Table S1. Highly significant SNPs in HMIP and *BCL11A* found in 619 Chinese β-thalassemia heterozygotes.

Gene	Gene position	Chromosome	SNP	Chromosome position	Minor/Major allele	Minor allele frequency	β [#]	*P*-value
ALDH8A1	5'	6q23	**rs4646869**	60,577,176	A / G	**0.28**	**– 0.22**	**2.80 E-08**
HMIP	Sequences 5' to *HBS1L* and *MYB*	6q23	**rs6913541**	135,450,809	G / A	**0.31**	**– 0.26**	**9.16 E-11**
			rs9494139	135,456,486	G / A	**0.32**	**– 0.23**	**1.19 E-09**
			rs7775698	135,460,328	T / C	**0.22**	**0.44**	**1.38 E-23**
			rs9399137	135,460,711	C / T	**0.21**	**0.45**	**1.39 E-24**
			rs4895441	135,468,266	G / A	**0.24**	**0.40**	**5.93 E-21**
			rs9376092	135,468,837	A / C	**0.24**	**0.40**	**2.73 E-21**
			rs9494145	135,474,245	C / T	**0.22**	**0.39**	**1.39 E-18**
			rs1320963	135,484,905	G / A	**0.37**	**– 0.24**	**1.11 E-10**
			rs1569534	135,493,273	T / C	**0.40**	**– 0.20**	**4.49 E-08**
			rs6569992	135,493,845	A / G	**0.21**	**0.34**	**3.91 E-14**
BCL11A	3' UTR	2p16	**rs1012585**	60,525,061	G / A	**0.29**	**0.23**	**1.29 E-08**
BCL11A	IVS II	2p16	**rs17331129**	60,533,446	C / T	**0.26**	**0.23**	**9.46 E-08**
			rs12477097	60,551,901	A / C	**0.26**	**0.30**	**4.36 E-12**
			rs6545816	60,568,365	A / C	**0.24**	**0.34**	**2.49 E-14**
			rs6545817	60,568,683	C / T	**0.24**	**0.33**	**2.43 E-14**
			rs766432	60,573,474	C / A	**0.24**	**0.34**	**2.40 E-15**
			rs6729815	60,577,176	T / C	**0.24**	**0.34**	**1.56 E-14**

[#] β equals % change in Hb F per copy of minor allele.

Table S2. Genotyped and imputed SNPs in LD with rs9399137 among Chinese and African Americans, their MAF and association *P*-value with HbF.

SNP	Chr. 6 Position	SNP Type	Chinese β-thalassemia heterozygotes			African Americans with sickle cell anemia (11)		Candidate for functional motif
			LD to rs9399137	MAF	*P*-value	MAF	*P*-value	
rs9399136	135,444,032	Imputed	0.97	0.23	2.84E-22	0.05	5.24E-02	Possible
rs1331309	135,447,871	Imputed	0.96	0.27	2.71E-20	0.05	5.81E-02	Possible
rs9376090	135,452,921	Imputed	0.97	0.23	1.68E-22	0.04	3.15E-03	Possible
rs7775698	135,460,328	Genotyped	0.99	0.22	1.36E-23	0.19	1.58E-01	No
rs7776054	135,460,609	Imputed	0.99	0.22	1.48E-23	0.25	3.76E-02	No
rs9399137	135,460,711	Genotyped	1.00	0.22	7.34E-24	0.05	4.20E-06	Possible
rs11321816	135,460,731	Imputed	0.99	0.24	1.94E-22	0.03	4.66E-03	Possible
rs56293029	135,460,732	Imputed	0.98	0.25	3.43E-22	0.17	6.86E-01	No
rs35786788	135,460,735	Imputed	0.99	0.22	7.78E-24	0.03	1.78E-03	Possible
rs9389268	135,461,324	Imputed	0.99	0.22	7.61E-24	0.19	1.36E-01	No
rs9376091	135,461,329	Imputed	0.93	0.24	1.93E-22	0.21	3.80E-01	No
rs9402685	135,461,381	Imputed	0.99	0.22	7.50E-24	0.28	2.10E-01	No
rs11969203	135,462,591	Imputed	0.88	0.17	3.50E-20	0.45	7.13E-04	No
rs11759553	135,463,989	Imputed	0.83	0.24	4.21E-21	0.32	6.12E-04	No
rs9373124	135,464,902	Imputed	0.84	0.24	5.22E-21	0.44	2.43E-03	No
rs35959442	135,465,872	Imputed	0.83	0.24	5.26E-21	0.31	1.12E-03	No
rs4895440	135,468,251	Imputed	0.83	0.24	5.85E-21	0.31	4.93E-04	No
rs4895441	135,468,266	Genotyped	0.83	0.24	5.94E-21	0.09	1.53E-05	Possible
rs9376092	135,468,837	Genotyped	0.83	0.24	5.92E-21	0.20	1.93E-02	No
rs9389269	135,468,852	Imputed	0.83	0.24	5.92E-21	0.08	1.87E-07	Possible
rs9402686	135,469,510	Imputed	0.84	0.24	5.02E-21	0.08	1.83E-07	Possible
rs7758845	135,470,230	Imputed	0.83	0.24	5.49E-21	0.30	2.42E-02	No
rs6920211	135,473,011	Imputed	0.76	0.24	2.32E-21	0.35	3.43E-03	No
rs11154792	135,473,333	Imputed	0.76	0.24	2.64E-21	0.12	2.17E-08	Possible
rs9494145	135,474,245	Genotyped	0.60	0.22	1.73E-18	0.07	4.10E-08	Unlikely
chr6:135474996	135,474,996	Imputed	0.48	0.32	1.65E-16	0.0	---	No
rs9483788	135,477,194	Imputed	0.52	0.25	4.59E-16	0.07	8.58E-08	Unlikely

Table S3

Illumina SNP assay for rs7775698

Illumina Infinium II Assay ID rs7775698-131_B_F_1863223089 was used to genotype rs7775698, at 6: 135,460,328.
The first step was hybridization of unlabeled DNA fragment to the 50-bp probe, colored in green in the following table, on the array.
The second step was enzymatic single base extension with labeled nucleotide to interrogate the rs7775698 genotype.

A. For rs7775698 with the major allele **C** as found in most subjects, the probe would identify the **C** nucleotide, colored in red below.

```
      CTCTACTGACACAACATCAGGATTAAATTCACTCTGGACAGCAGATGTTA
TAAGAAACTCTACTGACACAACATCAGGATTAAATTCACTCTGGACAGCAGATGTTACTATATCAAAACCA
```

B. For rs7775698 with the minor allele **T**, without the 3-bp deletion as found in most African subjects, the probe would identify the **T** nucleotide, colored in red below.

```
      CTCTACTGACACAACATCAGGATTAAATTCACTCTGGACAGCAGATGTTA
TAAGAAACTCTACTGACACAACATCAGGATTAAATTCACTCTGGACAGCAGATGTTATTATATCAAAACCA
```

C. The nucleotide sequences with the 3-bp deletion, 6: 135,460,326 to 135,460,328 as found in the Chinese population for example, are shown below with the " - - -" representing the 3-bp deletion.
The 'final' sequences with the 3-bp deletion are shown below on the last line.
The probe used to genotype rs7775698 would still identify a **T** nucleotide, in spite of the presence of the 3-bp deletion.

```
      CTCTACTGACACAACATCAGGATTAAATTCACTCTGGACAGCAGATGTTA
TAAGAAACTCTACTGACACAACATCAGGATTAAATTCACTCTGGACAGCAGATGT---TATATCAAAACCA
TAAGAAACTCTACTGACACAACATCAGGATTAAATTCACTCTGGACAGCAGATGTTATATCAAAACCA
```

Table S4

MAF of and LD between rs7775698 and rs9399137 among Different Populations from HapMap database.

	Populations outside of Africa	r^2 between rs7775698 and rs9399137	rs7775698 MAF	rs9399137 MAF	Frequency of 3-bp deletion based on haplotype rs7775698(T) rs9399137(C)
CEU	Utah residents with Northern and Western European ancestry	1.00	0.20	0.20	0.20
CHB	Han Chinese in Beijing, China	0.88	0.29	0.27	0.27
CHD	Chinese in Metropolitan Denver, Colorado	0.94	0.32	0.29	0.29
GIH	Gujarati Indians in Houston, Texas	0.95	0.13	0.13	0.13
JPT	Japanese in Tokyo, Japan	1.00	0.32	0.32	0.32
MEX	Mexican ancestry in Los Angeles, California	0.85	0.19	0.15	0.15
TSI	Toscans in Italy	0.97	0.25	0.24	0.24
	Mean ± SD	**0.94 ± 0.06**	**0.24 ± 0.07**	**0.23 ± 0.07**	**0.23 ± 0.07**

	Populations of African Descent	r^2 between rs7775698 and rs9399137	rs7775698 MAF	rs9399137 MAF	Frequency of haplotype rs7775698(T) rs9399137(C)
ASW	African ancestry in Southwest USA	0.22	0.18	0.08	0.08
LWK	Luhya in Webuye, Kenya	0.22	0.21	0.06	0.06
MKK	Maasai in Kinyawa, Kenya	0.16	0.25	0.05	0.05
YRI	Yoruba in Ibadan, Nigeria	0.14	0.23	0.04	0.04
	Mean ± SD of LWK, MKK, YRI	**0.17 ± 0.04**	**0.22 ± 0.02**	**0.05 ± 0.01**	**0.05 ± 0.01**

Table S5

Transcription factor binding sites near candidate functional variants.

SNP and Deletion	Chr. 6 Position	Nearby Putative Transcription Factor Binding Sites (Score > 10)	GATA motif
rs9399136	135,444,032	---	---
rs1331309	135,447,871	HMG-1Y; SQUA	---
rs9376090	135,452,921	---	GATA
3-bp deletion	135,460,326 – 135,460,328	TAL1/SCL; E47; RUNX1/AML1	GATA
rs9399137	135,460,711	TAL1/SCL; E47	---
rs11321816	135,460,731	---	GATA
rs35786788	135,460,735	---	GATA
rs4895441	135,468,266	---	---
rs9389269	135,468,852	---	---
rs9402686	135,469,510	RORalfa-1; Broad complex 1	---
rs11154792	135,473,333	---	GATA

Figure S1A Linkage disequilibrium plot for the HMIP region on chromosome 6q23.

Figure S1B Linkage disequilibrium plot for the *BCL11A* region on chromosome 2p16.

Figure S2 Phylogenetic alignment of sequences in region surrounding the TAC 3-bp deletion, showing conservation among many vertebrates, as displayed by the 44-Species Conservation Track from UCSC Genome Browser. The region is flanked by 2 highly conserved sequences with log odds conserved element scores of 64 and 84.

Figure S3 Chromatin immunoprecipitation (ChIP) assay.

A Nucleotide sequences encompassing the 3-bp deletion polymorphism showing the PCR primers used in the ChIP assay.

B Real-time PCR standard curve. IgG, control using IgG for immunoprecipitation.

C Relative real-time PCR signal intensities of the ChIP assay, showing significant binding of TAL1, E47, GATA2, RUNX1 to the immediate vicinity of the 3 bp deletion polymorphism. H3, positive control using antibody against histone H3, and RT-PCR using exon 3 primer set.

Figure S1A

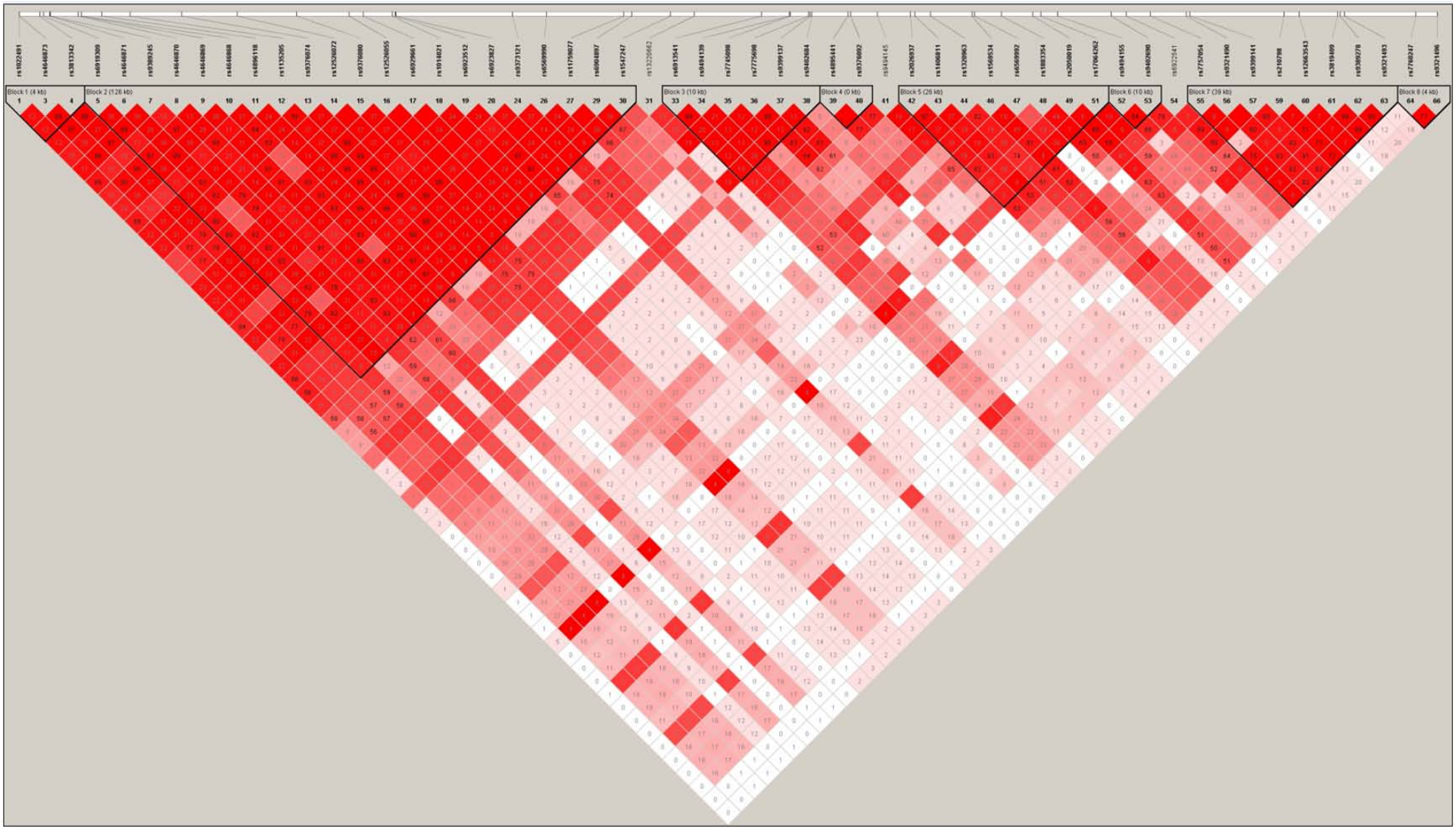

Block 1 (4 kb)
Block 2 (126 kb)
Block 3 (10 kb)
Block 4 (0 kb)
Block 5 (26 kb)
Block 6 (10 kb)
Block 7 (39 kb)
Block 8 (4 kb)

Figure S1B

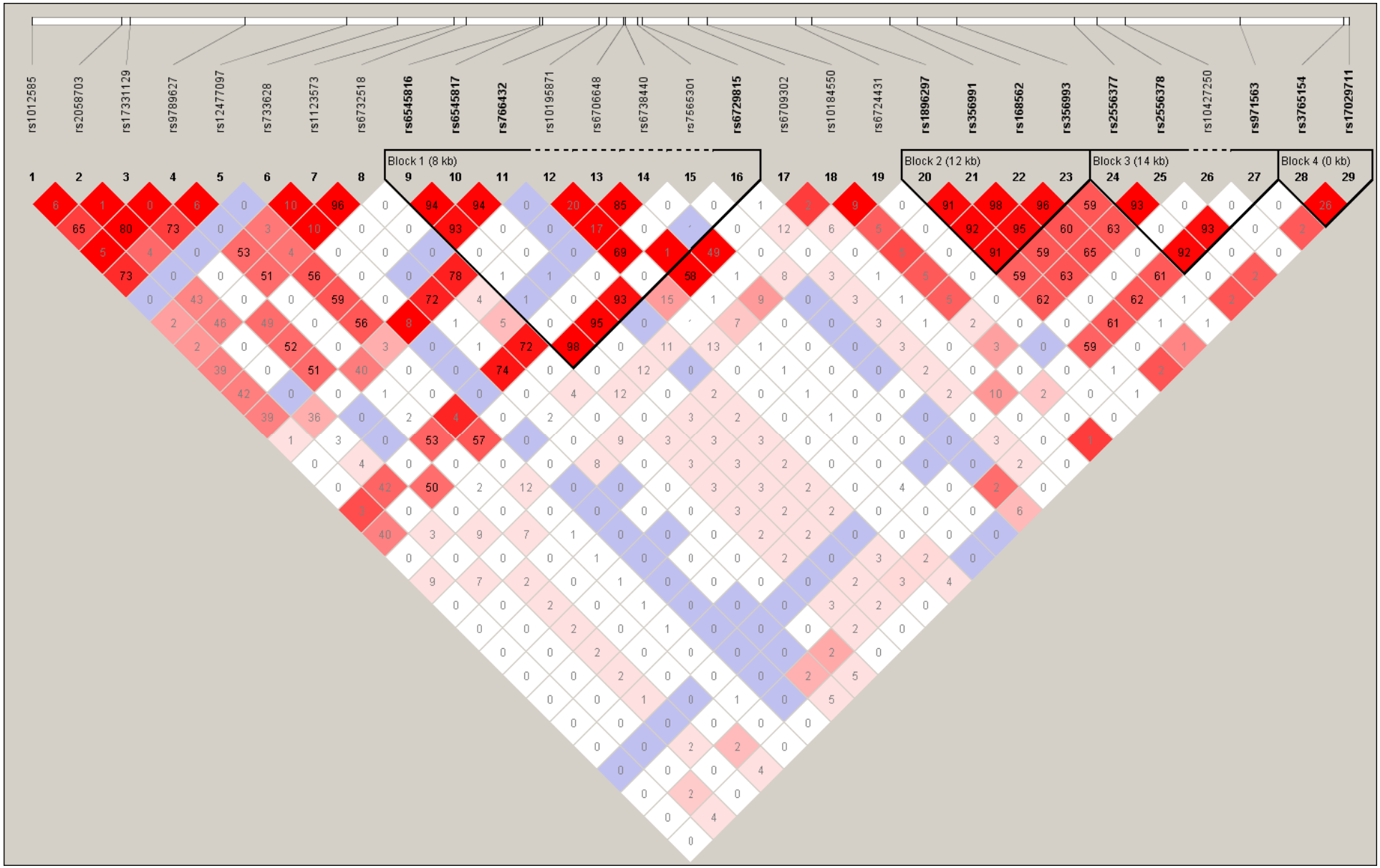

rs1012585
rs2058703
rs17331129
rs9789627
rs12477097
rs733628
rs1123573
rs6732518
rs6545816
rs6545817
rs766432
rs10195871
rs6706648
rs6738440
rs7566301
rs6729815
rs6709302
rs10184550
rs6724431
rs1896297
rs356991
rs168562
rs356993
rs2556377
rs2556378
rs10427250
rs971563
rs3765154
rs17029711
Block 1 (8 kb)
Block 2 (12 kb)
Block 3 (14 kb)
Block 4 (0 kb)

Figure S2

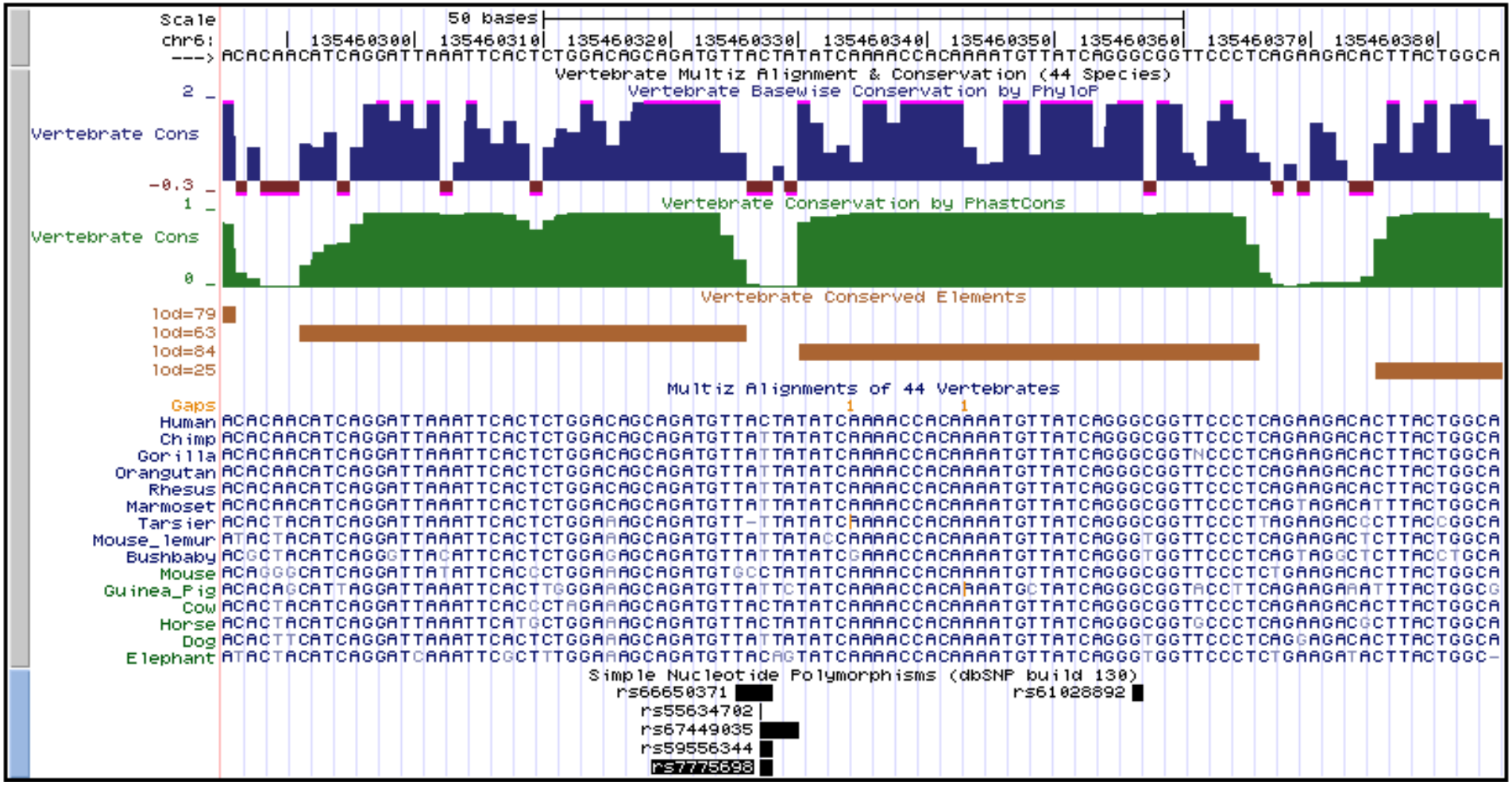
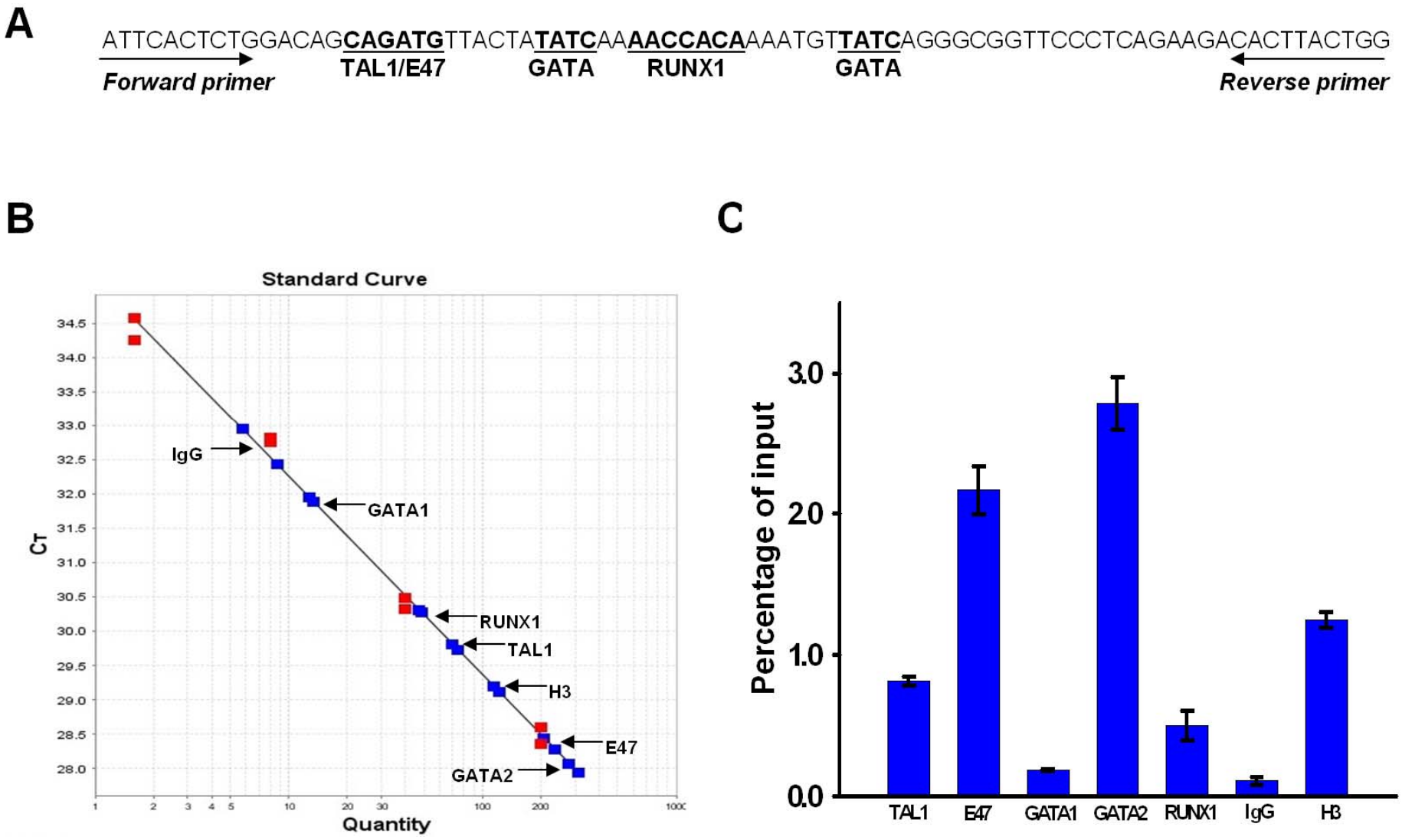

Scale
50 bases
chr6:
135460300 135460310 135460320 135460330 135460340 135460350 135460360 135460370 135460380
ACACAACATCAGGATTAAATTCACTCTGGACAGCAGATGTTACTATATCAAAACCACAAAATGTTATCAGGGCGGTTCCCTCAGAAGACACTTACTGGCA
Vertebrate Multiz Alignment & Conservation (44 Species)
Vertebrate Basewise Conservation by PhyloP
Vertebrate Cons
2
-0.3
Vertebrate Conservation by PhastCons
Vertebrate Cons
1
0
Vertebrate Conserved Elements
lod=79
lod=63
lod=84
lod=25
Multiz Alignments of 44 Vertebrates
Gaps
Human
Chimp
Gorilla
Orangutan
Rhesus
Marmoset
Tarsier
Mouse_lemur
Bushbaby
Mouse
Guinea_Pig
Cow
Horse
Dog
Elephant
Simple Nucleotide Polymorphisms (dbSNP build 130)
rs66650371
rs61028892
rs55634702
rs67449035
rs59556344
rs7775698

Figure S3


A
ATTCACTCTGGACAGCAGATGTTACTATATCAAAACCACAAAATGTTATCAGGGCGGTTCCCTCAGAAGACACTTACTGG
TAL1/E47
GATA
RUNX1
GATA
Forward primer
Reverse primer
B
Standard Curve
Ct
Quantity
IgG
GATA1
RUNX1
TAL1
H3
E47
GATA2
C
Percentage of input
TAL1
E47
GATA1
GATA2
RUNX1
IgG
H3