GATA1 erythroid-specific regulation of SEC23B expression and its implication in the pathogenesis of congenital dyserythropoietic anemia type II

Roberta Russo,[*1,2] *Immacolata Andolfo,*[*1,2] *Antonella Gambale,*[1,2] *Gianluca De Rosa,*[1,2] *Francesco Manna,*[1,2] *Alessandra Arillo,*[2] *Farooq Wandroo,*[3] *Maria Grazia Bisconte*[4] *and Achille Iolascon*[1,2]

***RR and IA contributed equally to this work.*

[1]Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Italy; [2]CEINGE Biotecnologie Avanzate, Napoli, Italy; [3]Department of Haematology Sandwell and West Birmingham Hospital, NHS trust West Midlands UK and [4]Dipartimento Oncoematologico, Azienda Ospedaliera di Cosenza, Italy

Correspondence: roberta.russo@unina.it
doi:10.3324/haematol.2016.162966

SUPPLEMENTARY DATA

Article title: GATA1 erythroid-specific regulation of SEC23B expression and its implication in the pathogenesis of Congenital Dyserythropoietic Anemia type II

Authors: Roberta Russo*, Immacolata Andolfo*, Antonella Gambale, Gianluca De Rosa, Francesco Manna, Alessandra Arillo, Farooq Wandroo, Maria Grazia Bisconte, Achille Iolascon

Corresponding author:

Roberta Russo, PhD - CEINGE – Biotecnologie Avanzate, Via Gaetano Salvatore 486, 80145 Naples (Italy) - email: roberta.russo@unina.it – Phone: +39 081 3737 736

Table of contents:

METHODS

Patients and genomic mutational screening

Diagnosis was based on history, clinical findings, laboratory data, and morphological analysis of aspirated bone marrow (BM). Whenever possible the hypoglycosylation of band 3 was evaluated by SDS-PAGE. Samples were obtained after informed consent for the studies, according to the Declaration of Helsinki. The study has been approved by local ethical committee of University Federico II, Napoli.

Genomic DNA preparation, mutational search, oligonucleotide primers design and direct sequencing were performed as previously described.[1] Sequence primers are available on request (roberta.russo@unina.it). Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of ATG translation initiation codon in the reference sequence (SEC23B Ensembl transcript ID: ENST00000377475; GATA1 Ensembl transcript ID: ENST00000376670), according to the nomenclature for the description of sequence variants of Human Genome Variation Society (www.hgvs.org/mutnomen). The initiation codon is codon 1. Prediction analysis for splice sites mutations was performed by web server tool Human Splicing Finder (http://www.umd.be/HSF/), as previously described.[2]

Gene and protein expression analyses

RNA isolation and reverse transcription. Total RNA was extracted from peripheral blood leukocytes (PBLs) and cell lines using Trizol reagent (Life Technologies). Synthesis of cDNA from total RNA (1 μg) was performed using cDNA synthesis kit (Applied Biosystems, Milan, Italy).

Quantitative real-time PCR analysis. Quantitative RT-PCR (qRT-PCR) using Power SYBR Green PCR Master Mix (Applied Biosystems) was performed to evaluate the gene expression of *GATA1* and *SEC23B* genes. Samples were amplified on Applied Biosystems 7900HT Sequence Detection System using standard cycling conditions. The primers were designed by the Primer Express 2.1 program (Applied Biosystems). B-actin was used as internal control. Relative gene expression was calculated by using the $2^{-\Delta Ct}$ method, while the mean fold change = $2^{-(\text{average } \Delta\Delta Ct)}$ was assessed using the mean difference in the ΔCt between the gene and the internal control.[2] Primer sequences are available upon request (roberta.russo@unina.it).

Protein isolation and western blotting analysis. Protein extraction from PBMCs and WB were performed as previously described.[3] Particularly, 30 μg of total extract proteins was loaded into each lane and was separated by 10% SDS PAGE bisacrylamide gel, followed by transfer to PVDF membranes (Biorad, Milan, Italy). A specific rabbit anti-SEC23B antibody (1:500) (BioLegend,

San Diego, CA) and anti-GATA1 antibody (1:500) (ab11963 – Abcam, Cambridge, UK) were used. Mouse anti-β-actin antibody (1:5000) (Sigma-Aldrich, Milan, Italy) was used as the control for equal loading.

Semi-quantitative analysis of protein expression was performed as previously described.[2] The bands were quantified by Quantity One software (Biorad) to obtain an integral optical density (OD) value, which then was normalized with respect to the β-actin value.

Vector cloning

In silico analysis of human *SEC23B* (HuSEC23B) upstream region (chr20:18487188-18491479: 4292-bp from ATG) was performed by ENCODE web tool implemented in the UCSC Genome Browser (http://genome.ucsc.edu/). Prediction analyses for CpG islands were performed by CpG Islands Track (UCSC Genome Browser) and by EMBOSS Cpgplot, at EMBL-EBI website (http://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/). The 5'upstream region of *SEC23B* gene (3438-bp from ATG) has been divided into 10 overlapping fragments (HuSEC23B/3.44, HuSEC23B/2.52, HuSEC23B/2.14, HuSEC23B/1.75, HuSEC23B/1.28, HuSEC23B/0.99, HuSEC23B/0.81, HuSEC23B/0.61, HuSEC23B/0.45, HuSEC23B/0.21), amplified from genomic DNA. Each construct was purified with QIAquick PCR Purification Kit (Qiagen, Milan, Italy), cloned upstream the luciferase gene into PGL3 vector (Life Technologies) in the Hind III and Xho I sites. Two-primer site-direct PCR mutagenesis was used to mutagenize the GATA1 binding sites in the two fragments HuSEC23B-2475 and HuSEC23B-463. The amplified region and ligation boundaries were verified by direct sequencing.

GATA1 cDNA product (BC009797_Clone ID 4048082, Invitrogen) was amplified from pDNR-LIB vector, purified with QIAquick PCR Purification Kit (Qiagen) and cloned into pcDNA3.1 vector (Invitrogen, Milan, Italy) in the EcoRV and XhoI sites.

All plasmids were transformed into the DH5α bacterial strain (Invitrogen) and subsequently purified with the QIAprep Miniprep Kit (Qiagen). Direct sequencing was performed using the BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems) and a 3730 DNA Analyzer (Applied Biosystems) with specific primers. Two-primer site-direct PCR mutagenesis was used to introduce GATA1-associated missense mutations G208R, R216W, D218G, and V205M. The amplified region and ligation boundaries were verified by direct sequencing.

Cell cultures and transfection

HEK-293 and K562 cells (ATCC, Manassas, VA, USA) were maintained according to the manufacturer's instructions. PGL3-HuSEC23B plasmids were transfected (2 μg/well) using the X-

tremeGENE HP DNA Transfection Reagent (Roche, Milan, Italy) according to the manufacturer's procedures. The cells were collected 48h after the transfection to luciferase assays. Co-transfection of GATA1bs/HuSEC23B-2475 and -463 mutants and GATA1-WT (1 μg for each plasmid) was performed in HEK-293 cell line as above described. Similarly, co-transfection of GATA1 mutants (GATA1-G208R, GATA1-R216W, GATA1-D218G, and GATA1-V205M) and the HuSEC23B/3.44 was also performed in HEK-293 cell line, as aforementioned. As control of transfection efficiency we evaluated the expression of *Neomycin* resistance gene (expressed by pCDNA3.1vector) in HEK293 cells transfected with pCDNA3.1-EV, pCDNA3.1-GATA1 and pCDNA3.1-GATA1-G208R.

Erythroid differentiation of K562 and HEL cells

Fifty μM hemin (Sigma-Aldrich) was added to the culture medium of the wild-type K562 and HEL cells ($2x10^5$/mL). Samples were collected at days 6 after hemin addition. Erythroid differentiation was assessed by FACS analysis for transferrin receptor 1 (CD71) and glycophorin A (CD235A) as previously described.[3]

Chromatin immunoprecipitation assay

GATA1 binding sites (GATA1bs) in the upstream region of *SEC23B* gene were predicted by the software tools MatInspector from Genomatix web server and by PROMO at the ALGGEN server. Chromatin immunoprecipitation (ChIP) studies were performed by Immunoprecipitation Kit - Dynabeads, following the manufacturer's instruction (Life Technologies). Isolated chromatin from K562 and HEL cells at 6 days of erythroid differentiation was immunoprecipitated using control immunoglobulin G or 5 μg rabbit polyclonal GATA1 antibody (ab11963 - Abcam). Immunoprecipitated DNA (10 μl) was used for the ChIP-quantitative PCR (ChIP-qPCR) analysis, to detect the presence of specific DNA segments. We used ETO [4] as a positive control and a non-binder GATA1-sequence of the HuSEC23B promoter as a negative control.

Promoter assay

Luciferase assays with HuSEC23B deletion mutants were performed in the HEK-293 and K562 cell lines. Cells were grown according to the manufacturer's protocol. For the promoter assay, HEK-293 and K562 cells were plated in 6-well plates and transfected with the HuSEC23B deletion mutants using Extreme gene (Roche). Transfected cells were harvested after 48 hrs. The PRL-CMV vector (100 ng) was used for normalization. Luciferase activities were analysed using a Dual Luciferase Reporter Assay system (Promega).

References

1. Russo R, Esposito MR, Asci R, et al. Mutational spectrum in congenital dyserythropoietic anemia type II: identification of 19 novel variants in SEC23B gene. Am J Hematol. 2010;85(12):915-920.

2. Russo R, Langella C, Esposito MR, et al. Hypomorphic mutations of SEC23B gene account for mild phenotypes of congenital dyserythropoietic anemia type II. Blood Cells Mol Dis. 2013;51(1):17-21.

3. Andolfo I, De Falco L, Asci R, et al. Regulation of divalent metal transporter 1 (DMT1) non-IRE isoform by the microRNA Let-7d in erythroid cells. Haematologica. 2010;95(8):1244-1252.

4. Ajore R, Dhanda RS, Gullberg U, Olsson I. The leukemia associated ETO nuclear repressor gene is regulated by the GATA-1 transcription factor in erythroid/megakaryocytic cells. BMC Mol Biol. 2010;11:38.

Table S1. Comparison of clinical phenotype between Case B-II.2 and CDAII biallelic patients

	Case 2 B-II.2	CDAII biallelic cases[§]	Reference values
Age at diagnosis	48 yrs	19.1 ± 3.0 (36)	-
Onset symptoms	40 yrs	3.7 ± 0.6 (79)	-
Complete blood count			
RBC (10^6/µL)	3.1	3.2 ± 0.1 (83)	4.2 – 5.6
Hb (g/dL)	10	9.6 ± 0.2 (143)	12.0 – 16.0
Ht (%)	25	28.6 ± 0.7 (84)	35.0 – 48.0
MCV (fL)	86	88.1 ± 1.0 (88)	80 – 97
Retics %	3.3	3.0 ± 0.2 (100)	0.5 – 2.0
Retics abs count (10^3/µL)	102	89.1 ± 5.2 (103)	-
PLT (10^3/µL)	198	381.3 ± 29.7 (56)	130.0 – 400.0
Laboratory data and iron balance			
Total bilirubin (mg/dL)	8.2	3.0 ± 0.5 (35)	0.2 – 1.1
Unconjugated bilirubin (mg/dL)	7.5	2.6 ± 0.5 (29)	0.2 – 0.8
Ferritin (ng/mL)	1157	275.9 ± 61.0 (36)	20 – 200
Transferrin saturation (%)	85	71.4 ± 5.5 (25)	14 – 45

[§]Data are presented as average ± SEM (n).

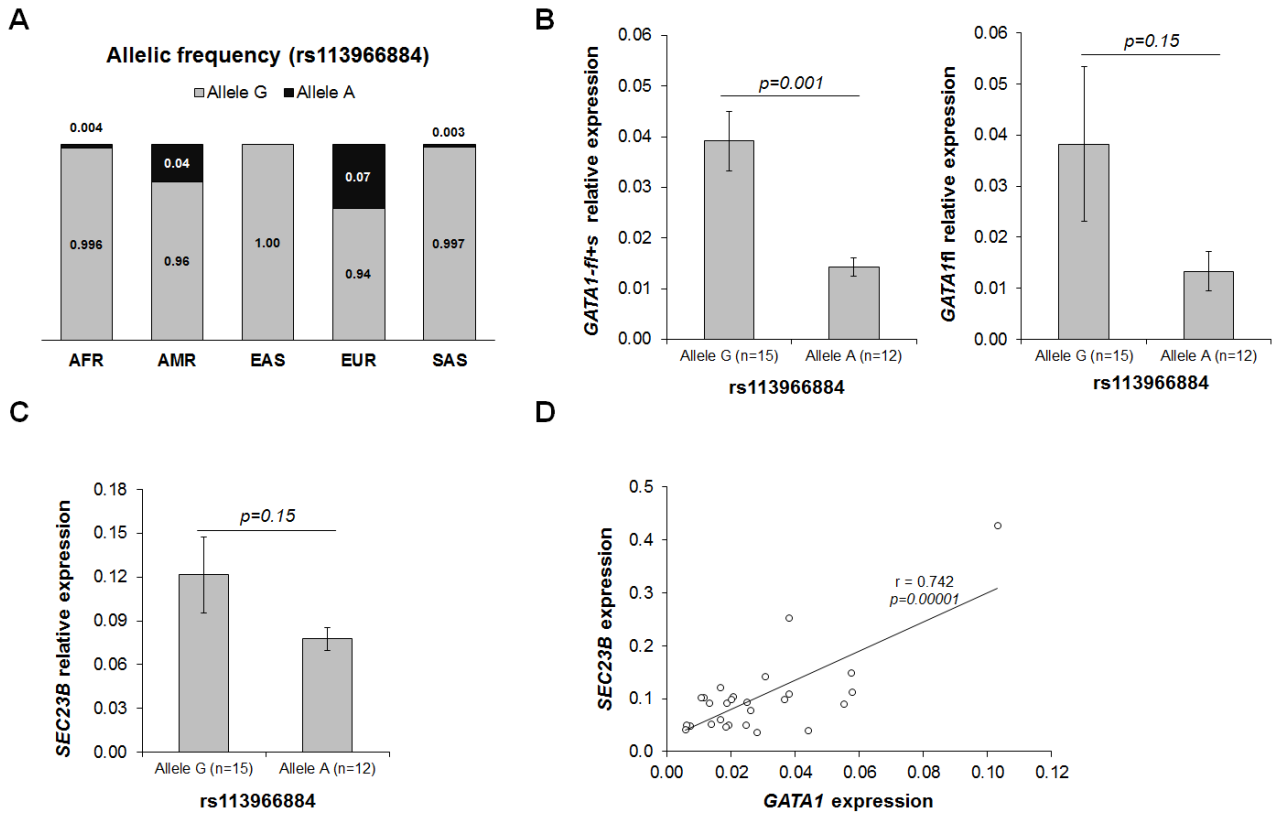

Figure S1. Genetic and functional description of GATA1-rs113966884 and correlation analysis between SEC23B and GATA1

(A) Absolute allele frequencies of GATA1-rs113966884 single nucleotide variant in control populations from 1000 Genomes project (http://www.1000genomes.org/; AFR, African; AMR, Ad Mixed American; EAS, East Asian; EUR, European; SAS, South Asian). (B) Gene relative expression of *GATA1* full length and short isoforms (*GATA1fl+s*) and *GATA1fl* transcripts respect to the reference gene, *β-actin*, in 27 healthy subjects stratified according to the genotype of GATA1-rs113966884 SNV is shown. Subjects with genotypes AG/AA/A0 showed a reduced expression of both GATA1 isoforms when compared with genotypes GG/G0 (GATA1fl+s: allele G, 0.04±0.006; allele A, 0.014±0.002; GATA1fl: allele G, 0.04±0.015; allele A, 0.013±0.004). Data are presented as mean ± SE. P value has been calculated by Student t test. (C) Gene relative expression of *SEC23B* respect to the reference gene, β-actin, in 27 healthy subjects stratified according to the genotype of GATA1-rs113966884 SNV is shown. Subjects with genotypes AG/AA/A0 showed a reduced expression of both GATA1 isoforms when compared with genotypes GG/G0 (SEC23B: allele G, 0.12±0.03; allele A, 0.08±0.008). Data are presented as mean ± SE. P value has been calculated by Student t test. (D) Correlation analysis between *SEC23B* and *GATA1* gene expression in PBL from 27 healthy controls. P value by Pearson correlation.

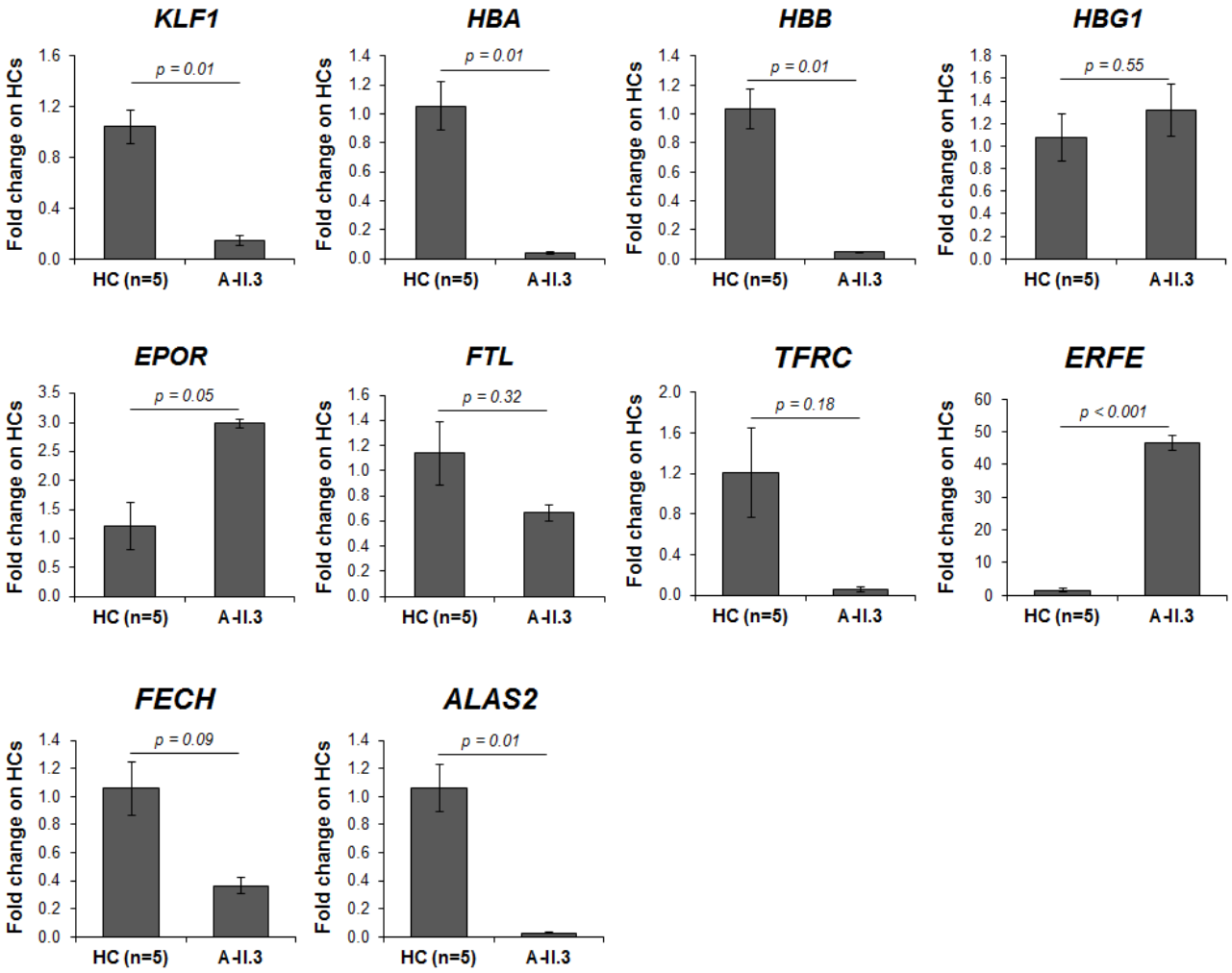

Figure S2. Gene expression profile of GATA1-regulated genes in GATA1-G208R patient

Gene relative expression of GATA1-dependent genes in PBL from A-II.3 proband and 5 healthy controls (HC). *KLF1*, Kruppel Like Factor 1; *HBA*, hemoglobin alpha; *HBB*, hemoglobin beta; *HBG1*, hemoglobin gamma 1; *EPOR*, erythropoietin receptor; *FTL*, ferritin light chain; *TFRC*, transferrin receptor; *ERFE*, erythroferrone; *FECH*, ferrochelatase; *ALAS2*, 5'-aminolevulinate synthase 2. As effect of ineffective erythropoiesis, the genes encoding for erythropoietin receptor (*EPOR*) and erythroferrone (*ERFE*) resulted marked up-regulated in the patient compared to healthy subjects. Data are presented as mean ± SE. P value has been calculated by Student t test.

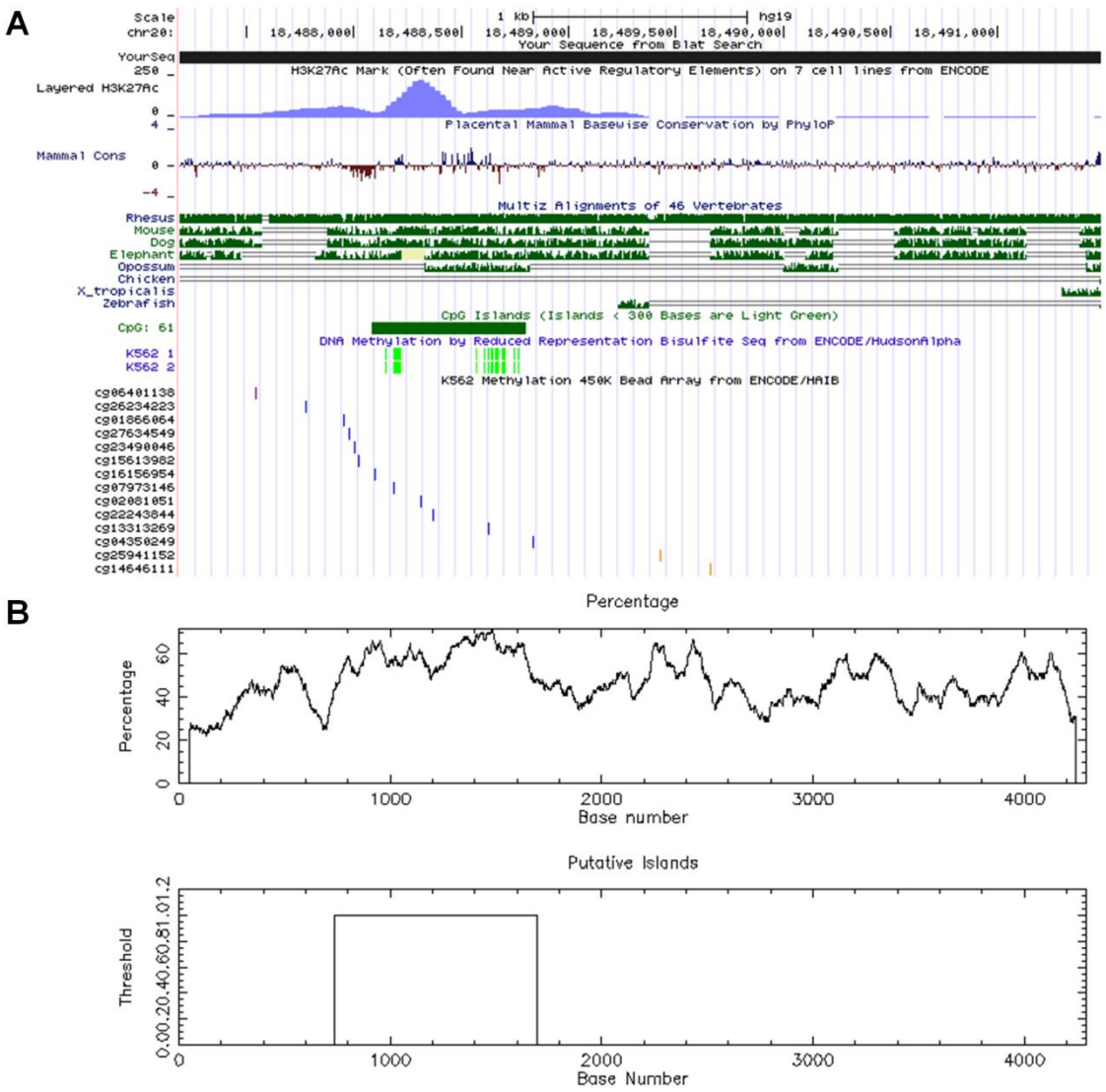

Figure S3. ***In silico*** **analysis of genomic sequence upstream HuSEC23B**

(A) *In silico* analysis of HuSEC23B upstream region is shown. The CpG Islands Track from UCSC Genome Browser (ENCODE web tool) predicts the presence of a 720 bp CpG island within the HuSEC23B/3.29-2.78 region. The track "DNA Methylation by Reduced Representation Bisulfite Seq from ENCODE/HudsonAlpha" from UCSC Genome Browser (ENCODE web tool) reports the percentage of DNA molecules that exhibit cytosine methylation at specific CpG dinucleotides. The 720 bp CpG island resulted unmethylated in K562 cell line (green = 0% of molecules sequenced are methylated). The track "CpG Methylation by Methyl 450K Bead Arrays from ENCODE/HAIB" displays the methylation status of specific CpG dinucleotides in K562 cells. Methylation status is color-coded as: orange = methylated; purple = partially methylated; bright blue = unmethylated. (B) Identification of the CpG island by EMBOSS Cpgplot at EMBL-EBI website (http://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/).

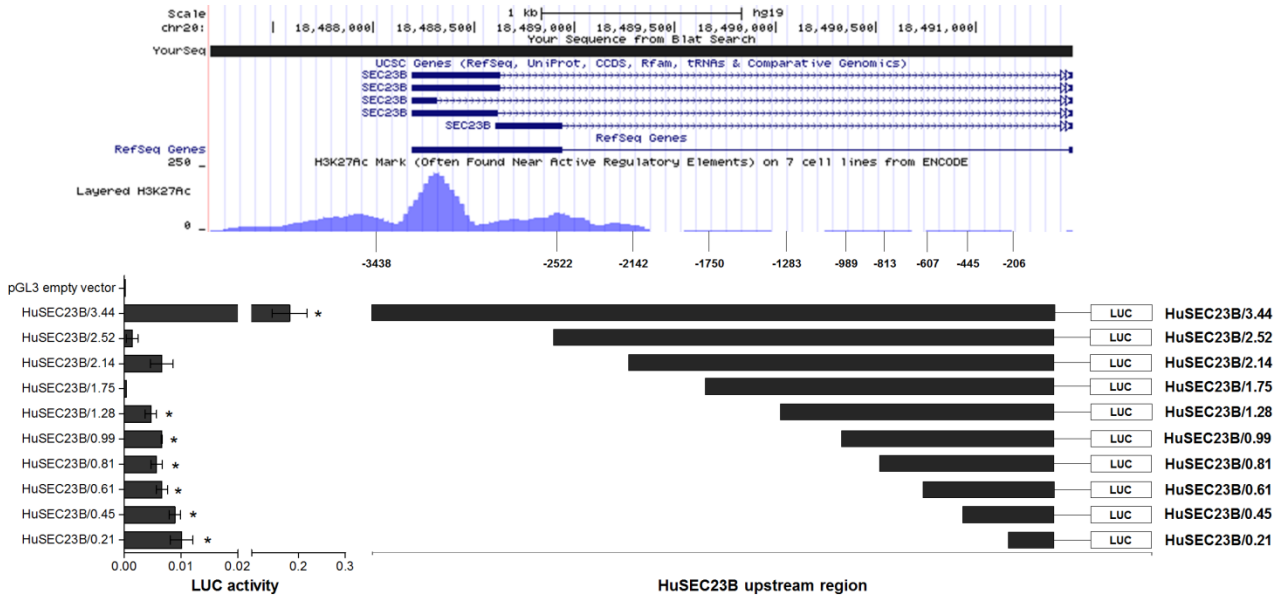

Figure S4. *In silico* and functional analyses of HuSEC23B upstream region

Above, *in silico* analysis of HuSEC23B upstream region is shown. The blue track shows the levels of enrichment of the H3K27Ac histone mark across the genome as determined by a ChIP-seq assay in K562 cell line from UCSC Genome Browser (ENCODE web tool). The H3K27Ac histone mark is the acetylation of lysine 27 of the H3 histone protein, and it is thought to enhance transcription possibly by blocking the spread of the repressive histone mark H3K27Me3.

Below, a schematic representation of 10 deletion mutants of HuSEC23B upstream region is shown (on the right). Luciferase normalized activity (LUC activity) for each construct is shown in the chart on the left. Only the larger fragment, HuSEC23B/3.44, containing the enhancer region HuSEC23B/3.29-2.78, showed a marked luciferase activity compared to the empty vector (fold = 933.8). Conversely, the fragments lacking the H3K27Ac enriched sequence exhibited a reduced luciferase expression compared to HuSEC23B/3.44. Particularly, the fragments HuSEC23B/2.52, HuSEC23B/2.14 and HuSEC23B/1.75 showed a non-statistically significant luciferase transactivation compared to empty vector. This is in agreement with the presence of the CpG sites that resulted methylated in K562 cells (see Figure S3). However, the smallest fragment (HuSEC23B/0.21) was still able to drive luciferase expression. As expected, this region is predicted to contain the vertebrate TATA motif. Data are presented as mean ± SD. P value by Student t test.

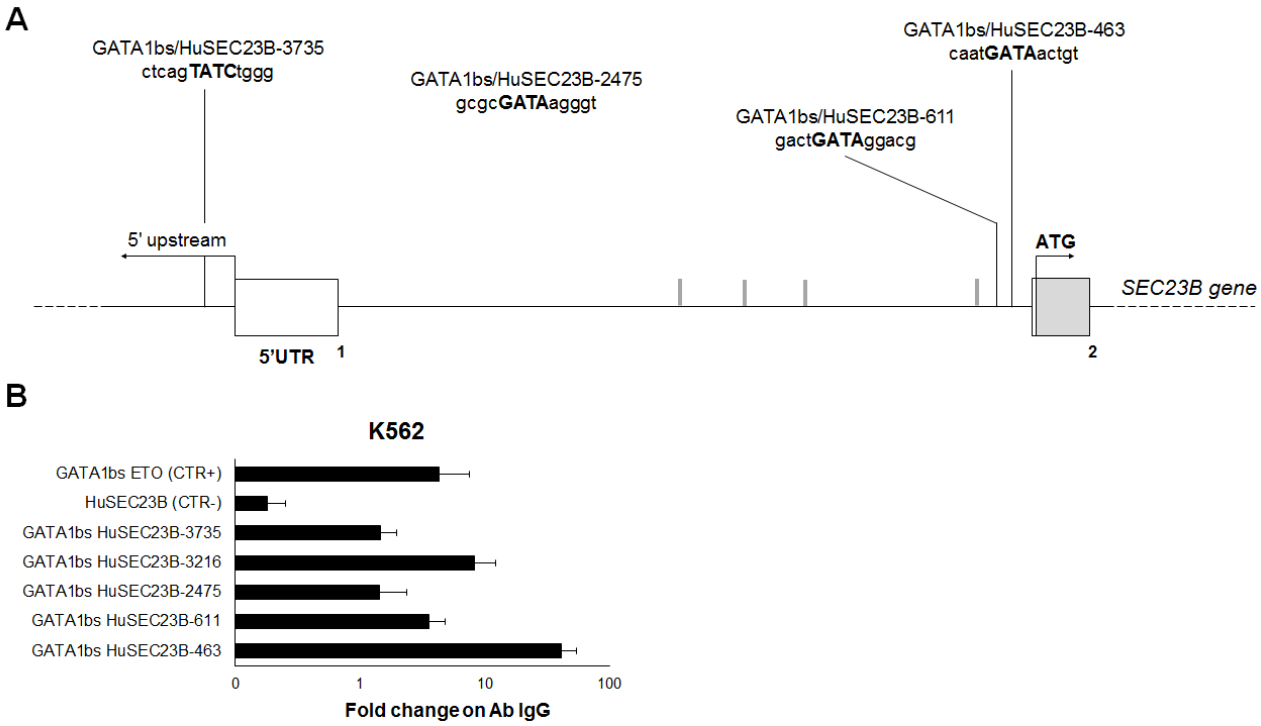

Figure S5. **GATA1 directly binds to *cis*-acting regulatory sequences of HuSEC23B upstream region**

(A) *In silico* analysis of HuSEC23B upstream region showed the presence of putative binding sites of the transcription factor GATA1. The binding sites predicted by both *in silico* tools (MatInspector and PROMO) were shown; light gray bar indicate the sites predicted by only one tool. We selected sense and antisense binding sites predicted by both tools for further analyses. The *cis*-acting regulatory elements not containing the consensus sequence W(A/T)GATAR(A/G) were excluded. (B) ChIP-qPCR analysis on GATA1 immunoprecipitated DNA in K562 cells at 6 days of differentiation by hemin. GATA1bs/ETO is the CTR+. HuSEC23B (CTR-) is a sequence flanking GATA1bs within HuSEC23B upstream region. Data are presented as mean ± SD.