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

Biallelic mutations in *SEC23B* gene cause congenital dyserythropoietic anemia type II (CDAII), an autosomal recessive disorder characterized by ineffective erythropoiesis, hemolytic anemia, and splenomegaly.[1] Approximately 80 *SEC23B* mutations have been described so far.[1-3] However, several cases show an incomplete inheritance pattern or no mutations within the gene. Mutations in cis-acting regulatory elements of *SEC23B* as well as digenic inheritance was hypothesized as additional genetic etiology of CDAII.[2] CDA variants have been associated to mutations in *GATA1*, encoding for the homonymous DNA binding protein that plays an essential role in development and maintenance of both megakaryocytic and erythroid lineages.[4] GATA1-related cytopenias are hallmarked by thrombocytopenia that can be associated to impaired erythropoiesis, dyserythropoietic anemia, congenital erythropoietic porphyria (CEP), thalassemia,[5] or Diamond Blackfan anemia-like disease.[6] The pathogenesis of these conditions is mostly related to altered GATA1 protein levels.[7,8]

In this study we described the functional interaction between *GATA1* and *SEC23B* genes in two patients with suspected CDAII (Table 1). See methods in *Online Supplementary Data.*

Case1 (A-II.3) was a 22-year-old Italian male from healthy parents (Figure 1A). At birth, respiratory distress, haemorrhagic manifestations, and hypospadias were observed. Family history was not indicative for anemia, but both brothers were monorchid. At 4 months, anemia with thrombocytopenia was observed, and blood and platelets (PLT) transfusion regimen was started. Bone marrow (BM) analysis revealed dyserythropoiesis with bi- and multi-nucleated erythroblasts, while electron microscopy (EM) highlighted erythroblasts with double membranes, a typical feature of CDAII. Additionally, ultrastructural abnormalities in megakaryocytes, such as fairly normal granulation pattern, lack of demarcation membranes, and presence of cytoplasmic vacuoles were observed. The clinical data affecting both erythroid and megakaryocytic lines prompted us to perform sequencing analysis of both *GATA1* and *SEC23B*. We identified the causative mutation of X-linked thrombocytopenia with dyserythropoietic anemia (XLTDA) *GATA1*-p.G208R,[9] but no causative mutations in *SEC23B* gene.

Case2 (B-II.2) was a 48-year-old Punjabi male from healthy parents (Figure 1B). Family history was not indicative of anemia. Overall clinical features mostly resembled those of CDAII biallelic patients, except for a slight reduction of PLT count (Table 1 and *Online Supplementary S1*). He showed jaundice, gallstones and liver iron overload with high transferrin saturation, low hepcidin/ferritin ratio and increased level of erythroferrone.[10] Consequently to iron loading, he developed diabetes and complained of impotence. Mutations neither in *HFE* nor in *HBA* and *HBB* genes were detected. Deficit of pyruvate kinase and G6PD, porphyria, sideroblastic anemia and hepatitis were also excluded. BM examination highlighted hypercellularity with severe dyserythropoiesis, multinucleated erythroblasts, and internuclear bridges. Accordingly, mutations in *CDAN1* and *C15ORF41* genes were ruled out. Additionally, no mutations in CDAIV causative gene, *KLF1*, were found. Pathognomonic features of CDAII were observed: EM analysis showed binucleated erythroblasts with double membrane, and SDS-PAGE displayed hypoglycosylated band 3. Mutational analysis in this proband revealed an intronic *SEC23B* mutation, c.1905+3G>T, in heterozygous state. This variant was neither annotated in 1000 Genomes nor in ExAC browser. Both *in silico* and *ex vivo* analyses demonstrated that it impairs splicing, leading to the skipping of exon 16 in *SEC23B* mRNA with subsequent mRNA decay of the mutated allele (Figure 1C). We excluded additional mutations in both coding and splice site regions of *SEC23B* as well as large gene deletions or mutations in the promoter. Thus, the mutational pattern in this patient resulted incomplete. Sequence analysis of *GATA1* showed the nucleotide transition c.-183G>A in the 5' upstream region. This is a common variant (rs113966884), annotated in 1000 Genomes project with a minor allele frequency 0.02 (*Online Supplementary Figure S1A*), affecting the expression of both *GATA1* isoforms (full length, fl and short, s), leading to a reduced expression of both transcripts in healthy controls (HCs) carrying rs113966884 A allele compared to G allele (*Online Supplementary Figure S1B*). Interestingly, we also observed a reduced SEC23B expression in control subjects with GATA1-rs113966884 A allele compared to G allele (*Online Supplementary Figure S1C*).

Expression analysis on peripheral blood leukocytes (PBL) from the 2 probands showed that both *GATA1* mutations, p.G208R and c.-183G>A, account for a reduced expression of *GATA1* itself at both mRNA and protein levels (Figure 1 D-E-F). Of note, in the *GATA1*-c.-183G>A patient we observed a slight reduction of *GATA1* gene expression, despite the marked downregulation of GATA1 protein. This apparent discrepancy can be explained by the GATA1 positive feedback loop. Indeed,

Table 1. Clinical data.

	Case 1	Case 2	Reference values
Age at diagnosis	22 yrs	48 yrs	–
Onset symptoms	4 mo	40 yrs	–
Complete blood count			
RBC (10^6/μL)	3.3	3.1	4.2-5.6
Hb (g/dL)	9.8	10.0	12.0-6.0
Ht (%)	33.0	25.0	35.0-48.0
MCV (fL)	100	86	80-97
MCH (pg)	29.0	–	25.0-34.0
MCHC (g/dL)	29.0	–	32.0-38.0
Retics %	2.6	3.3	0.5-2.0
Retics abs count (10^3/μL)	85.8	102.0	–
PLT (103/μL)	25.0	198.0	130.0 – 400.0
Laboratory data and iron balance			
Total bilirubin (mg/dL)	3.3	8.2	0.2-1.1
Unconjugated bilirubin (mg/dL)	2.6	7.5	0.2-0.8
LDH (U/L)	1121	171	227-450
Ferritin (ng/mL)	313	1157	20-00
Transferrin saturation (%)	25	85	14-45
Hepcidin (nM)	1.3	1.6	–
Hepcidin/ferritin	0.004	0.001	–

RBC: red blood cell count; Hb: hemoglobin; Ht: hematocrit; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; Retics: reticulocyte count; PLT: platelet count; LDH: lactic dehydrogenase.

the upstream variant leads to a quantitative defect of GATA1 itself but not to a qualitative one, as in the case of missense p.G208R. Similarly to those observed in HCs carrying *GATA1*-rs113966884 A allele, in case B-II.2 we did not detect unbalanced expression of the two GATA1 isoforms; moreover, no abnormal megakaryopoiesis was observed, supporting the observation that GATA1s expression promotes mostly megakaryopoiesis, but not erythropoiesis.[11]

Although sequencing analysis revealed no mutations in *SEC23B* coding sequence in case A-II.3, expression analysis of this gene showed the same reduced trend of *GATA1* (Figure 1D), with comparable results at protein evaluation (Figure 1 E-F). According to the *ex vivo* data, we observed a reduced *GATA1* and *SEC23B* expression in HEK-293 cells transfected with GATA1-G208R mutant compared to cells transfected with GATA1 wild type (WT) (Figure 2 A-B-C). As expected, the reduced expression of GATA1-G208R in the proband A-II.3 also affected several other GATA1-regulated genes, similarly to *SEC23B* (*Online Supplementary Figure S2*). Thus, the phenotype of this patient could result from altered expression of many erythroid genes, although the specific abnormality of double membranes in erythroblasts could arise form altered *SEC23B* expression.

Also in case B-II.2, a reduced expression of SEC23B at both mRNA and protein levels was observed when compared to HC (Figure 1 D-E-F). Additionally, a correlation analysis on PBL from 27 HC revealed a significant direct correlation between *GATA1* and *SEC23B* expression (P=0.00001) (*Online Supplementary Figure S1D*). Thus, we postulated a functional interaction between these two genes as a novel genetic basis of CDAII.

In order to corroborate our hypothesis, we characterized 4292-bp 5' upstream sequence of human *SEC23B* (HuSEC23B) using a twofold approach: both *in silico* and *in vitro* data demonstrated that this genomic region is highly accessible to transcription in K562 cell line. *In silico* evaluation by ENCODE web-tool identified a 500-bp conserved region, named HuSEC23B/3.29-2.78 (3300-bp upstream ATG), which exhibited an enrichment of the H3K27Ac histone mark in K562 cells, suggesting a high accessibility of the chromatin to transcription in this locus. This genomic sequence is predicted to contain a 720-bp CpG island, partially unmethylated in K562 cells (*Online Supplementary Figure S3*). We tested 10 deletion mutants of HuSEC23B upstream region for the ability to drive luciferase expression by transient transfection of K562 cells. According to *in silico* data, only the larger fragment, containing the enhancer region HuSEC23B/3.29-2.78, showed a marked luciferase activity compared to the empty vector (*Online supplementary Figure S4*).

We demonstrated the direct binding of GATA1 to 5 predicted cis-elements (GATA1bs) within HuSEC23B upstream region by chromatin immunoprecipitation (ChIP) assay in both K562 and HEL cells at 6 days of erythroid differentiation (Figures 2D and *Online Supplementary Figure S5*), confirming the ChIPseq data showing GATA1 occupancy at *SEC23B* in primary human erythroid cells.[12] Particularly, two sites, GATA1bs/HuSEC23B-2475 and GATA1bs/HuSEC23B-463, showed high relative occupancy of the immunoprecipitated factor (Figure 2D). Of note, when we transiently co-transfected GATA1-WT with both GATA1 binding sites mutated in the core GATA1 binding sequence, we observed a reduction of approximately 30-40% of

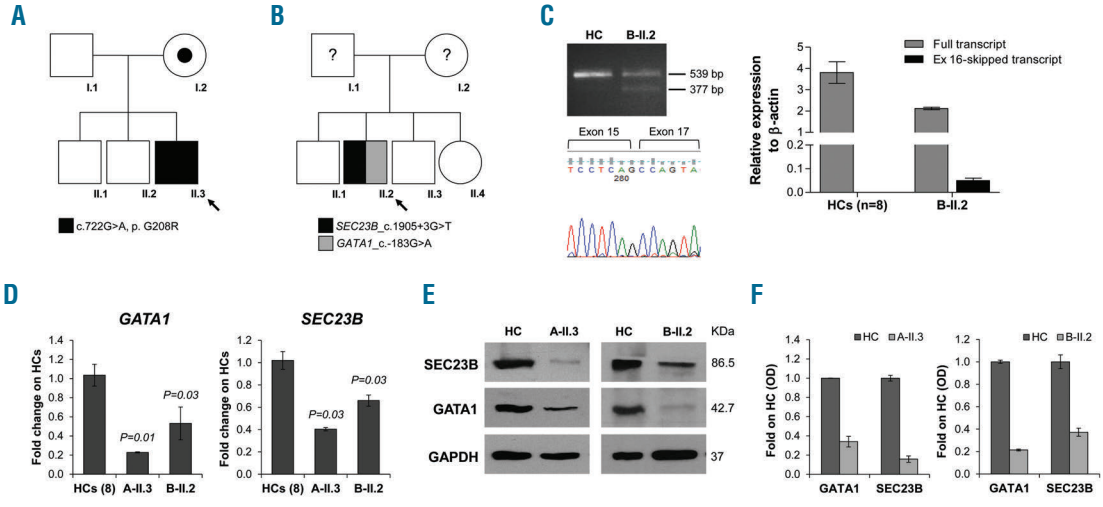

Figure 1. Genetic study and *ex vivo* analysis of GATA1 and SEC23B expression. (A) The proband A-II.3 was hemizygous for the GATA1 mutation c.722G>A, p.G208R; according to X-linked inheritance pattern, the mutation was inherited from the mother. (B) Proband B-II.2 showed digenic inheritance pattern, with *SEC23B* intronic mutation, c.1905+3G>T, and GATA1 c.-183G>A polymorphic variant. DNA from relatives was not available. (C) *In silico* analysis of c.1905+3G>T mutation predicted a slight reduction of the score between WT and mutated donor site sequence (WT score = 95.15, mut score = 91.28, variation = -4.06%). Accordingly, amplification by PCR of the exon region 13-17 of SEC23B cDNA from PBL of the patient B-II.2 highlighted the presence of two bands on agarose gel: the top band (expected size fragment) of 539-bp, and the bottom band in lane 2 of 377-bp (exon-16 skipped transcript). Sequencing analysis of the 377-bp PCR product is shown. The chart shows the expression analysis of the exon-16-skipped transcript compared to the full length transcript by qRT-PCR analysis. In agreement with *in silico* prediction, the exon-16-skipped transcript is produced at very low level compared to full transcript (2.6%) HC(s), healthy control(s). (D) GATA1 and SEC23B expression in both patients compared to HC (n = 8) is shown. We observed a marked downregulation (80%) of GATA1-G208R mRNA expression in the proband A-II.3 compared to HC (P=0.01), and a slight reduction (50%) of GATA1-c.-183G>A gene expression in the proband B-II.2 compared to HC (P=0.03). Also SEC23B gene expression resulted down-regulated in both probands compared to HC (P=0.03). Data are presented as mean ± SE. P value by Student t-test. (E) WB analysis of GATA1 and SEC23B in PBL from both patients compared to a HC. Sizes (in kDa) are on the right. (F) The charts show the densitometric quantification of WB presented in the panel E. A marked reduction of GATA1 level in both probands A-II.3 (70%) and B-II.2 (80%) compared to HC was observed. SEC23B protein levels similarly decreased in both patients. OD, optical density. Data are presented as mean ± SD.

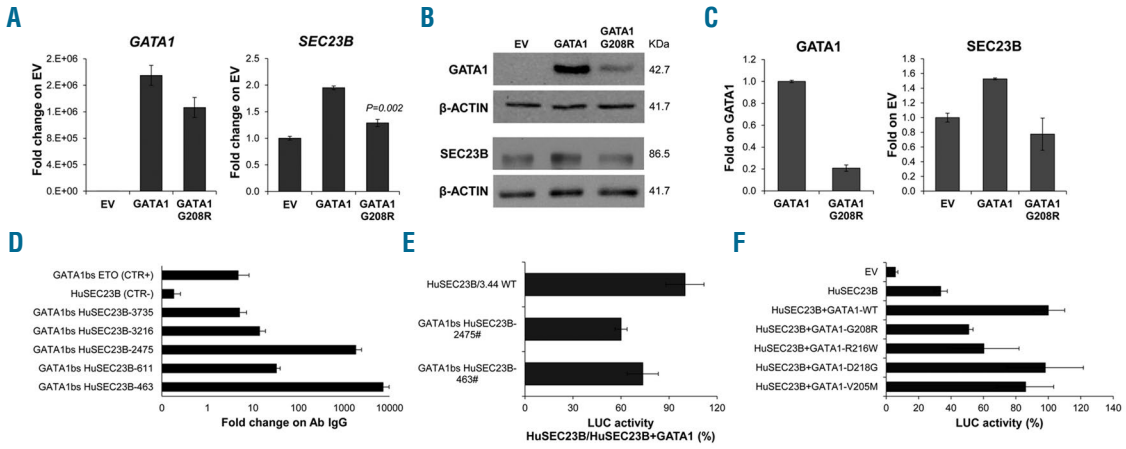

Figure 2. Functional interaction between *GATA1* and *SEC23B* genes. (A) *GATA1* and *SEC23B* expression of HEK-293 cells overexpressing pCDNA3.1-GATA1 and pCDNA3.1-GATA1-G208R compared to those transfected with empty vector (EV) were shown. Data are presented as mean ± SD. *P* value by Student *t*-test. (B) WB analysis of GATA1 and SEC23B in HEK-293 cells transfected with pCDNA3.1-EV, pCDNA3.1-GATA1 and pCDNA3.1-GATA1-G208R. Sizes (in kDa) are on the right. (C) Densitometric quantification of WB presented in the panel B; OD, optical density. Data are presented as mean ± SD. (D) ChIP-quantitative PCR analysis on GATA1 immunoprecipitated DNA in HEL at 6 days of differentiation by hemin. GATA1bs/ETO is the positive control (CTR+). HuSEC23B is the negative control (CTR-), a sequence flanking GATA1bs within HuSEC23B upstream region. Data are presented as mean ± SD. (E) Luciferase normalized activity (LUC activity) of HEK-293 cells co-transfected with PGL3-GATA1bs/HuSEC23B-2475 and -463 mutants and GATA1-WT showed a reduction of approximately 30 and 40% compared to GATA1bs-WT sequence. Data are presented as mean ± SD. (F) LUC activity of HEK-293 cells co-transfected with GATA1 mutants (GATA1-G208R, -R216W, -D218G, -V205M) and PGL3-HuSEC23B/3.44 in HEK-293 cell line. Data are presented as mean ± SD.

luciferase activity compared to those observed in co-transfected cells with GATA1bs-WT sequence (Figure 2E). Similarly, when we transiently co-transfected GATA1 mutants associated to X-linked cytopenias (GATA1/G208R, GATA1/R216W, GATA1/D218G, and GATA1/V205M)[5] and the HuSEC23B/3.44 in HEK-293 cells, we observed a reduction of HuSEC23B/3.44 luciferase activity induced by two specific mutants, GATA1-G208R (50%) and GATA1-R216W (40%), which are the causative variants of XLTDA and CEP, respectively, compared to that observed in co-transfected cells with GATA1-WT (Figure 2F). These results are in agreement with a CEP case recently reported, in which the severity of the anemia has been explained by the co-inheritance of the CEP-related GATA1-R216W with a rare SEC23B missense substitution.[13]

The identification of cis-acting regulatory elements in *SEC23B* upstream region could allow the definitive diagnosis of CDAII patients with both peculiar clinical phenotypes and incomplete mutation pattern. Although it has been classically conceived that phenotypes segregating in a Mendelian fashion are caused by alterations in a single gene or a unique locus, exceptions to this notion have been identified.[14] Indeed, it now seems more likely that a lot of Mendelian disorders could be explained with the combinations of multiple disease-causing alleles, or their combination with polymorphic variants. It is likely that common genetic variants associated with small changes in erythrocyte production act by slightly altering the binding and activity of GATA1 and co-factors, resulting in mild to moderate changes in the expression of target genes.[15] In the study herein we demonstrated that the clinical features of the case B-II.2 are the result of a complex genotype SEC23B-GATA1, suggesting a novel genetic etiology underlying CDAII. We described a *GATA1* polymorphism as a genetic modifier, which exacerbates the phenotype induced by mutations in GATA1-dependent genes. This is a novel paradigm that paves the way to new investigations for evaluating the role of this *GATA1* common variant as gene modifier in many other pathological conditions.

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

Funding: this work was supported by grants from the Italian Ministry of University and Research, by PRIN to AI (20128PNX83), by SIR to RR (RBSI144KXC), by grants from Regione Campania (DGRC2362/07).

Acknowledgments: the authors thank the CEINGE Service Facility platforms of Sequencing Core.

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Gambale A, Iolascon A, Andolfo I, Russo R. Diagnosis and management of congenital dyserythropoietic anemias. Expert Rev Hematol. 2016;9(3):283-296.
2. Russo R, Gambale A, Langella C, Andolfo I, Unal S, Iolascon A. Retrospective cohort study of 205 cases with congenital dyserythropoietic anemia type II: definition of clinical and molecular spectrum and identification of new diagnostic scores. Am J Hematol. 2014; 89(10):E169-175.
3. Bianchi P, Schwarz K, Högel J, et al. Analysis of a cohort of 101

CDAII patients: description of 24 new molecular variants and genotype-phenotype correlations. Br J Haematol. 2016;175(4):696-704.
4. Gao J, Chen YH, Peterson LC. GATA family transcriptional factors: emerging suspects in hematologic disorders. Exp Hematol Oncol. 2015;4:28.
5. Ciovacco WA, Raskind WH, Kacena MA. Human phenotypes associated with GATA-1 mutations. Gene. 2008;427(1-2):1-6.
6. Sankaran VG, Ghazvinian R, Do R, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. J Clin Invest. 2012;122(7):2439-2443.
7. Arlet JB, Ribeil JA, Guillem F, et al. HSP70 sequestration by free alpha-globin promotes ineffective erythropoiesis in beta-thalassaemia. Nature. 2014;514(7521):242-246.
8. Ludwig LS, Gazda HT, Eng JC, et al. Altered translation of GATA1 in Diamond-Blackfan anemia. Nat Med. 2014;20(7):748-753.
9. Del Vecchio GC, Giordani L, De Santis A, De Mattia D. Dyserythropoietic anemia and thrombocytopenia due to a novel mutation in GATA-1. Acta Haematol. 2005;114(2):113-116.
10. Russo R, Andolfo I, Manna F, et al. Increased levels of ERFE-encoding FAM132B in patients with congenital dyserythropoietic anemia type II. Blood. 2016;128(14):1899-1902.
11. Chlon TM, McNulty M, Goldenson B, Rosinski A, Crispino JD. Global transcriptome and chromatin occupancy analysis reveal the short isoform of GATA1 is deficient for erythroid specification and gene expression. Haematologica. 2015;100(5):575-584.
12. Ulirsch JC, Lacy JN, An X, Mohandas N, Mikkelsen TS, Sankaran VG. Altered chromatin occupancy of master regulators underlies evolutionary divergence in the transcriptional landscape of erythroid differentiation. PLoS Genet. 2014;10(12):e1004890.
13. Di Pierro E, Russo R, Karakas Z, et al. Congenital erythropoietic porphyria linked to GATA1-R216W mutation: challenges for diagnosis. Eur J Haematol. 2015;94(6):491-497.
14. Lupski JR. Digenic inheritance and Mendelian disease. Nat Genet. 2012;44(12):1291-1292.
15. Ulirsch JC, Nandakumar SK, Wang L, et al. Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. Cell. 2016;165(6):1530-1545.