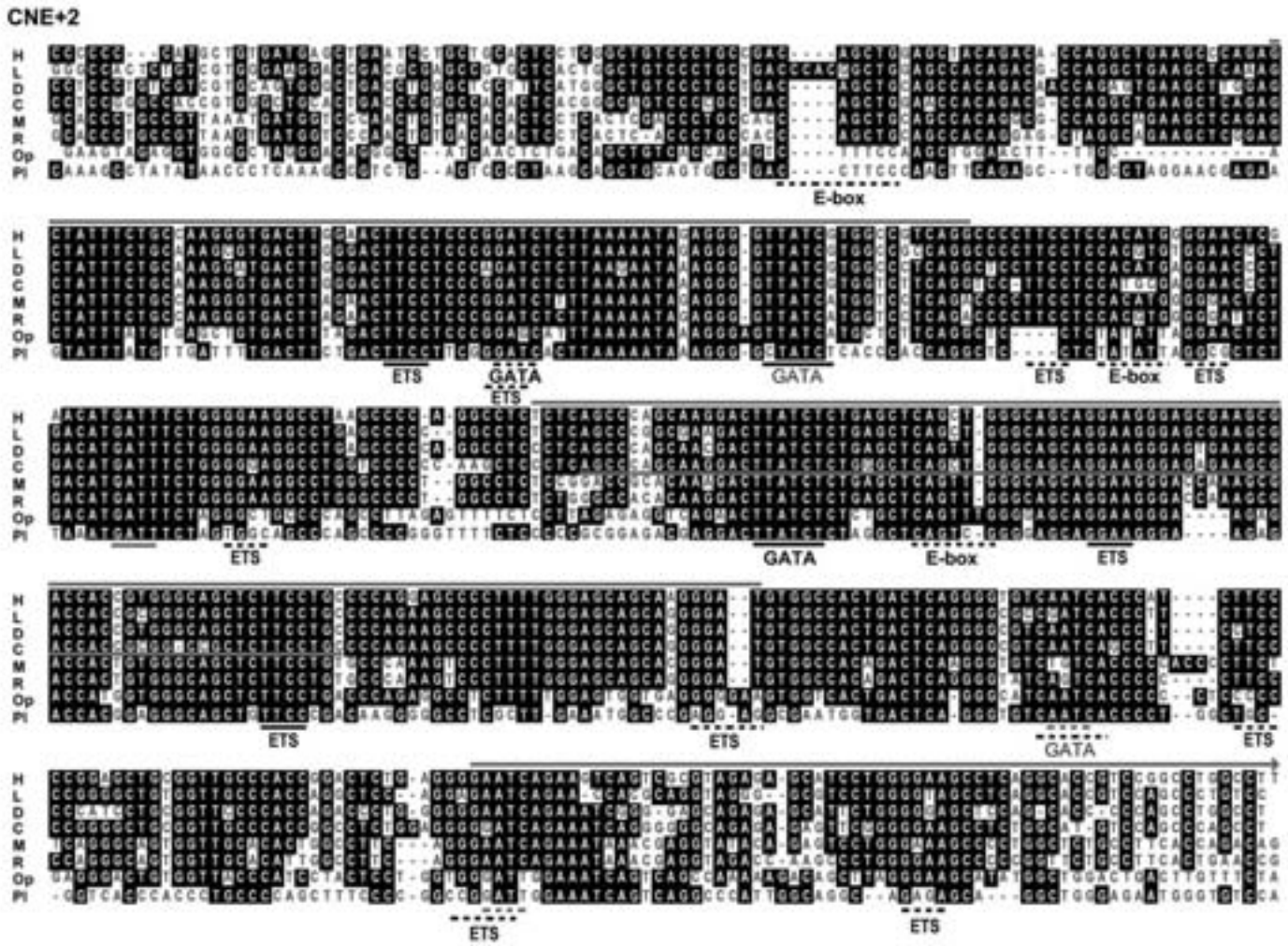
GFI1B controls its own expression binding to multiple sites

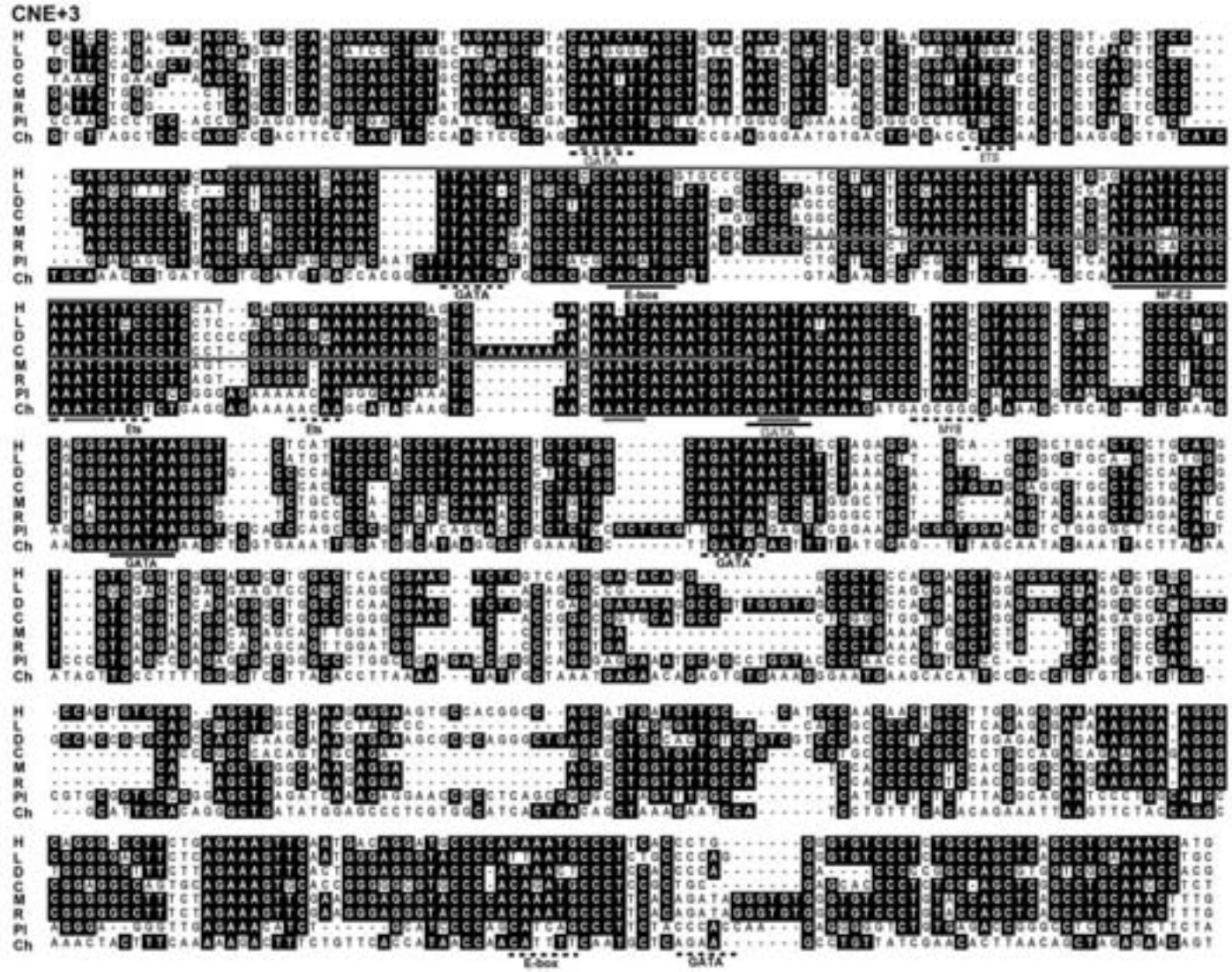
Eduardo Anguita,[1,2] Ana Villegas,[1,2] Francisco Iborra,[3] and Aurora Hernández[2]

[1]University Complutense, Madrid, Spain; [2]Hematology Department, Hospital Clinico San Carlos, Madrid, Spain, and [3]Medical Research Council Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, UK

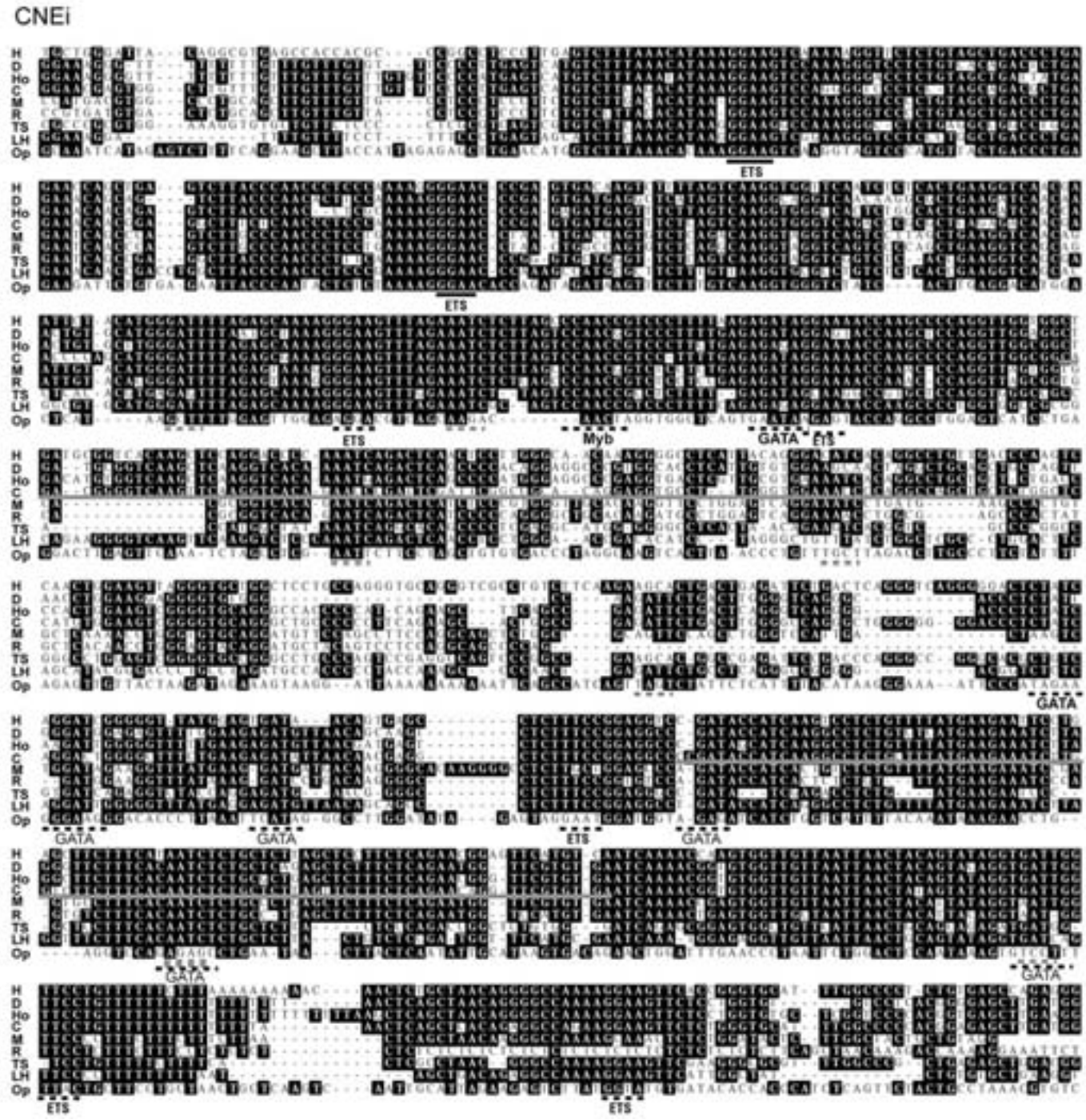
Citation: Anguita E, Villegas A, Iborra F, and Hernández A. GFI1B controls its own expression binding to multiple sites. Haematologica. 2010;95:36-46. doi:10.3324/haematol.2009.012351



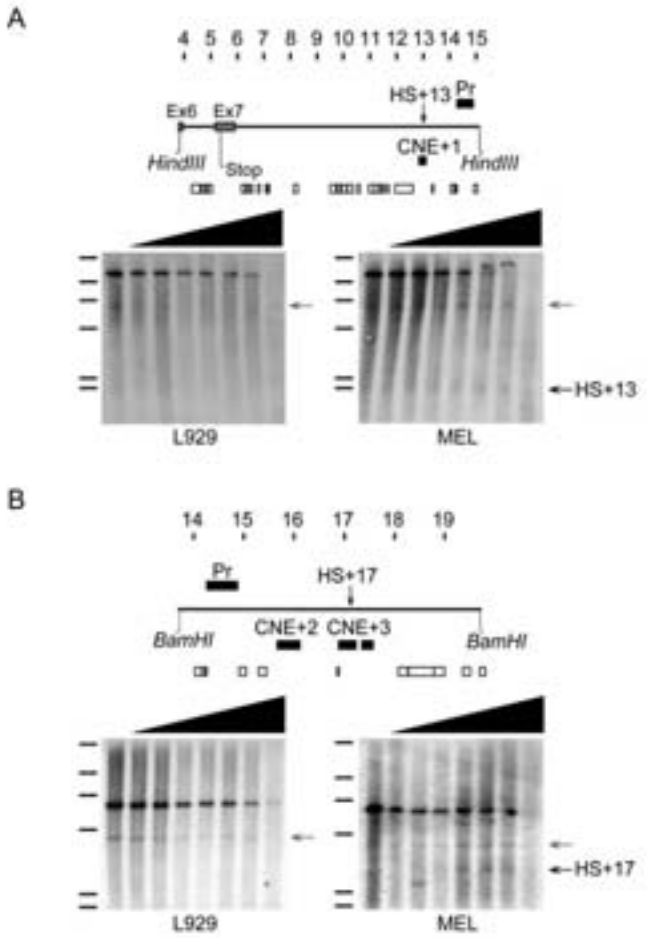
Online Supplementary Figure S1. Conserved transcription factor binding sites corresponding to the alignment of eight species at CNE+2. A GATA binding site with perfect match with the consensus (A/T GATA A/G) in all species and one GATC site (GATA1 N-terminal zinc finger binding motif)[39] in all but one species are shown (black letter font). Two potential GATA binding sites, with up to one mismatch outside the GAT core in some species are indicated (gray letter font). Three E-boxes conserved from human to rodents are present; one is part of an Ebox/GATA motif. Three AATC motifs conserved in all or some species are pointed with gray lines below the sequences. ETS binding sites are also indicated. A dashed line indicates that a site is not conserved in all species. Real-time PCR amplicons are shown as a gray line over corresponding sequence. H indicates human (*Homo sapiens*); L, mouse lemur (*Microcebus murinus*); D, dog (*Canis familiaris*); C, cow (*Bos taurus*); M, mouse (*Mus musculus*); R, rat (*Rattus norvegicus*); Op, opossum (*Monodelphis domestica*); Pl, platypus (*Ornithorhynchus anatinus*).

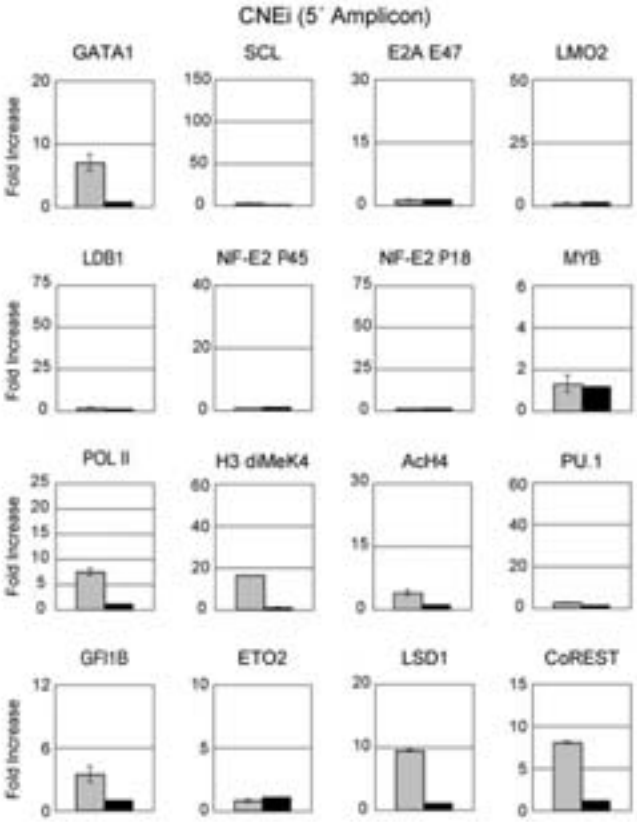


Online Supplementary Figure S2. Conservation analysis of transcription factor binding sites at CNE+3, alignment of eight species. Two areas of homology are distinguished. At the first one there is a GATA binding site with perfect match with the consensus conserved in all species. There are also two other GATA sites with one mismatch in one or two species and two additional sites with one mismatch outside the GAT core in all or most species. An E-box conserved in all cases (part of an Ebox/GATA motif), and an NF-E2 site with one mismatch with respect to the consensus sequence ($^{T}/_{G}$GCTGA$^{C}/_{G}$TCA$^{T}/_{C}$), are indicated. A MYB site (YAACNG) conserved in most cases is also shown. AATC motifs are indicated with gray lines below the sequence. In the most downstream fragment of homology there is an E-box conserved in some species including human and rodents and a perfect match GATA site only present in rodents that constitutes an Ebox/GATA motif. A dashed line indicates that a site is not conserved in all species. Real-time PCR amplicons are shown as a gray line over the sequence. H indicates human (*Homo sapiens*); L, mouse lemur (*Microcebus murinus*); D, dog (*Canis familiaris*); C, cow (*Bos taurus*); M, mouse (*Mus musculus*); R, rat (*Rattus norvegicus*); Pl, platypus (*Ornithorhynchus anatinus*); Ch, chicken (*Gallus gallus*).



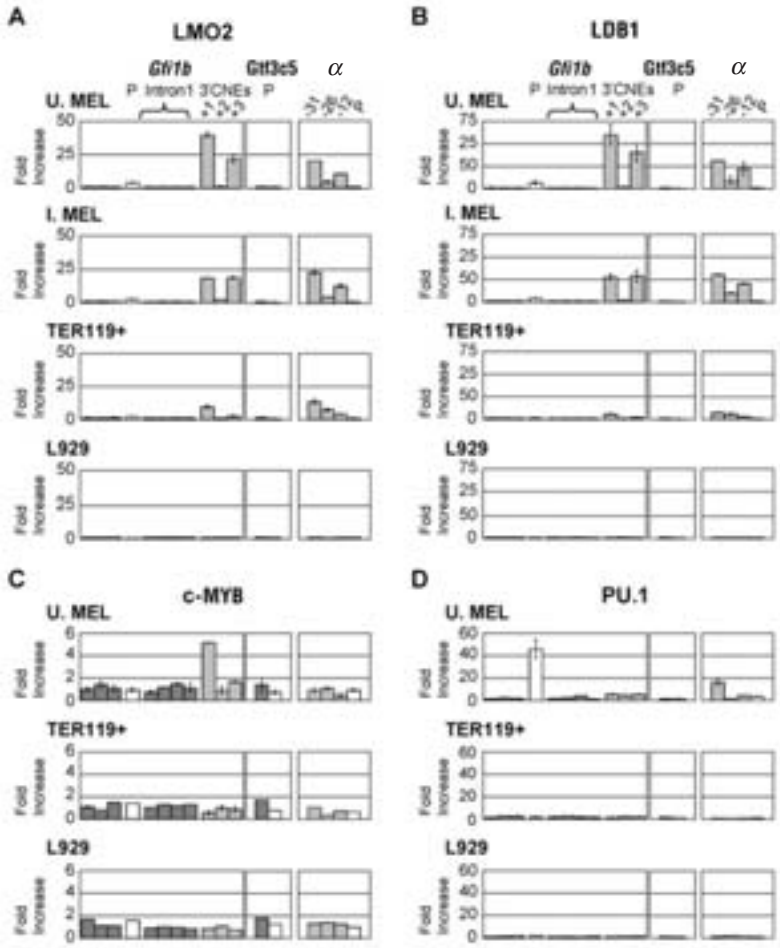
Online Supplementary Figure S3. Sequence conservation at the first *GFI1B* intron (CNEi), alignment of nine species. Plot follows Supplementary Figures S1 and S2. One GATA binding site with perfect match with the consensus is present in most species. Other potential GATA sites are shown. A MYB site (YAACNG) is shown. AATC motifs are underlined in gray, including one contained in the most similar sequence to the *GFI1B* consensus binding site, 75%, (indicated with a thicker line). Upstream and downstream amplicons are shown with gray lines over the sequence. H, human (*Homo sapiens*); L, mouse lemur (*Microcebus murinus*); D, dog (*Canis familiaris*); Ho, horse (*Equus caballus*); C, cow (*Bos taurus*); M, mouse (*Mus musculus*); R, rat (*Rattus norvegicus*); TS, northern tree shrew (*Tupaia belangeri*); LH, lesser hedge tenrec (*Echinops telfairi*); Op, opossum (*Monodelphis domestica*).[3]



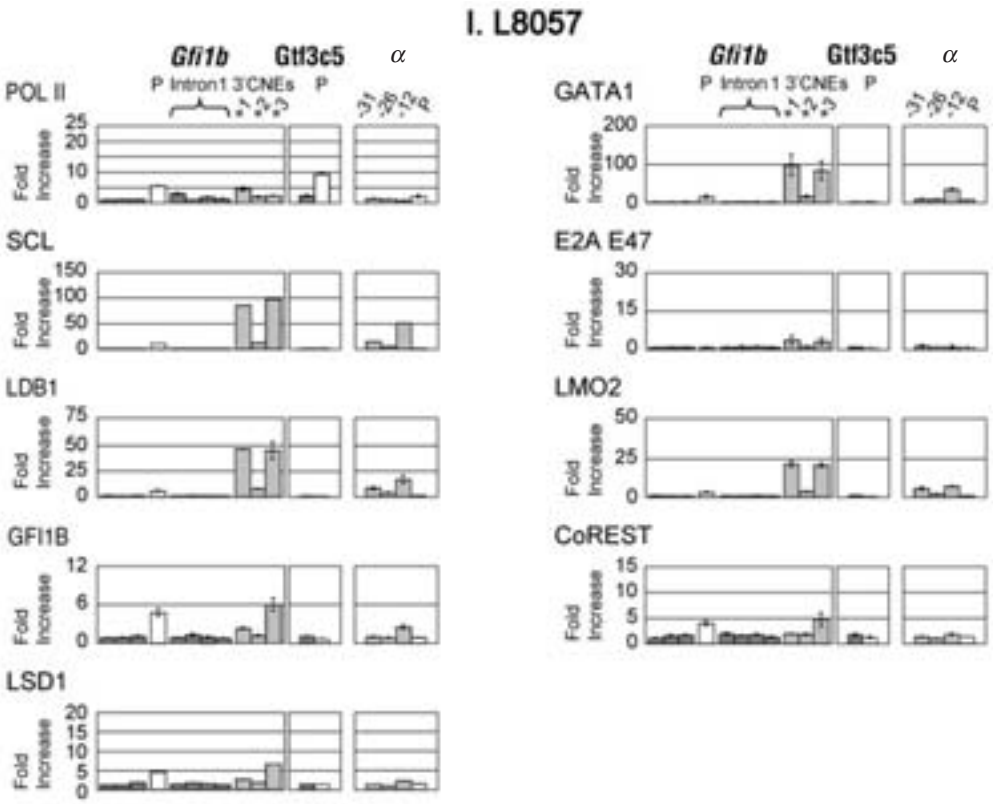


Online Supplementary Figure S4. CNE associate to DNase1 hypersensitive sites (HSs). Representative Southern blots for HSs analysis. Increasing concentrations of DNase1 are indicated as black triangles, first line sample was incubated at 37°C with no DNase1. Left to each blot position of lambda DNA digested with Hind III marker bands are shown. Above: the schematic representation of the area analyzed in each experiment, DNA fragment cut by each restriction enzyme is represented as a black line. Restriction enzyme used; probe (Pr) represented by a black box over the sequence; *GFI1B* exons (Ex), gray boxes on the sequence; CNE (black boxes below the sequence) and repeats (white boxes below the sequence) are indicated. On the top, coordinates in relation to ATG start codon. Black arrows indicate the position of HSs. With DNase1 digestion HS+13 and +17 appear in MEL cells at CNE+1 and CNE+3 locations, but not in L929. Gray arrows indicate sub-bands detectable in all cell types.

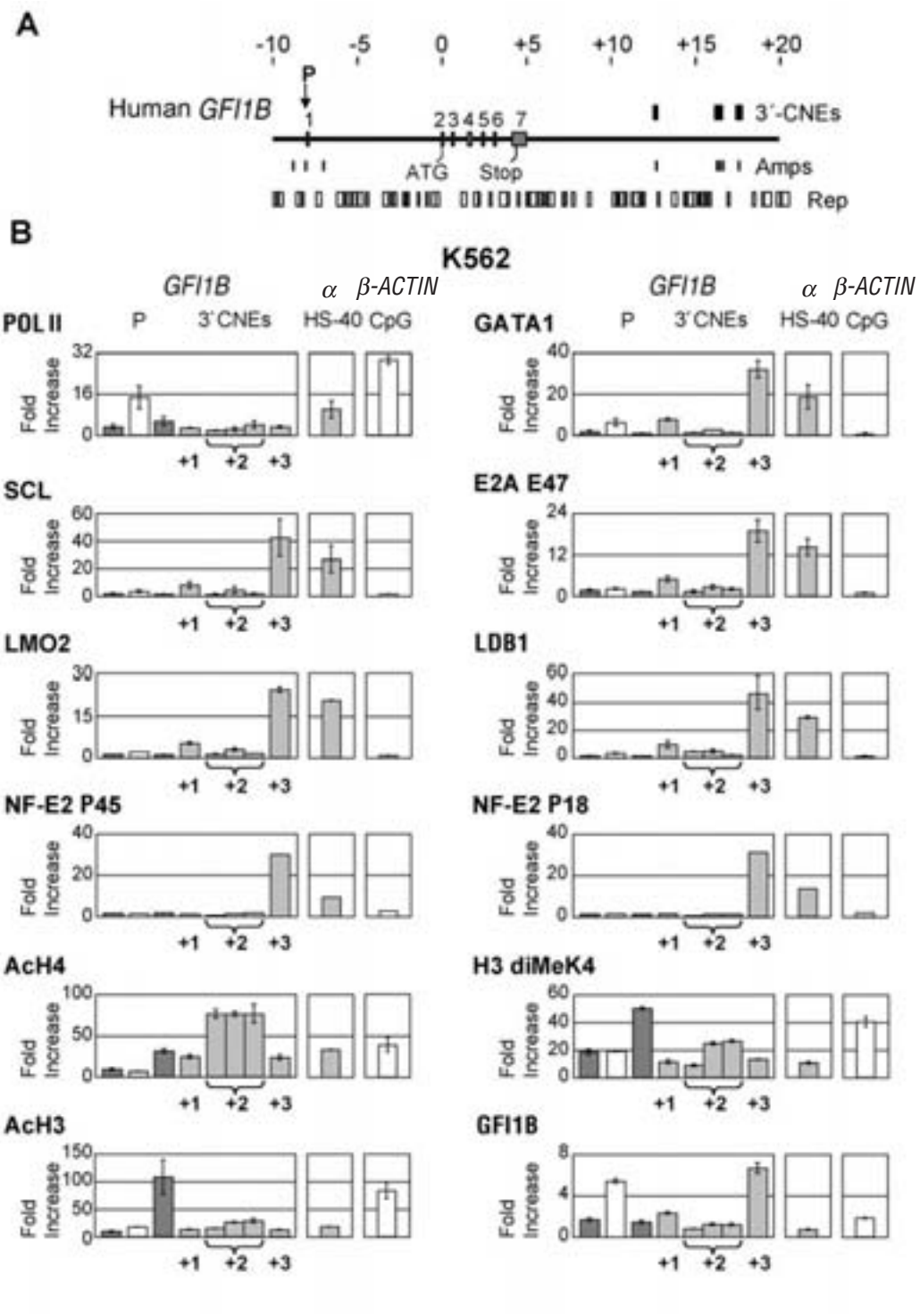
Online Supplementary Figure S5. CNEi binds GATA1, together with repressors, but no other members of the SCL complex. ChIP analysis of CNEi with a Taqman amplicon close to the only perfectly matched GATA binding sequence in this area. This GATA site is approximately 275 bp upstream of a previously used Taqman amplicon, which could be in the limit of the detection. Chromatin modifications, transcription factors and cofactors binding analyses at the upstream side of CNEi in uninduced MEL cells are shown in gray by the left bars and L929 data in black at the right bars, for each antibody. Note that the scale of GATA1 has been reduced to a tenth of Figure 2 E to facilitate the view of the enrichment in MEL cells.[4]



Online Supplementary Figure S6. The remaining components of the pentameric complex bind to mouse *Gfi1b* locus at the promoter and downstream CNE with a similar pattern to GATA1. In contrast, transcription factor binding at specific sites suggests differential roles for downstream CNE. (A) ChIP analysis of LMO2. (B) LDB1 ChIP study. (C) C-MYB analysis demonstrates the binding at CNE+1. (D) PU.1 ChIP showing a high enrichment of *Gfi1b* promoter sequence compared to the low enrichment at the downstream CNE in MEL cells, contrasting with the opposite pattern to GATA1. Plotted as in Figure 2.



Online Supplementary Figure S7. *Gfi1b* locus transcription factor binding in megakaryocytic cells shows a similar pattern to that in erythroid cells. ChIP study of L8057 cells induced to differentiate into the megakaryocytic lineage with TPA performed with the antibodies indicated. Plot follows Figure 2. *α-globin* locus shows some protein binding, mainly at HS-12 in relation to its low level of expression in these cells (*E. Anguita unpublished data*).



Online Supplementary Figure S8. Transcription factor binding at the downstream CNE is conserved between human and mouse erythroid cells. (A) Schematic representation of the human *GFI1B* locus. On the top, coordinates in relation to the ATG start codon. Promoter (P) is pointed with a black arrow; downstream CNE are shown with black boxes over the DNA sequence (black line); gray boxes on the sequence are *GFI1B* exons; black boxes below indicate the position of amplicons used in real-time PCR (Amps); white boxes are repetitive sequences (Rep). Start and stop codons are shown. (B) Real-time PCR analysis of ChIP on K562 cells with antibodies against RNA POL II, GATA1, SCL, E2A (E47), LMO2, LDB1, NF-E2 P45, P18, acetylated histone H4, (AcH4), dimethylated lysine 4 of histone H3 (H3 diMeK4), acetylated histone H3, (AcH3) and GFI1B, as indicated. Enrichment of ChIP DNA over input is normalized to the *18S* control sequence. Results of two amplicons at the promoter (P) are shown. Due to the proximity of CNE+2 and CNE+3, 1 kb apart, ChIP results at the first one could be attributed to the lack of resolution between both, although chromatin sonication rendered fragments below 500 bp for ChIP assay. To confirm that this corresponds to real binding and to localize it accurately, we designed three different Taqman probe and primers sets along the CNE+2 (shown over a brace), which showed an obvious relative enrichment of the central amplicon for all the members of the pentameric complex, suggesting that the binding is weak but real. Error bars correspond to ± 1 SD from two independent ChIP assays. Controls on *α-GLOBIN* HS-40 regulatory element and *β-ACTIN* CpG island are shown.

Online Supplementary Table S1. Oligonucleotides used in real-time PCR analysis of FAIRE and ChIP assays at mouse *Gfi1b* locus. Coordinates are relative to the first base of the start codon. When the amplicon localizes to a relevant site, this is indicated.

Amplicon	Oligonucleotide Coordinate	Oligonucleotide sequence
CNE-2		
	Sense (-15050/-15026)	5'-AGGGTGGAATTTCTCAGAGTCTCTT-3'
	Probe (-14991/-15020)	FAM-5'-AAACTTCTAGGGACTTCCGCCTTTGCATTT-3'-TAMRA
	Antisense (-14967/-14989)	5'-CCTGGCATTTGGTTACTGCTCTA-3'
-8317/-8235		
	Sense (-8317/-8298)	5'-GAGCCACCGTTGAACTGGTT-3'
	Probe (-8285/-8259)	FAM-5'-AGCTTGAGGCCAGAGCTCAGCCTAACA-3'-TAMRA
	Antisense (-8235/-8257)	5'-GCTTTGGCTTTCTAAATGCACAA-3'
CNE-1		
	Sense (-7553/-7532)	5'-GCAAGAGCATGCAAATCAGCTA-3'
	Probe (-7498/-7527)	FAM-5'-TGCATACAGAGAGAACTCCACCAAACAGCA-3'-TAMRA
	Antisense (-7471/-7496)	5'-GGATGACTGATTTATTCCAGGAAGTT-3'
Gfi1b promoter		
	Sense (-6744/-6719)	5'-GAAGCTTCTGTGAAGTTTTGATAAGC-3'
	Probe (-6717/-6695)	FAM-5'-AATGTGGCTGCACTTTCGCGCTC-3'-TAMRA
	Antisense (-6674/-6691)	5'-TGCAGCGTCAGCCAATGA-3'
-5720/-5650		
	Sense (-5720/-5699)	5'-GAGTTGGACAGCCTGGGTACTG-3'
	Probe (-5697/-5671)	FAM-5'-CAGAGCTGTGCCTATCCACCTGTGCTC-3'-TAMRA
	Antisense (-5650/-5669)	5'-GAGGGCTTTGCAGACTTGGA-3'
CNEi (5'side)		
	Sense (-3582/-3565)	5'-TGCACGCGGTCACAGAAT-3'
	Probe (-3563/-3537)	FAM-5'-AGACTCATCTCCCGTGGGTTGCACAAG-3'-TAMRA
	Antisense (-3502/-3523)	5'-CAGTGGCTTCATCAGGATTTCC-3'
CNEi (3'side)		
	Sense (-3358/-3330)	5'- AGATACCATCACTGTCTCTGTTTTATGAA-3'
	Probe (-3327/-3300)	FAM-5'-AATCCGAGTGCCTTTCACAATCTCTGCC-3'-TAMRA
	Antisense (-3270/-3293)	5'- TCACACGAACCATTCTGAGAAAGA-3'
5' Ebox Intron 1		
	Sense (-2806/-2789)	5'- CGCCCAGCTGCTTTTGAT-3'
	Probe (-2787/-2759)	FAM-5'-CAATCATCTATACCAGAGGCCGCTGTGAG-3'-TAMRA
	Antisense (-2736/-2754)	5'- TCCCCACCACACCATCTGT-3'
3' Ebox Intron 1		
	Sense (-271/-248)	5'-CTCCCTGCTTTTGATACACAAGTC-3'
	Probe (-241/-220)	FAM-5'-CCTGCGCCCAGCTTTCCCTCTT-3'-TAMRA
	Antisense (-198/-218)	5'-CCCAGATGTCCCTCCCTACAA-3'
CNE+1		
	Sense (+12970/+12991)	5'-GCTCAGGTTTGGCCTGATAAGT-3'
	Probe (+13012/+13035)	FAM-5'-CCAACTGTCAGCATGGAGCCTGGC-3'-TAMRA
	Antisense (+13073/+13054)	5'-CGCTATCGACCCAGGAGTGT-3'
CNE+2		
	Sense (+15911/+15934)	5'-GACCGCACAAAGACTTATCTCTGA-3'
	Probe (+15958/+15937)	FAM-5'-TCCCTTCCTGCTGCCCAACTGA-3'-TAMRA
	Antisense (+15990/+15970)	5'-ACAGGAAGAGCTGCCCACAGT-3'
CNE+3		
	Sense (+17070/+17086)	5'-CCCAACCCCCTCAACCA-3'
	Probe (+17093/+17120)	FAM-5'-CAGCATGACACAGCAAATCTTCCCTCAG-3'-TAMRA
	Antisense (+17155/+17130)	5'-TGACATTGTGATTTCTCATCCTTGTT-3'
+31012/+31088		
	Sense (+31012/+31031)	5'-ATCCCAGGCAGGACCAAGTC-3'
	Probe (+31036/+31063)	FAM-5'-TCTCTCAGAAGCTTGGCCTGAGGACCTC-3'-TAMRA
	Antisense (+31088/+31067)	5'-GGGAGACAAAGACAAGGACCAA-3'
Gtf3c5		
	Sense (+32261/+32281)	5'-GGCCTTCGAGTCTCCATGATT-3'
	Probe (+32311/+32284)	FAM-5'-CCGGTCCTCCTTACCTTTCTCGTCTCCT-3'-TAMRA
	Antisense (+32335/+32318)	5'-CAAAGGTTCCGCGAAGGA-3'
c16orf8 or *Dist*		
	Sense (55413/55434)	5'-AACATTGACCCTGTCATGAGCA -3'
	Probe (55437/55463)	FAM-5'-CAGGGAATGCCAAACACTCCAACAGCT-3'-TAMRA
	Antisense (55484/55484)	5'-GCATGTGATGTGGCCTCTTG-3'

Online Supplementary Table S2. Oligonucleotides used in real-time PCR at human *GFI1B* locus. Coordinates relative to the first base of the start codon. When the amplicon localizes to a relevant site, this is indicated.

Amplicon	Coordinates	Oligonucleotide sequence
-8819/-8722		
	Sense (-8819/-8797)	5′-TTCAATCTGTATGGCAGCTTCAG-′3
	Probe (-8771/-8747)	FAM-5′-AACTCACTGCCACACCTCCAGGGTG-3′-TAMRA
	Antisense (-8722/-8745)	5′-GATATGACCTTTGAGCGGGTAGAG-′3
***GFI1B* promoter**		
	Sense (-8124/-8101)	5′-GTCTTGTGTCCTGGAAAGTTTTGA-′3
	Probe (-8099/-8076)	FAM-5′-AAGCAAATACGGCTGAGCTCCCGC-3′-TAMRA
	Antisense (-8050/-8072)	5′-GTGAACAAGCAGCCAATGAAGAG-′3
-7073/-7000		
	Sense (-7073/-7056)	5′-GCTGCCCTCTCCCACACA-′3
	Probe (-7051/-7025)	FAM-5′-CCAGGCTCCAGGAACAATAGGACCGAC-3′-TAMRA
	Antisense (-7000/-7023)	5′-CGAGCTGAAACAGTACCTCTCCAT-′3
CNE+1		
	Sense (+12648/+12670)	5′-GCTCAGGTTTGGCCTGATAAGTT-′3
	Probe (+12683/+12701)	FAM-5′-CCGGCTGCCAACCGTCGGT-3′-TAMRA
	Antisense (+12725/+12705)	5′-GCTTGGCCCATGGTTCTACTC-′3
CNE+2a		
	Sense (+16259/+16281)	5′-GCTATTTCTGCCAAGGGTGACTT-′3
	Probe (+16286/+16312)	FAM-5′-CTTCCTCCCGGATCTCTTAAAAATAGA-3′-TAMRA
	Antisense (+16333/+16315)	5′-CCTGACGGCCACGATAACC-′3
CNE+2b		
	Sense (+16396/+16417)	5′- TCTCAGCCCAGCAAGGACTTAT-′3
	Probe (+16466/+16447)	FAM-5′-CCCACGGTGGTCGCTTCGCT-3′-TAMRA
	Antisense (+16511/+16493)	5′- ATCCCTTGCTGCTCCCAAA-′3
CNE+2c		
	Sense (+16583/+16606)	5′- GAATCAGAAGTCAGTCGCGTAGAG-′3
	Probe (+16620/+16639)	FAM-5′-AGCCTCAGGCACCGTCCGGC-3′-TAMRA
	Antisense (+16662/+16644)	5′- AGGGCTTGGAGCAGAAAGG-′3
CNE+3		
	Sense (+17554/17574)	5′- CCCGGCCTGAGACTTATCACT-′3
	Probe (+17595/+17619)	FAM-5′-CTCCTCCTCCAACCACCTCACCCTG-3′-TAMRA
	Antisense (+17645/+17623)	5′- ATGGAGGGAAGATTTGCTGAATC-′3
***β-ACTIN* CpG island**		
	Sense	5′- CGGCCAACGCCAAAACT-′3
	Probe	FAM-5′-TCCCTCCTCCTCTTCCTCAATCTCGC-3′-TAMRA
	Antisense	5′-CCCTCTCCCCTCCTTTTGC-3

Onine Supplementary Table S3. Oligonucleotides used in real-time PCR analysis of 3C at mouse *Gfi1b* locus (coordinates relative to the first base of the start codon) and *α-globin* HS-31 control (coordinates from the telomere).[31]

Element	Coordinates	Oligonucleotide sequence
Gfi1b promoter		
	-6119/-6100	5´-CACCAGAGGAGCCCAAATTG-3´
	-6095/-6075	FAM-5´-CCCCCTGGGCTAGAGCTGCCC-3´-TAMRA
Exon 7		
	+5467/+5444	5´-CTCAAAAGAAGCAGAGAGGCTCAT-3´
+8802/+9399		
	+8827/+8803	5´-CCAAGATTTTTCCTCATCCTCAGAT-3´
CNE+1		
	+12940/+12919	5´-GATTTCCCTCCAGTTGCCTTTC-3´
CNE+3		
	+17823/+17848	5´-GCTGGAATGGAGATATAAATGGTCAT-3´
HS-31		
	85870/85892	5'-TTCTGACCTCACCTCAGCTAAGC-3'
	85894/85919	FAM-5'-TCTTCCTCCTCTGAGAATCCGCCATG-3'-TAMRA
	85943/85925	5'-TGTGTGGGCAGAGGACACA-3'