including the *HBBP1* pseudogene ( $\psi\beta$ -globin) and the *BGLT3* non-coding gene that is between *HBG2* and *HBBP1*, contribute to long range interaction [51, 52] Our data show that the  $\delta\beta$  promoter region is a novel *cis*-regulatory element that is well downstream of *HBBP1* pseudogene, but upstream of the *HBD* gene.

A note of caution must be added to these studies since the locus under study is a transgene. Incongruent results have often been obtained when comparisons are made between mutations in the human  $\beta$ -globin transgene locus and seemingly similar mutations in the endogenous mouse  $\beta$ -globin locus. Part of this may be due to the well-documented fact that human transgenes in mouse erythroid cells do not exactly recapitulate the regulatory pattern of the endogenous locus. Our analysis of *ex-vivo* erythroid precursors from patients with a naturally occurring deletion of the  $\delta\beta$  promoter helps to illustrate this point [19]. The Corfu thalassemia mutation is a 7.2 kb deletion that removes the  $\delta\beta$  promoter and the 5' end of the *HBD* gene. Unlike the situation in transgenic mice, the deletion does not lead to a general shut down of the affected globin locus and LCR. Instead we found that intergenic transcripts appeared to initiate from an intergenic promoter upstream of the fetal-specific *HBG* genes. This correlated with an open chromatin structure over the *HBG* genes and the downstream *HBB* gene, and high-level transcription of both *HBG* genes and *HBB* in the adult cell environment. It is tempting to speculate that binding sites for BCL11A or the PYR

repressor complex, factors implicated in *HBG* silencing, are deleted in the larger Corfu mutation [53, 54], which may partly explain the differences in *HBG* silencing.

## **Material and Methods**

### **Human primary erythroid cell culture**

Samples of venous blood (50 ml) were collected from informed and consenting volunteers using EDTA vacuette tubes and spun on Histopaque-1077 (Sigma) to isolate the mononuclear cells. Following six PBS platelet removal washes these were resuspended at  $1 \times 10^6$  cells/ml in phase one media: Stemspan medium (Stemcell Tech),  $1 \mu\text{g/ml}$  cyclosporin A (Sigma),  $20 \text{ ng/ml}$  IL-3 (Peprotech),  $20 \text{ ng/ml}$  IL-6 (Peprotech) and  $50 \text{ ng/ml}$  stem cell factor (Miltenyi Biotech). These were incubated for 24 hours at  $37^\circ\text{C}$ ; non-adherent cells were transferred to a new flask and incubated for a further 6 days at  $37^\circ\text{C}$  in 5%  $\text{CO}_2$ . The cells were washed in PBS and then seeded at  $1 \times 10^5$  cells/ml in phase two media: Stemspan medium,  $2 \text{ U/ml}$  EPO (Espex),  $5 \text{ ng/ml}$  IL-3,  $2 \times 10^{-6} \text{ M}$  dexamethasone (Sigma),  $10^{-6} \text{ M}$   $\beta$ -estradiol (Sigma) and  $20 \text{ ng/ml}$  stem cell factor. Cells were collected at different time-points of phase two growth to monitor erythroid differentiation and extract RNA using Tri-reagent (Sigma). Differentiation was assessed using qRT-PCR (with Taqman probes for HBB and HBG2, and GAPDH to normalise) and May-Grunwald morphological staining (data not shown).

### **Analysis of RNA extracted from phase two erythroid culture**

Microarrays were designed by Nimblegen Roche to cover the  $\beta$ -globin locus at high resolution.  $100 \text{ ng}$  of RNA was treated with DNase (Sigma) before being converted to cDNA using the Transplex complete whole transcriptome amplification kit (Sigma) as per the manufacturer's guidelines. The cDNA was RNase (Sigma) treated and cDNA precipitated and labelled using the Nimblegen Expression Array kit using the manufacturer's protocol. The labelled cDNA was

hybridised to a single chamber in a Nimblegen 12plex microarray slide following the manufacturer's guidelines. The slide was washed and scanned at 3  $\mu\text{m}$  resolution using an InnoScan 700 Microarray Scanner and TIFF images were generated using Mapix ver 5.1 software. The TIFF images were aligned and converted into probe intensity values using the Nimblegen DEVA software with background correction and uploaded onto UCSC genome browser.

To confirm the peak of transcription over the  $\delta\beta$ -promoter region in human erythroid cells quantitative RT-PCR was performed; cDNA was first generated from 150 ng of DNase-treated (Sigma) RNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems). Quantitative PCR was carried out on the BioRad CFX96 QPCR machine using IQ SYBR green supermix (BioRad), 0.5  $\mu\text{M}$  concentration of each primer and 0.9  $\mu\text{l}$  of cDNA. No-RT controls and no-template controls verified the lack of genomic or PCR contamination respectively (not shown). Standard curves for the U-amplicon and D-amplicon were generated by amplifying genomic DNA at a range of concentrations and used to quantify the relative levels of transcription in the region. The ratio of downstream over upstream signal was used to estimate the changing transcriptional pattern during erythroid differentiation.

Primer sequences:

Downstream-F (forward primer): 5'-ATGCCAATGTGGGTTAGAATG-3'

Downstream-R (reverse primer): 5'-CATACCATGTGGCTCATCCTC-3'

Upstream-F (forward primer): 5'-TTAAATGCCAGTGCTCTCCAC-3'

Upstream-R (reverse primer): 5'-CAGAAGGAACTAGGATGTGTCC-3'

For strand-specific RT-PCR of the D and U amplicons cDNA was generated from 150 ng of DNase-treated RNA using the GoScript™ Reverse Transcription System (Promega) with either the Downstream-F or Downstream-R primer. PCR was then performed using Downstream-F and

Downstream-R primers. To assess the RNA integrity an additional cDNA reaction was performed by using random-primers in the RT step.

### **YAC transgenic mice**

264W transgenic mice carrying a single copy of a 150 kb human  $\beta$ -globin locus YAC were previously described [32]. A 213 kb YAC containing the human  $\beta$ -globin locus [28] was used to generate the rest of the transgenic lines. The YAC was modified using the pop-in-pop-out recombination procedure [55] to introduce *loxP* sites upstream and downstream of the minimal  $\delta\beta$  intergenic promoter. Briefly, a gene targeting vector was generated by inserting *loxP* sites at the *BseRI* and *ClaI* restriction sites of the 2.5 kb *BglIII* restriction fragment (see Figure 3). The modified YACs were then used to generate transgenic mice as previously described [26]. The structural integrity of the  $\beta$ -globin loci in the obtained transgenic animals was analysed as described in [27]. Briefly, *SfiI*-digested DNA was fractionated by pulsed-field gel electrophoresis and Southern blots were prepared. The membrane was cut into strips comprising individual gel lanes. Each strip was hybridized with a single probe. Probes utilised include: probe 1, 0.7 kb *PstI* fragment upstream of HS3; probe 2, 1.9 kb *HindIII* fragment upstream of HS2; probe 3, 3.7 kb *EcoRI* fragment spanning the *HBE* gene; probe 4, 2.4 *EcoRI* fragment downstream of the *HBG1* gene; probe 5, 1.9 kb *XbaI* fragment downstream of the beta-like pseudogene; probe 6, 2.1 kb *PstI* fragment upstream of the *HBD* gene; probe 7, 0.9 *EcoRI-BamHI* *HBB* fragment; probe 8, 1.4 kb *XbaI* DF10 3'HS1 fragment; probe 9, 1.9 kb *BglIII* HPFH-3 fragment; probe 10, 0.5-kb *HindIII* H500 fragment; probe 11, 1.5-kb *EcoRI-BglIII* HPFH-6 fragment (all described in [27]). Transgene copy numbers were estimated by quantitative PhosphorImager analysis of Southern blots. To generate a line with a deletion of the  $\delta\beta$  intergenic promoter, mice were bred to Deleter Cre recombinase-expressing mice [31]. After several generations, animals carrying a single, complete copy of the human  $\beta$ -globin locus with a

deletion of the  $\delta\beta$  promoter were identified. The integrity of the transgene after recombination was confirmed by pulsed-field gel electrophoresis of *Sfi*I-digested DNA and subsequent Southern blotting and hybridisation with probe D, 2.5 kb *Bgl*II fragment spanning the  $\delta\beta$  promoter. Deletion of the  $\delta\beta$  promoter and copy-number reduction were verified by Southern hybridization of *Bgl*II-digested genomic DNA with probes D and *Thy1*, followed by PhosphorImager quantification.

### **Reporter constructs, transfection of cultured cells, and FACS analysis**

50 ng of mouse genomic DNA from lines 264W, FX115 and  $\Delta$ 115 were amplified by PCR with primers 5'-CTTGAGTGCATTGACAAAATTACC-3' and 5'-CAGACTTGGACCATGACGGTG-3'. The PCR products were cloned into pGEM-T easy (Promega) and the sequence of the resulting plasmids was verified. The  $\delta\beta$  promoter fragments were released by *Eco*RI/*Cla*I (264W) or *Eco*RI/*Sal*I (FX115 and  $\Delta$ 115) digestions, blunted with T4 DNA polymerase (New England Biolabs) and ligated into vector pEGFP-C1 (promoterless EGFP, Clontech) at the *Sma*I site. Plasmids were verified by sequencing.

Cells were grown in RPMI 1640 supplemented with 10% foetal calf serum and Penicillin/Streptomycin.  $5 \times 10^7$  cells were transfected by electroporation with 50  $\mu$ g of linearized plasmid DNA at 350 V, 1000  $\mu$ F in 500  $\mu$ l of 1x HEPES-Buffered Saline. Plasmid pEGFP-N1 (Clontech), expressing EGFP under the control of a CMV promoter, was also used as a positive control. After transfection cells were transferred to non-selective medium for 24 hours followed by 14 days in medium with 600  $\mu$ g/ml G418. EGFP expression was analysed by FACS performed on a FACSCalibur. Background was determined by analysing non-transfected cells and percentage of green cells as well as mean fluorescence intensity was determined using the CellQuest software.

## RNA FISH

RNA FISH was performed as previously described [9]. Cocktails of hapten-labelled oligonucleotides (Oswel) were used as probes, sequences provided below:

Dinitrophenol-labelled human *HBE* intron 2 probes:

5'-AATCTTGAGGACTTTCCCAATCAACTTGCT-3'

5'-GTTCTAACATCAAGCTCGACCCATGATTTC-3'

5'-CCTTCACTTGGTCACAATACTGATGGGAAA-3'

5'-CTGACCTCAAAGTGTCCAAGGTTTGTGCC-3'

Digoxigenin-labelled human *HBG* intron 2 probes:

5'-CGACCTGGACTTTTGCCAGGCACAGGGTCC-3'

5'-TCACTCCCAACCCCAGTATCTTCAAACAGC-3'

5'-GCATCTTTTAAACGACCATACTTGTCTGC-3'

5'-ACAGAGCTGACTTTCAAATCTACTCCAGC-3'

Dinitrophenol-labelled human *HBB* intron 2 probes:

5'-TTCCACACTGATGCAATCATTCGTCTGTTT-3'

5'-TGTGTACACATATTAACATTACTTTA-3'

5'-ATTAGCAATATGAAACCTCTTACATCAGTT-3'

5'-AGTAATGTACTAGGCAGACTGTGTAAAGTT-3'

Dinitrophenol-labelled mouse *Hbb-b1* intron 2 probes:

5'-TCAACTCCACACACAGTCATGGAGACTG-3'

5'-GATATCAGGATGGGAAGTAAATAACCAGCT-3'

5'-AAACAAACGTAGATCAAAGAAGAAGAAATG-3'

5'-AGCTATGAGAAGAAACAGGGACATATCTTC-3'

Digoxigenin-labelled human *HBB* exon 3 probes:

5'-TTTAATAGAAATTGGACAGCAAGAAAGCG-3'

5'-AGGCCCTTCATAATATCCCCCAGTTTAGTA-3'

Dinitrophenol-labelled probes were detected with Texas Red-labelled antibodies (primary antibody from Serotec, secondary antibodies from Jackson ImmunoResearch); Digoxigenin-labelled probes were detected with FITC-labelled antibodies (primary antibody from Roche, secondary antibodies from Jackson ImmunoResearch). Slides were examined on an Olympus BX41 epifluorescence microscope. To determine the percentage of cells with human transgene signals, a minimum of 300 loci were counted for each data point.

### **RT-PCR**

Total RNA was isolated from adult anaemic spleen 5 days after phenylhydrazine injections [56] using the RNeasy mini kit (Qiagen) according to the manufacturer's instructions. RNA samples were DNase I-treated with 1U of RQ1 DNase (Promega) per  $\mu\text{g}$  of RNA and purified through a second RNeasy column. 1 to 5 $\mu\text{g}$  of total RNA was reverse-transcribed with Superscript III reverse transcriptase (Invitrogen) following manufacturer's protocol at 50°C in the presence of 2 U/ $\mu\text{l}$  RNasin (Promega). RT negative controls in which the reverse transcriptase enzyme was omitted were set up in parallel. Quantitative RT-PCR was performed by real-time PCR in an ABI PRISM 7000 Sequence Detection System using SYBR green PCR Master Mix (Applied Biosystems). All PCR reactions were carried out in duplicates and carbonic anhydrase II was used as a normalisation control. Serial dilutions of genomic DNA (primary and intergenic transcripts) or of a plasmid containing the carbonic anhydrase II cDNA PCR product were used as standard curves for absolute quantification.

The following primers were used for PCR amplification of cDNA:

*HBD* gene, intron 2 forward primer: 5'-TGGGGATCAGTTTTGTCTAAGATTTGGG-3'

*HBD* gene, intron 2 reverse primer: 5'-GCGGAGAAGAGGTAGGCAGATACATGC-3'

*HBB* gene, intron 2 forward primer: 5'-GACGAATGATTGCATCAGTGTGGAAGTC-3'

*HBB* gene, intron 2 reverse primer: 5'-TGCGGAGAAGAAAAAAAAAAGAAAGCAAG-3'

intergenic RT3B forward primer: 5'-TCATCTGGGATTTTGGGGACTATGTC-3'

intergenic RT3B reverse primer: 5'-CAGACTTGGGAATTGGGATTATACAGGC-3'

intergenic RT4 forward primer: 5'-CTTGTGTTACGACTGACATCACCG-3'

intergenic RT4 reverse primer: 5'-GGAGACAAATAGCTGGGCTTCTGTTG-3'

mouse carbonic anhydrase II forward primer: 5'-GAGTTTGATGACTCTCAGGACAATGCAG-3'

mouse carbonic anhydrase II reverse primer: 5'-AGATGAGCCCCAGTGAAAGTGAAAC-3'

### **Chromatin immunoprecipitation**

Histone modification profiles across the human  $\beta$ -globin locus were assessed by native chromatin immunoprecipitation [57] and performed exactly as described by [10]. Briefly, single-cell suspensions of erythroid cells were resuspended to  $2 \times 10^7$  cells/ml in ice cold RSB buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl<sub>2</sub>), 0.1% Triton X-100, 0.5 mM DTT, 0.1 M sucrose, 0.1 mM PMSF, 5 mM Na butyrate, supplemented with protease inhibitor cocktail (Sigma). The cells were homogenised in a cold B-type Dounce homogeniser and diluted with an equal volume of the same buffer with 0.25 M sucrose. The suspension was layered onto a sucrose cushion of 0.33 M sucrose, 5 mM MgCl<sub>2</sub>, 10 mM Tris pH 8.0, 0.5 mM DTT, 0.1 mM PMSF, 5 mM Na butyrate. This was then centrifuged at  $800 \times g$  for 5 min at 4°C to pellet nuclei. Chromatin was digested with micrococcal nuclease generating DNA predominantly mononucleosomal in length. CHIP was carried out using the following rabbit polyclonal antibodies: anti-trimethyl-histone H3



(K4) (Abcam), anti-dimethyl-histone H3 (K4) (Upstate Biotechnology), anti-acetyl-histone H3 (K9/K14) (Upstate Biotechnology).

The representation of multiple sequences across the human  $\beta$ -globin locus in the Input and Bound fractions was determined by quantitative real-time PCR. Real-time PCR was performed in an ABI PRISM 7000 Sequence Detection System using SYBR green PCR Master Mix (Applied Biosystems). All PCR reactions were carried out in duplicates using 3 ng Input or Bound DNA as template. For primer sequences, see Miles et al. [10]. The ratio of Bound to Input DNA was calculated using the comparative  $C_T$  method,  $\text{Bound/Input} = 2^{(\text{Input } C_T - \text{Bound } C_T)}$  [58]. Data were normalised to the most 3' data point, downstream of 3'HS1, which shows low enrichment for all samples.

### **Acknowledgments**

We thank L. Mercer, M. Anderton, A. Nanou, and D. Dimitrova for their assistance. This work was supported in part by the Medical Research Council and the Biotechnology and Biological Sciences Research Council, UK, by USPHS NIH grants DK053510, HL067336, and HL111264 awarded to KRP and by a grant from Sparks awarded to DRFC.

## Figure Legends

**Figure 1.** Schematic map of the human  $\beta$ -globin locus and characterization of the  $\delta\beta$  intergenic promoter. **(A)** Schematic diagram of the human  $\beta$ -globin locus and the  $\delta\beta$  intergenic promoter region. Black boxes represent globin genes; grey box corresponds to the  $\beta$ -like pseudogene, open boxes – to olfactory receptor genes flanking the  $\beta$ -globin locus. Vertical arrows indicate DNase hypersensitive sites, including sites 1 to 5 of the locus control region (LCR); horizontal arrows denote intergenic transcription start sites. The detailed map of the  $\delta\beta$  promoter region shows the positions of the intergenic transcription start site and upstream Alu element (left horizontal open arrow). Restriction fragments from the  $\delta\beta$  promoter region were cloned upstream of a promoterless EGFP gene as indicated and described in Materials and Methods. Relevant restriction sites are indicated. **(B)** EGFP reporter assays. Constructs were stably transfected into K562 cells and the percentage of EGFP-expressing cells was determined by flow cytometry. The Apo construct, which showed the highest percentage of EGFP-positive cells, was assigned an arbitrary value of 100.

**Figure 2.** Analysis of transcription in the  $\delta\beta$  intergenic promoter region in human erythroid cells. **(A)** Schematic diagram of the human  $\beta$ -globin locus and the  $\delta\beta$  intergenic promoter region. Black boxes represent globin genes; grey box corresponds to the  $\beta$ -like pseudogene, open boxes correspond to olfactory receptor genes flanking the  $\beta$ -globin locus. Vertical arrows indicate DNase hypersensitive sites, including sites 1 to 5 of the locus control region (LCR); horizontal arrows denote putative intergenic transcription start sites. The detailed map of the  $\delta\beta$  promoter region shows the positions of the intergenic transcription start site and upstream Alu element (left horizontal open arrow). Restriction enzyme sites are indicated by vertical lines. The graph below the  $\delta\beta$  intergenic promoter schematic indicates microarray signal intensity for various probes (indicated under the graph with horizontal black boxes) following labelling and hybridisation of

RNA extracted from primary erythroid cells at different phases (day 4 in light grey, day 5 in dark grey and day 7 in black) of phase II culture. The results shown are representative of five biological replicates (B) The RNA samples used for microarray analysis in A were converted into cDNA using a WTA kit and amplified by PCR. The PCR amplicons are indicated beneath the microarray signal graph as U (upstream) amplicon and D (downstream) amplicon. The amounts of each amplicon were calculated by comparing signals to a standard curve generated with genomic DNA and the ratio of downstream over upstream was calculated. The ratio increases through phase II indicating a relative increase in transcripts downstream of the putative intergenic promoter. Results are representative of two biological replicates. Pearson correlation analysis indicates a statistically significant correlation ( $p < 0.01$ ). (C) RNA from day 7 of erythroid differentiation was converted into cDNA using either gene-specific primer (the downstream-amplicon forward primer [Downstream-F], the downstream-amplicon reverse primer [Downstream-R] or a combination of the two primers [Downstream-F+R] as indicated to the left of each panel). PCR was then carried out using both Downstream-F and Downstream-R primers (to amplify the downstream amplicon). GAPDH primers were used as a control to confirm RNA integrity. PCR products were run on a 2% gel, stained with SYBR Gold, and visualised on a Typhoon Imager. The strand-specific results show that the transcripts detected downstream of the  $\delta\beta$  intergenic promoter are sense transcripts (travelling towards HBD).

**Figure 3.** Generation of YAC transgenic mice carrying a floxed  $\delta\beta$  intergenic promoter and production of the deletion lines. (A) Schematic diagram of the human  $\beta$ -globin locus YAC used to generate transgenic lines FX14, FX101, and FX115. Black boxes represent globin genes; the grey box corresponds to the  $\beta$ -like pseudogene; open boxes indicate olfactory receptor genes. Vertical arrows denote DNase hypersensitive sites. Horizontal arrow refers to the  $\delta\beta$  intergenic transcription

start site. The  $\delta\beta$  promoter region is shown at a larger scale below the main map. Black triangles denote *loxP* sites inserted upstream and downstream of the  $\delta\beta$  promoter. White horizontal arrows indicate Alu repeat elements flanking the  $\delta\beta$  promoter. *SfiI* and selected *BglIII* restriction sites are indicated, and probes are shown as grey boxes below the lines (probes 1 to 4 and probe D described in Materials and Methods). **(B)** YAC transgenic lines carrying the human  $\beta$ -globin locus, before Cre-mediated deletion of the intergenic promoter. Genomic DNA from three transgenic lines, FX14, FX101, and FX115, was digested with *SfiI*, subjected to pulsed-field electrophoresis and Southern transfer. Several identical DNA samples from each line were run on the same gel, blots were cut into strips and each strip was hybridized with a different probe as indicated. Probes throughout the  $\beta$ -globin locus identify identical *SfiI* fragments. **(C)** Transgenic mice with deleted  $\delta\beta$  promoter. Genomic DNA from line FX115 and 5 animals carrying the deletion,  $\Delta 1$  to  $\Delta 5$ , was digested with *SfiI*, subjected to pulsed-field electrophoresis, transferred and hybridized with probe D. A single band corresponding to the  $\beta$ -globin *SfiI* fragment is detected in  $\Delta 1$  to  $\Delta 5$ . **(D)** Efficient deletion of the  $\delta\beta$  intergenic promoter. Southern blot analysis of *BglIII*-digested genomic DNA from parent line FX115 and  $\Delta 1$  to  $\Delta 5$  hybridized to probe D. Note size reduction of *BglIII* band after deletion of the  $\delta\beta$ -promoter.

**Figure 4.** Human  $\beta$ -globin transcription in transgenic mice before and after deletion of the  $\delta\beta$  intergenic promoter. **(A)** RNA FISH analysis of human *HBE* and *HBG* transcription in E10.5d blood cells from heterozygous transgenic embryos. *HBE* primary transcripts are in red, *HBG* transcripts in green; for scale, see (C), scale bar, 5  $\mu\text{m}$ . **(B)** Normal transcription of *HBE* and *HBG* genes in E10.5d embryonic blood cells by RNA FISH. Cells positive for human globin gene transcription signals were scored; in each line nearly 100% of the embryonic blood cells have transcription signals for *HBE*, *HBG* or both genes indicating normal high-level expression. **(C)**

RNA FISH analysis of *HBB* transcription in adult erythroid cells. *HBB* primary transcripts are in red, mouse *Hbb-b1* transcripts detected in green as an internal control to identify erythroid precursors; scale bar, 5  $\mu$ m. **(D)** Reduced transcription of *HBB* in adult erythroid cells carrying a  $\delta\beta$  promoter deletion. Heterozygous mice from wild-type lines FX14 and 264W, lines FX101 and FX115 with ‘floxed’  $\delta\beta$  promoter, and  $\Delta$ 115 with deleted  $\delta\beta$  promoter were analyzed by RNA FISH. Nuclei with a *HBB* signal were scored and expressed as a percentage of *Hbb-b1* positive nuclei. *HBB* transcription is impaired in lines FX101 and FX115, and further reduced in line  $\Delta$ 115. **(E)** Reduced transcription of human *HBD* and *HBB* genes in adult erythroid cells carrying a  $\delta\beta$  promoter deletion. Animals from wild-type lines FX14 and 264W, and line  $\Delta$ 115 were analyzed by real-time RT-PCR, as described in Materials and Methods. Primary transcript levels are corrected for transgene copy number; wild-type *HBB* level in 264W is assigned an arbitrary value of 100.

**Figure 5.** Intergenic transcription in the  $\beta$ -globin locus before and after deletion of the  $\delta\beta$  intergenic promoter. **(A)** RT-PCR to detect intergenic transcripts downstream of the  $\delta\beta$  intergenic promoter in adult erythroid cells. The positions of PCR amplicons RT3B and RT4 are indicated below the map; black boxes, *HBD* and *HBB* genes; horizontal arrow, normal position of the  $\delta\beta$  intergenic promoter and transcript start site. Intergenic transcription is reduced in lines FX115 and  $\Delta$ 115 with ‘floxed’ and deleted  $\delta\beta$  intergenic promoter, respectively. **(B)** Transfection assays to assess intergenic promoter activity. Fragments containing the  $\delta\beta$  intergenic promoter or corresponding fragments from the deletion lines were PCR amplified from lines 264W, FX115, and  $\Delta$ 115, and cloned upstream of a promoterless EGFP gene as described in Materials and Methods. Reporter constructs were stably transfected into MEL cells and the number of EGFP-expressing cells in the population was determined by flow cytometry. The percentage of cells

expressing EGFP is severely decreased compared to wild-type  $\delta\beta$  promoter (assigned an arbitrary value of 100) and CMV promoter when driven by FX115- and  $\Delta 115$ -derived sequences.

**Figure 6.** *HBB* expression is variegated in transgenic mice carrying a deletion of the  $\delta\beta$  intergenic promoter. RNA FISH analysis of *HBB* transcripts in adult erythroid cells; primary and mature *HBB* transcripts are detected in green by an exon probe; as an internal control, *Hbb-b1* primary transcripts are detected in red to identify erythroid cells; scale bar, 5  $\mu\text{m}$ . Greater than 95% of FX14 erythroid cells have nuclear primary *HBB* signals and cytoplasmic *HBB* mRNA staining (top row). Conversely, the majority of FX115 and  $\Delta 115$  cells lack primary *HBB* signals and do not contain cytoplasmic *HBB* mRNA (cells on the left). The few FX115 and  $\Delta 115$  cells that do express *HBB* mRNA in the cytoplasm also have nuclear *HBB* primary transcript signals and vice-versa (cells on the right) showing that *HBB* expression is restricted to a small subset of adult erythroid cells.

**Figure 7.** Histone modifications across the human  $\beta$ -globin locus in wt and  $\delta\beta$  promoter deleted adult erythroid cells. Chromatin immunoprecipitation assays for histone H3 acetylation (H3ac), dimethylated lysine 4 of histone H3 (H3K4me2), and tri-methylated lysine 4 of histone H3 (H3K4me3); ●, wild-type adult erythroid cells from line 264W [10]; ■, adult erythroid cells from line  $\Delta 115$ . Coordinates in bp are shown along the x axis with *HBE* start site as +1. Bound versus input ratios were calculated and normalized to the most 3' primer pair, located just downstream of 3'HS1 in the ORG cluster. A map of the locus is shown below the graphs.

## References

1. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. *Science*. 2005;309(5740):1559-63. Epub 2005/09/06. doi: 309/5740/1559 [pii] 10.1126/science.1112014. PubMed PMID: 16141072.
2. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, et al. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*. 2005;308(5725):1149-54. PubMed PMID: 15790807.
3. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, et al. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*. 2004;116(4):499-509. PubMed PMID: ISI:000189184200005.
4. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 25(18):1915-27. Epub 2011/09/06. doi: gad.17446611 [pii] 10.1101/gad.17446611. PubMed PMID: 21890647.
5. Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, et al. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res*. 2005;15(7):987-97. PubMed PMID: 15998911.
6. Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*. 1987;51(6):975-85.
7. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. Long-range chromatin regulatory interactions in vivo. *Nat Genet*. 2002;32(4):623-6. PubMed PMID: 12426570.
8. 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. PubMed PMID: 12504019.
9. Gribnau J, Diderich K, Pruzina S, Calzolari R, Fraser P. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell*. 2000;5(2):377-86.
10. Miles J, Mitchell JA, Chakalova L, Goyenechea B, Osborne CS, O'Neill L, et al. Intergenic transcription, cell-cycle and the developmentally regulated epigenetic profile of the human beta-globin locus. *PLoS ONE*. 2007;2(7):e630. PubMed PMID: 17637845.
11. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 2005;120(2):169-81. PubMed PMID: 15680324.
12. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007;129(7):1311-23. PubMed PMID: 17604720.
13. Tuan D, Kong S, Hu K. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A*. 1992;89(23):11219-23.
14. Long QM, Bengra C, Li CH, Kutlar F, Zeng S, Bollag R, et al. An LTR-retrotransposon of the-human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human beta-globin locus control region. *Blood*. 1998;92(10):1367. PubMed PMID: ISI:000077121301359.
15. Leach KM, Nightingale K, Igarashi K, Levings PP, Engel JD, Becker PB, et al. Reconstitution of human beta-globin locus control region hypersensitive sites in the absence of chromatin assembly. *Mol Cell Biol*. 2001;21(8):2629-40.

16. Routledge SJ, Proudfoot NJ. Definition of transcriptional promoters in the human beta globin locus control region. *J Mol Biol.* 2002;323(4):601-11. PubMed PMID: 12419253.
17. Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ. Intergenic transcription and transinduction of the human beta-globin locus. *Genes & Development.* 1997;11(19):2494-509.
18. Calzolari R, McMorrow T, Yannoutsos N, Langeveld A, Grosveld F. Deletion of a region that is a candidate for the difference between the deletion forms of hereditary persistence of fetal hemoglobin and deltabeta-thalassemia affects beta- but not gamma-globin gene expression. *Embo J.* 1999;18(4):949-58.
19. Chakalova L, Osborne CS, Dai YF, Goyenechea B, Metaxotou-Mavromati A, Kattamis A, et al. The Corfu deltabeta thalassemia deletion disrupts gamma-globin gene silencing and reveals post-transcriptional regulation of HbF expression. *Blood.* 2005;105(5):2154-60. Epub 2004/11/13. doi: 2003-11-4069 [pii] 10.1182/blood-2003-11-4069. PubMed PMID: 15536151.
20. Nagano T, Fraser P. No-nonsense functions for long noncoding RNAs. *Cell.* 2011;145(2):178-81. Epub 2011/04/19. doi: S0092-8674(11)00291-1 [pii] 10.1016/j.cell.2011.03.014. PubMed PMID: 21496640.
21. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science.* 2008;322(5908):1717-20. Epub 2008/11/08. doi: 1163802 [pii] 10.1126/science.1163802. PubMed PMID: 18988810.
22. Redrup L, Branco MR, Perdeaux ER, Krueger C, Lewis A, Santos F, et al. The long noncoding RNA *Kcnq1ot1* organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development.* 2009;136(4):525-30. Epub 2009/01/16. doi: dev.031328 [pii] 10.1242/dev.031328. PubMed PMID: 19144718.
23. Koziol MJ, Rinn JL. RNA traffic control of chromatin complexes. *Curr Opin Genet Dev.* 2010;20(2):142-8. Epub 2010/04/07. doi: S0959-437X(10)00037-7 [pii] 10.1016/j.gde.2010.03.003. PubMed PMID: 20362426.
24. Xu J, Shao Z, Glass K, Bauer DE, Pinello L, Van Handel B, et al. Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev Cell.* 2012;23(4):796-811. Epub 2012/10/09. doi: 10.1016/j.devcel.2012.09.003. PubMed PMID: 23041383; PubMed Central PMCID: PMC3477283.
25. Peterson KR, Clegg CH, Huxley C, Josephson BM, Haugen HS, Furukawa T, et al. Transgenic mice containing a 248-kb yeast artificial chromosome carrying the human beta-globin locus display proper developmental control of human globin genes. *Proceedings of the National Academy of Sciences of the United States of America.* 1993;90(16):7593-7.
26. Peterson KR, Clegg CH, Li Q, Stamatoyannopoulos G. Production of transgenic mice with yeast artificial chromosomes. *Trends Genet.* 1997;13(2):61-6. PubMed PMID: 9055607.
27. Peterson KR, Li QL, Clegg CH, Furukawa T, Navas PA, Norton EJ, et al. Use of yeast artificial chromosomes (YACs) in studies of mammalian development: production of beta-globin locus YAC mice carrying human globin developmental mutants. *Proc Natl Acad Sci U S A.* 1995;92(12):5655-9. PubMed PMID: 7539923.
28. Peterson KR, Navas PA, Li Q, Stamatoyannopoulos G. LCR-dependent gene expression in beta-globin YAC transgenics: detailed structural studies validate functional analysis even in the presence of fragmented YACs. *Hum Mol Genet.* 1998;7(13):2079-88. PubMed PMID: 9817925.
29. Peterson KR, Navas PA, Stamatoyannopoulos G. beta-YAC transgenic mice for studying LCR function. *Ann N Y Acad Sci.* 1998;850:28-37. PubMed PMID: 9668524.



30. Peterson KR, Zitnik G, Huxley C, Lowrey CH, Gnirke A, Leppig KA, et al. Use of yeast artificial chromosomes (YACs) for studying control of gene expression: correct regulation of the genes of a human beta-globin locus YAC following transfer to mouse erythroleukemia cell lines [published erratum appears in Proc Natl Acad Sci U S A 1994 Mar 1;91(5):1979]. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(23):11207-11.
31. Schwenk F, Baron U, Rajewsky K. A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. Nucleic acids research. 1995;23(24):5080-1. PubMed PMID: 8559668.
32. Tanimoto K, Liu Q, Bungert J, Engel JD. Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. Nature. 1999;398(6725):344-8.
33. Milot E, Strouboulis J, Trimborn T, Wijgerde M, de Boer E, Langeveld A, et al. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. Cell. 1996;87(1):105-14.
34. Wijgerde M, Grosveld F, Fraser P. Transcription complex stability and chromatin dynamics in vivo. Nature. 1995;377(6546):209-13.
35. Trimborn T, Gribnau J, Grosveld F, Fraser P. Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci. Genes Dev. 1999;13(1):112-24.
36. Farrell CM, West AG, Felsenfeld G. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. Mol Cell Biol. 2002;22(11):3820-31. PubMed PMID: 11997516.
37. Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet. 2006;15 Spec No 1:R17-29. PubMed PMID: 16651366.
38. Kim A, Kiefer CM, Dean A. Distinctive signatures of histone methylation in transcribed coding and noncoding human beta-globin sequences. Mol Cell Biol. 2007;27(4):1271-9. PubMed PMID: 17158930.
39. 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. PubMed PMID: 16094312.
40. McKittrick E, Gafken PR, Ahmad K, Henikoff S. Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci U S A. 2004;101(6):1525-30. PubMed PMID: 14732680.
41. Sims RJ, 3rd, Belotserkovskaya R, Reinberg D. Elongation by RNA polymerase II: the short and long of it. Genes Dev. 2004;18(20):2437-68. PubMed PMID: 15489290.
42. Schmitt S, Prestel M, Paro R. Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. 2005;19(6):697-708. PubMed PMID: 15741315.
43. Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, et al. Long noncoding RNAs with enhancer-like function in human cells. Cell. 2010;143(1):46-58. Epub 2010/10/05. doi: [S0092-8674\(10\)01011-1 \[pii\] 10.1016/j.cell.2010.09.001](https://doi.org/10.1016/j.cell.2010.09.001). PubMed PMID: 20887892.
44. Akhtar A, Becker PB. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol Cell. 2000;5(2):367-75. PubMed PMID: ISI:000085645100016.
45. Bender MA, Bulger M, Close J, Groudine M. Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. Mol Cell. 2000;5(2):387-93.
46. Schubeler D, Francastel C, Cimborá DM, Reik A, Martin DI, Groudine M. Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. Genes Dev. 2000;14(8):940-50.

47. Xiang P, Fang X, Yin W, Barkess G, Li Q. Non-coding transcripts far upstream of the epsilon-globin gene are distinctly expressed in human primary tissues and erythroleukemia cell lines. *Biochemical and biophysical research communications*. 2006;344(2):623-30. PubMed PMID: 16620781.
48. 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. PubMed PMID: 15361872.
49. Ragozy T, Bender MA, Telling A, Byron R, Groudine M. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev*. 2006;20(11):1447-57. PubMed PMID: 16705039.
50. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*. 2006;16(10):1299-309. Epub 2006/09/07. doi: gr.5571506 [pii] 10.1101/gr.5571506. PubMed PMID: 16954542.
51. Huang P, Keller CA, Giardine B, Grevet JD, Davies JOJ, Hughes JR, et al. Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes Dev*. 2017;31(16):1704-13. Epub 2017/09/17. doi: 10.1101/gad.303461.117. PubMed PMID: 28916711; PubMed Central PMCID: PMC65647940.
52. Ivaldi MS, Diaz LF, Chakalova L, Lee J, Krivega I, Dean A. Fetal gamma-globin genes are regulated by the BGLT3 long noncoding RNA locus. *Blood*. 2018;132(18):1963-73. Epub 2018/08/29. doi: 10.1182/blood-2018-07-862003. PubMed PMID: 30150205; PubMed Central PMCID: PMC6213316.
53. Sankaran VG, Xu J, Byron R, Greisman HA, Fisher C, Weatherall DJ, et al. A functional element necessary for fetal hemoglobin silencing. *N Engl J Med*. 2012;365(9):807-14. Epub 2011/09/02. doi: [10.1056/NEJMoa1103070](https://doi.org/10.1056/NEJMoa1103070). PubMed PMID: 21879898.
54. Jawaid K, Wahlberg K, Thein SL, Best S. Binding patterns of BCL11A in the globin and GATA1 loci and characterization of the BCL11A fetal hemoglobin locus. *Blood Cells Mol Dis*. 2010;45(2):140-6. Epub 2010/06/15. doi: 10.1016/j.bcmd.2010.05.006. PubMed PMID: 20542454.
55. Rothstein R. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods in enzymology*. 1991;194:281-301. PubMed PMID: 2005793.
56. Dickerman HW, Cheng TC, Kazazian HH, Jr., Spivak JL. The erythropoietic mouse spleen—a model system of development. *Arch Biochem Biophys*. 1976;177(1):1-9. PubMed PMID: 999280.
57. O'Neill LP, Turner BM. Immunoprecipitation of native chromatin: NChIP. *Methods*. 2003;31(1):76-82. PubMed PMID: 12893176.
58. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods*. 2001;25(4):402-8. PubMed PMID: ISI:000173949500003.

Figure 5  
Debrand *et al.*

bioRxiv preprint doi: <https://doi.org/10.1101/639807>; this version posted May 15, 2019. 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.

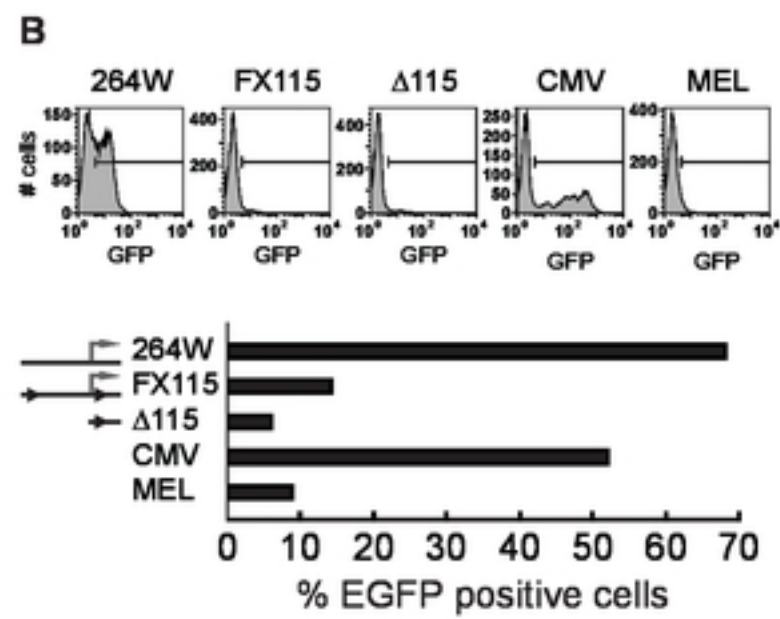
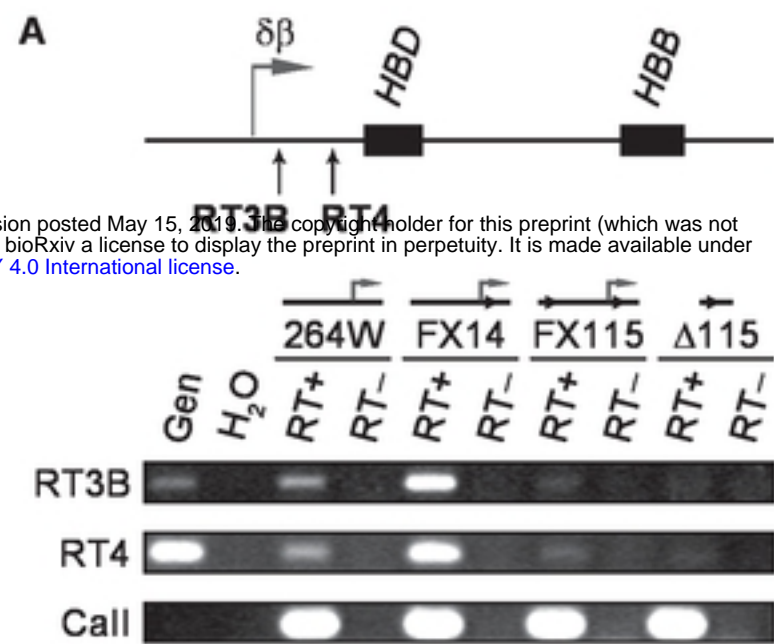


Figure 6  
Debrand *et al.*

bioRxiv preprint doi: <https://doi.org/10.1101/639807>; this version posted May 15, 2019. 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.

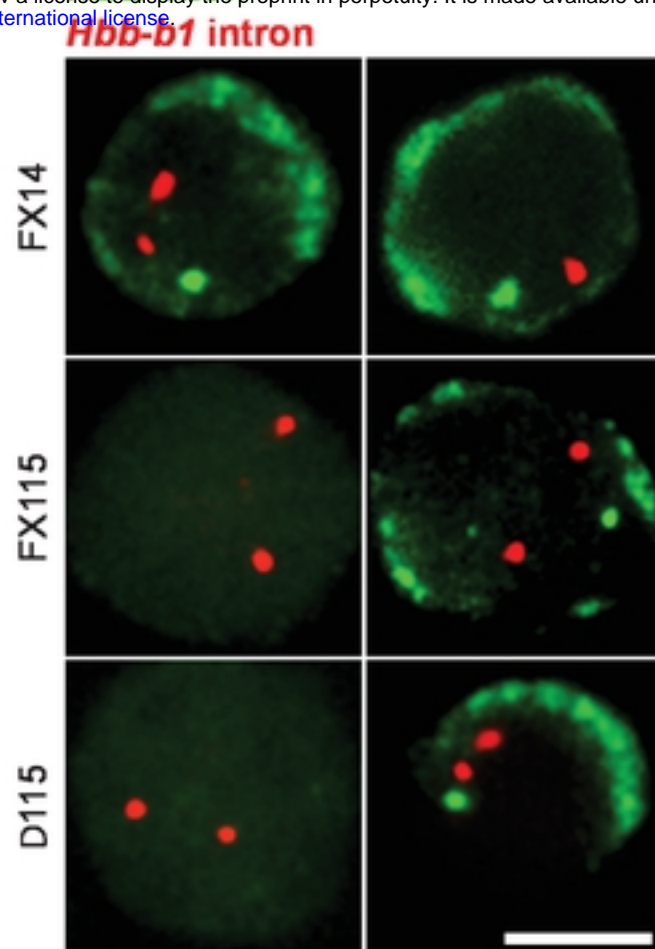


Figure 7  
Debrand *et al.*

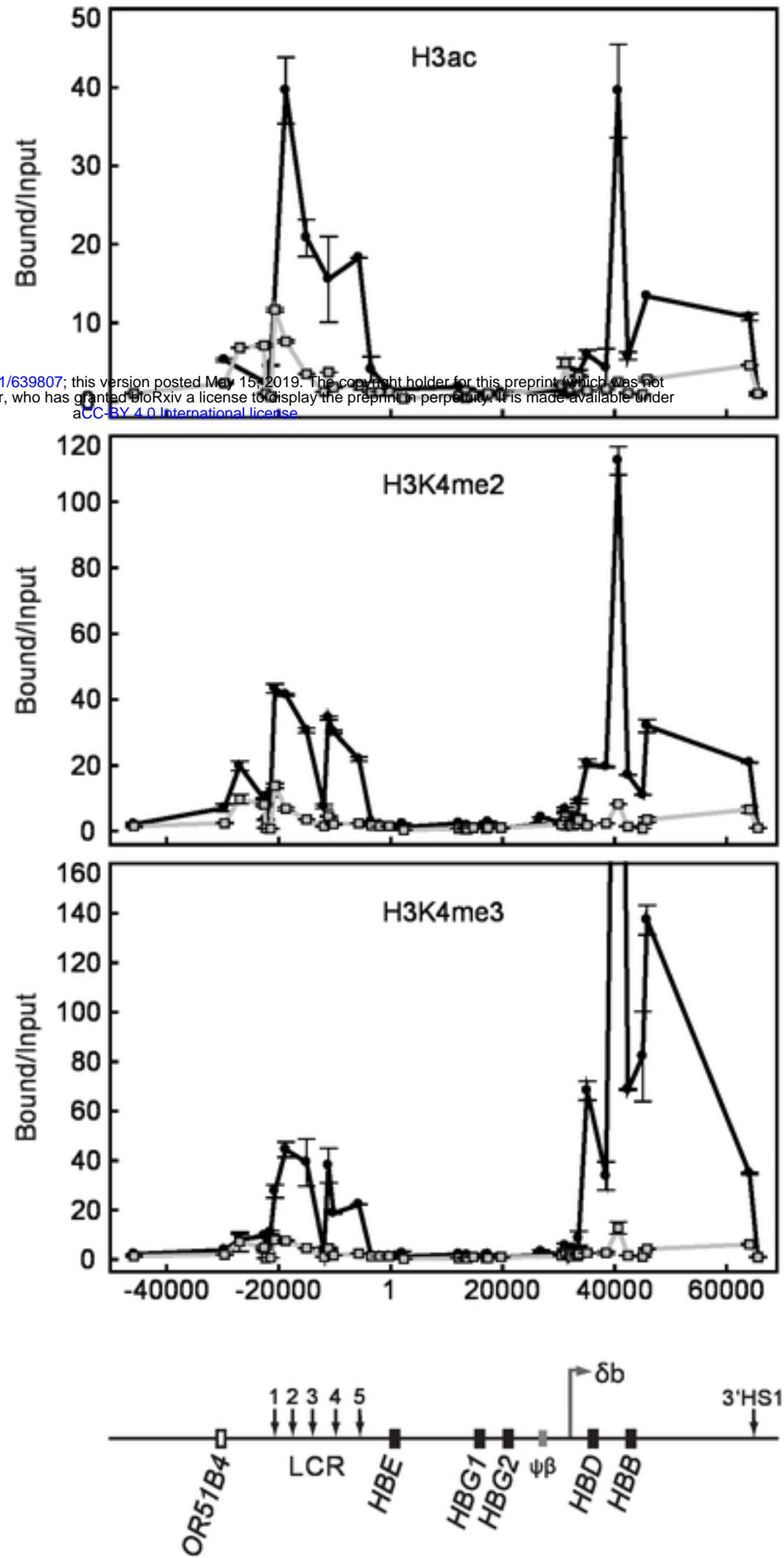


Figure 1  
Debrand *et al.*

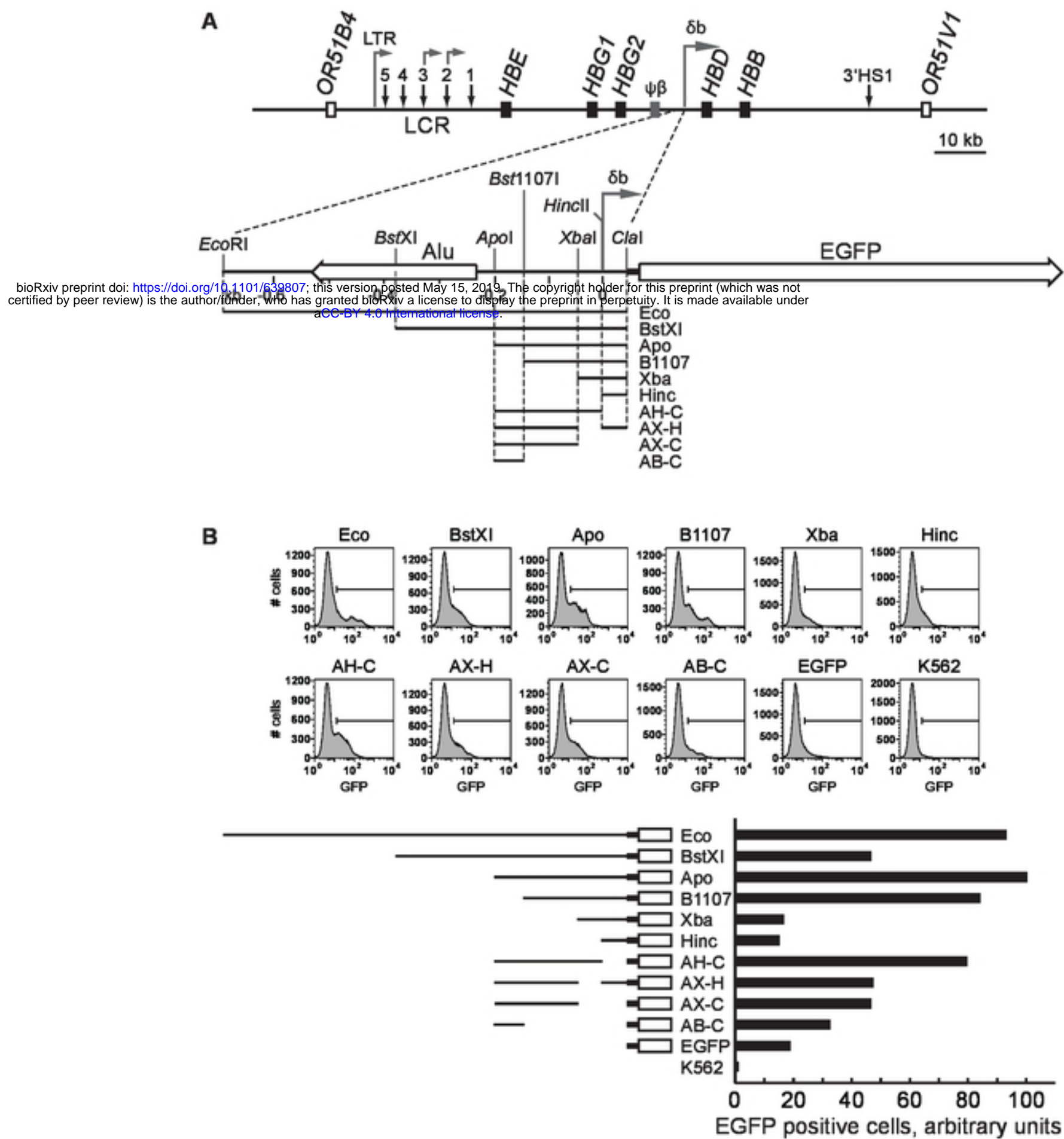
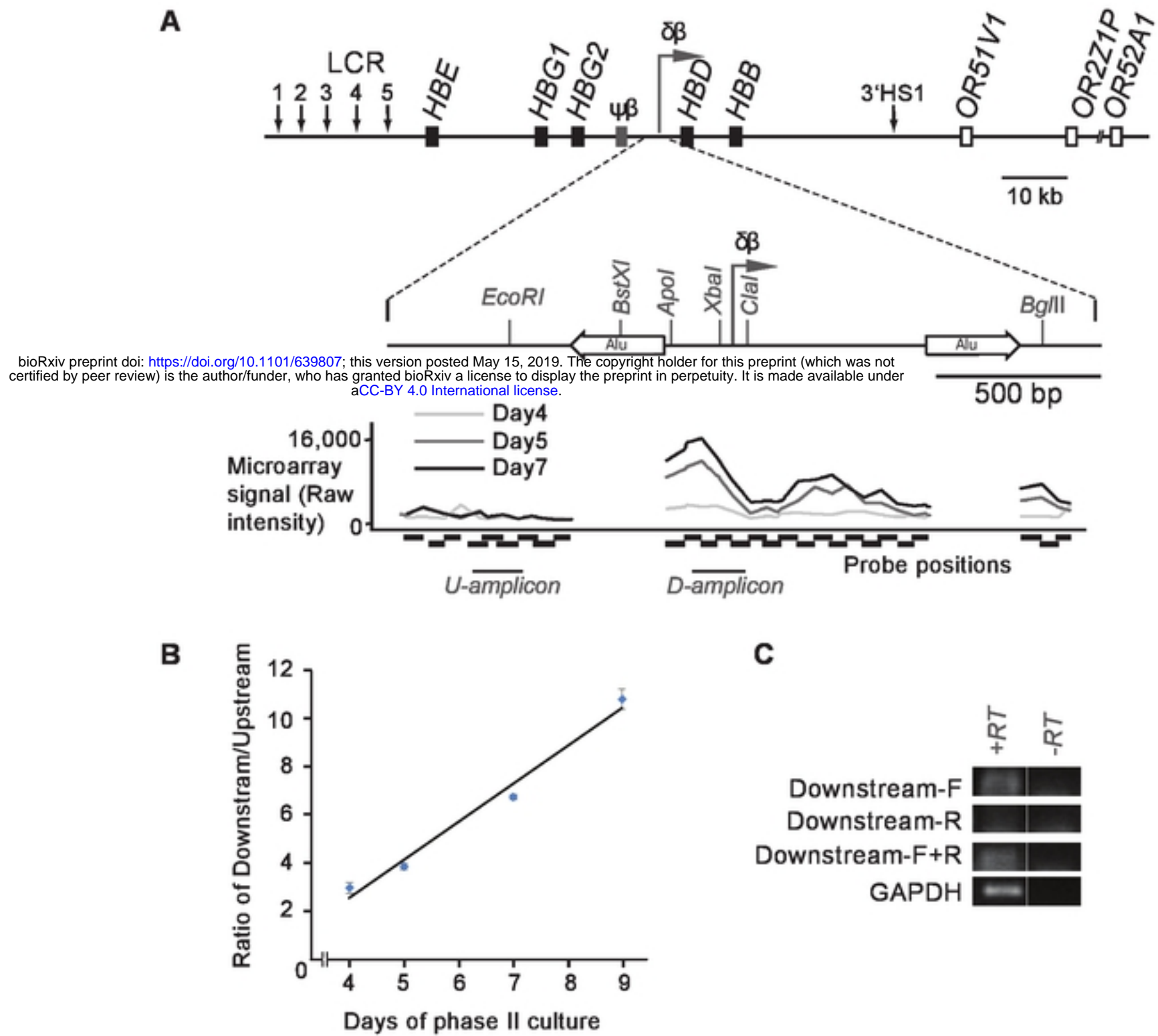


Figure 2  
Debrand *et al.*



bioRxiv preprint doi: <https://doi.org/10.1101/639807>; this version posted May 15, 2019. 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.

Figure 3  
Debrand *et al.*

bioRxiv preprint doi: <https://doi.org/10.1101/639807>; this version posted May 15, 2019. 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.

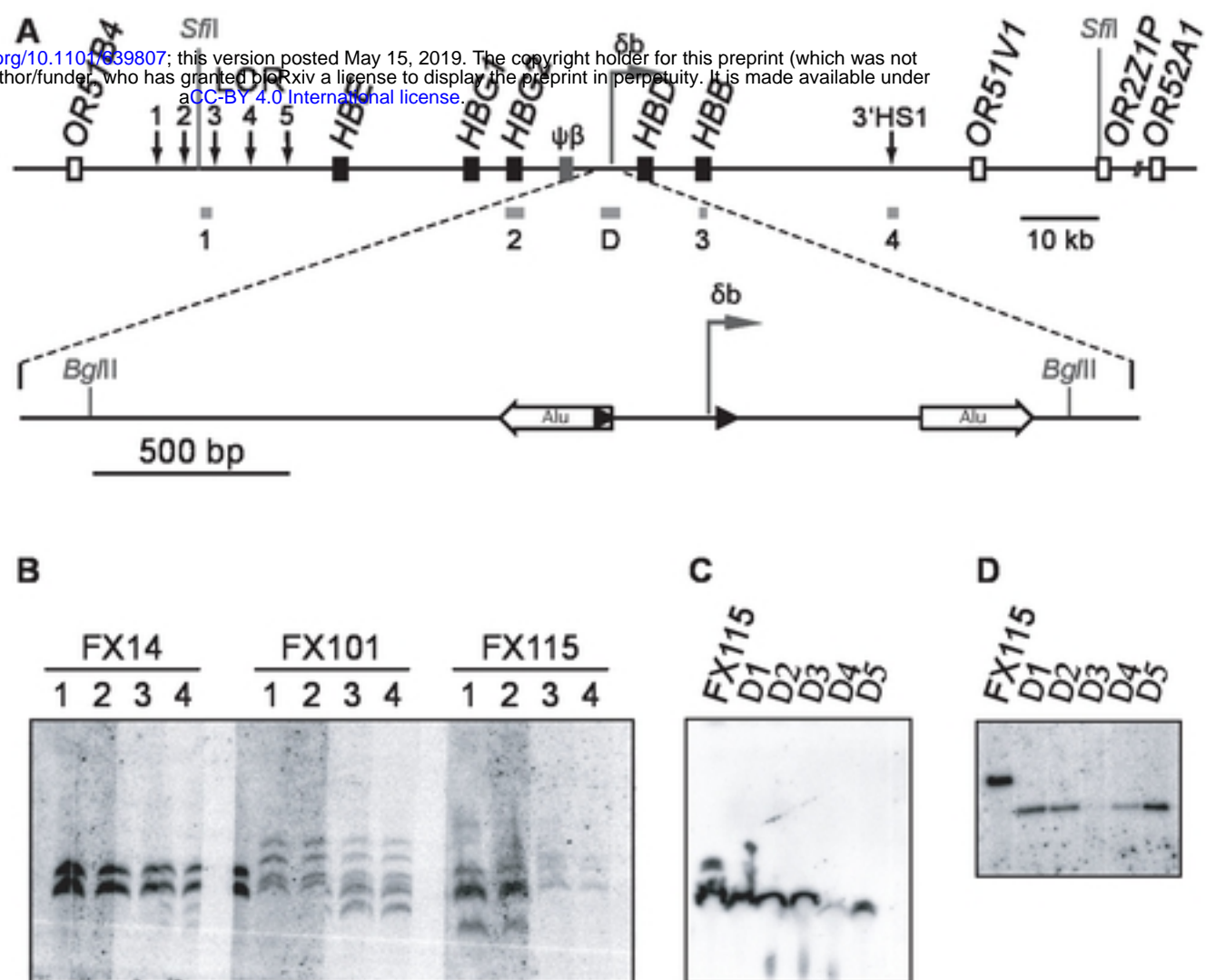




Figure 4  
Debrand *et al.*

