

Beta-thalassemia: from genotype to phenotype

Fabrice Danjou, Franco Anni, Renzo Galanello

Dipartimento di Scienze Biomediche e Biotecnologie, Università di Cagliari, Ospedale Regionale Microcitemie ASL8, Cagliari, Italy

E-mail: renzo.galanello@mcweb.unica.it doi:10.3324/haematol.2011.055962

(Related Original Article on page 1712)

Beta-thalassemias are heterogeneous autosomal recessive hereditary anemias characterized by reduced or absent β globin chain synthesis. The resulting relative excess of unbound α globin chains precipitate in erythroid precursors in the bone marrow, leading to their premature death and, hence, to ineffective erythropoiesis. β -thalassemia phenotypes are variable, ranging from the severe transfusion dependent thalassemia major to the mild form of thalassemia intermedia. Patients with the major form of the disease have severe anemia, microcytic and hypochromic anemia, hepatosplenomegaly, and usually come to medical attention within the first two years of life. Without treatment, affected children have severely compromised growth and development and shortened life expectancy. Treatment with a regular transfusion program and chelation therapy, aimed at reducing transfusional iron overload, allows for normal growth and development and extends life expectancy into the third to fifth decade.¹ Individuals with thalassemia intermedia present later in life, have milder anemia (that never or only rarely requires transfusion), liver and spleen enlargement, typical bone modifications, and mild to moderate jaundice.² Occasionally patients with thalassemia intermedia are completely asymptomatic until adult life with only mild anemia. The major and intermedia forms of the disease are the two extremes of a wide range of clinical variability. Each group includes a continuous scale of severity, as demonstrated by the variability in age at which thalassemia major patients need transfusion; from months to years of life.

Beta-thalassemias are also very heterogeneous at the molecular level, with more than 200 disease-causing mutations so far identified; a complete updated list is available at the Globin Gene Server Web Site - <http://globin.cse.psu.edu/>. In most cases, mutations are single nucleotide substitutions, deletions or insertions of single nucleotides or small oligonucleotides leading to frameshift. Their diversity and the consequent variable degree of globin chain imbalance are the main determinants for milder phenotypes, the coinheritance of homozygosity or compound heterozygosity for mild β -thalassemia alleles being responsible for a consistent residual output of β chains from the affected β globin locus. However, much of the phenotypic variability is also explained by other genetic determinants capable of reducing the α /non- α chain imbalance thereby resulting in a lesser degree of α chain precipitation.³

One of the first discovered mechanisms able to reduce this imbalance is the coinheritance with homozygous β -thalassemia of α -thalassemia determinant. In this case, the severity of the clinical phenotype correlates with the α globin chain deficiency and with the improved α /non-

α chain imbalance as a consequence of reduced α chain output.⁴

A substantial decrease in α /non- α chain imbalance can also be obtained through the coinheritance of genetic determinants able to sustain a continuous production of gamma chains which, binding the excess α chains, result in a persistent fetal hemoglobin (Hb F) production measurable in adult life. In delta- β 0 thalassemia, this ability is due to deletions of variable extent within the β globin cluster,⁵ while in other cases it depends on the co-transmission of point mutations at A-gamma or G-gamma promoters (-196 C→T A-gamma; -158 C→T G-gamma).^{6,7} A mild phenotype may also be determined by coinheritance of genetic determinants associated with increased gamma chain production mapping outside the β globin cluster. Different polymorphisms at the *BCL11A* gene on 2p16.1⁸ and *HBS1L-MYB* intergenic region on 6q23.3 have been described.^{9,10} Several polymorphisms within intron 2 of the *BCL11A* gene have been strongly associated with Hb F levels: rs766432, rs4671393, rs1427407 and rs11886868, all in high linkage disequilibrium (LD) with each other.¹¹⁻¹⁵ Other independent signals in the same area were also identified with rs10189857 and rs7599488, in high LD, as well as rs7606173 and rs6706648,¹³ also in high LD with each other.^{12,13} In the *HBS1L-MYB* intergenic region, different SNPs have been described as being associated with Hb F variations in different studies: rs9399137, as well as rs4895441, rs9402686 and rs28384513.^{11,12} However, evidence has been reported of other contributing loci that have not been validated in recent genome-wide association studies, such as the 8q26-28 and Xp22.2-22.3 loci.^{16,17}

Together with these discoveries, the interest for prediction of Hb F levels and β -thalassemia phenotype has naturally grown in recent years, and the three loci previously mentioned have now been reported to be responsible for 20 to 50% of the Hb F trait variance in patients with β -thalassemia or sickle cell disease, and in healthy Europeans.¹⁸ Meanwhile, Galanello *et al.* reported the impact of variants in the *BCL11A* and *HBS1L-MYB* loci together with α gene defects on the clinical severity of β 0-39-thalassemia, quantifying their overall contribution to 75% of the variation differences between β 0-39-thalassemia major and intermedia phenotypes.¹⁴

The work by Badens *et al.*, presented in this issue of the Journal, extends previous studies by integrating the -158 C→T G-gamma polymorphism and β 0/ β + status, in addition to rs11886868 in the *BCL11A* gene, rs9389268 in the *HBS1L-MYB* intergenic region and α globin genes defects, to define β -thalassemia severity.¹⁹ Multivariate analysis including these five genetic modifiers was carried out and an accurate prediction has been made regarding major/intermedia status in more than 80% of

patients. The heterogeneity of this 106 patient cohort, with thirty different β globin gene mutations, might introduce variability not accountable for in the model, but it certainly provides a large amount of information on the prediction ability of the $\beta 0/\beta +$ status. Also, while it is likely that future studies will better define the genetic polymorphisms that modulate the effect of the *BCL11A* and *HBS1L-MYB* intergenic region loci and eventually uncover causal variants, results from Badens *et al.* already provide clinically relevant information to practitioners, clarifying the impact of genetic modifiers on the clinical severity of the disease. Furthermore, future studies will probably expand this predictive ability, including the effect of the different strongest independent predictors known to date for each gene (or even identified causal variants), and will eventually relate genetic modifiers to a more detailed measurement of clinical severity.

Very few other complex disease phenotypes can be explained in such depth, and prediction of patient risk as a function of their personal genetic background already offers support in clinical settings for β -thalassemia as opposed to most complex diseases. Extended molecular diagnosis can be carried out in patients affected by homozygous β -thalassemia to define their genotype for different modifiers and to better understand their phenotypic modulation abilities. Before long, β -thalassemia modifiers should allow us to redefine, on a genetic basis, the phenotypic definition actually in use.

Linkage analysis and genome-wide association studies have greatly contributed to such results, and next generation sequencing might further improve prediction ability and eventually guide the development of new therapies. At present, recent studies on Hb F modifier genes have produced mixed results: while a 3bp deletion associated with Hb F levels has been recently identified in the *HBS1L-MYB* intergenic region,²⁰ no such results have been obtained with the sequencing of the *BCL11A* gene. We hope that finally, in addition to providing detailed information to help promote enhanced β -thalassemia prenatal screening, achieving these objectives will also help to identify the mechanisms responsible for fetal hemoglobin control, since reactivation of fetal hemoglobin can provide major therapeutic benefits to people affected by β -hemoglobinopathies.

Fabrice Danjou, MD, specialized in epidemiology with a PhD in pediatric pharmacology. In recent years, he has mainly been involved in statistical genetics for genome-wide studies and clinical research. He currently works at both the Ospedale Microcitmico of Cagliari, Italy, and the Institute of Genetic and Biomedical Research for the Italian Research Council. Current interests are the analysis of high-throughput genotyping technology data for the clinical aspects of hematologic diseases. Franco Anni is a biologist with a PhD in pediatric pharmacology and a masters in bioinformatics. He is currently working as a researcher at the Department of Biological Sciences and Biotechnology of the University of Cagliari, Italy. In recent years, he has mainly been working on genome-wide studies both in Italy and at the Jefferson University Kimmel Cancer Center. He is now particularly interested in deep exome-sequencing data production and analysis. Professor Galanello is Professor of Pediatrics at the University of Cagliari. He is Director of the

Pediatric Clinic II of the Thalassemia Unit at Ospedale Regionale per le Microcitemie (a WHO Collaborating Center for Community Control of Hereditary Disease) and of the Department Biomedical Sciences and Biotechnologies, Cagliari, Italy. His main research interests are the clinical aspects and molecular genetics of thalassemias and other hemoglobinopathies, glucose-6-phosphate dehydrogenase deficiency, and Gilbert syndrome. Over the last ten years he has been the principal investigator of several clinical trials of new oral iron chelators. He is author of over 240 peer reviewed publications and is co-author of the books "Prevention of thalassemias and other haemoglobin disorders" (vols. 1 and 2), edited by the Thalassemia International Federation, and of the chapter "The Thalassemias" in Wintrobe's Clinical Hematology book (11th edition).

Financial and other disclosures provided by the author using the ICMJE (www.icmje.org) Uniform Format for Disclosure of Competing Interests are available with the full text of this paper at www.haematologica.org.

References

- Borgna-Pignatti C, Rugolotto S, De Stefano P, Zhao H, Cappellini MD, Del Vecchio GC, et al. Survival and complications in patients with thalassemia major treated with transfusion and deferoxamine. *Haematologica*. 2004;89(10):1187-93.
- Weatherall DJ, Clegg JB. The thalassemia syndromes. 4th Ed, Blackwell Science Ltd. 2001.
- Cao A, Galanello R, Rosatelli MC. Genotype-phenotype correlations in beta-thalassemias. *Blood Rev*. 1994;8(1):1-12.
- Galanello R, Cao A. Relationship between genotype and phenotype. *Thalassemia intermedia*. *Ann N Y Acad Sci*. 1998;850:325-33.
- Loudianos G, Cao A, Ristaldi MS, Pirastu M, Tzeti M, Kannavakis E, et al. Molecular basis of delta beta-thalassemia with normal fetal hemoglobin level. *Blood*. 1990;75(2):526-8.
- Tasiopoulou M, Boussiou M, Sinopoulou K, Moraitis G, Loutradi-Anagnostou A, Karababa P. G gamma-196 C-->T, A gamma-201 C-->T: two novel mutations in the promoter region of the gamma-globin genes associated with nondeletional hereditary persistence of fetal hemoglobin in Greece. *Blood Cells Mol Dis*. 2008;40(3):320-2.
- Labie D, Dunda-Belkhdja O, Rouabhi F, Pagnier J, Ragusa A, Nagel RL. The -158 site 5' to the G gamma gene and G gamma expression. *Blood*. 1985;66(6):1463-5.
- Menzel S, Garner C, Gut I, Matsuda F, Yamaguchi M, Heath S, et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet*. 2007;39(10):1197-9.
- Thein SL. Genetic modifiers of beta-thalassemia. *Haematologica*. 2005;90(5):649-60.
- Menzel S, Jiang J, Silver N, Gallagher J, Cunningham J, Surdulescu G, et al. The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. *Blood*. 2007;110(10):3624-6.
- Solovieff N, Milton JN, Hartley SW, Sherva R, Sebastiani P, Dworkis DA, et al. Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. *Blood*. 2010;115(9):1815-22.
- Galarnau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat Genet*. 2010;42(12):1049-51.
- Bhatnagar P, Purvis S, Barron-Casella E, Debaun MR, Casella JF, Arking DE, et al. Genome-wide association study identifies genetic variants influencing F-cell levels in sickle-cell patients. *J Hum Genet*. 2011;56(4):316-23.
- Galanello R, Sanna S, Perseu L, Sollaino MC, Satta S, Lai ME, et al. Amelioration of Sardinian beta0 thalassemia by genetic modifiers. *Blood*. 2009;114(18):3935-7.
- Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci USA*. 2008;105(5):1620-5.
- Dover GJ, Smith KD, Chang YC, Purvis S, Mays A, Meyers DA, et al. Fetal hemoglobin levels in sickle cell disease and normal individuals

- are partially controlled by an X-linked gene located at Xp22.2. *Blood*. 1992;80(3):816-24.
17. Garner C, Silver N, Best S, Menzel S, Martin C, Spector TD, et al. Quantitative trait locus on chromosome 8q influences the switch from fetal to adult hemoglobin. *Blood*. 2004;104(7):2184-6.
 18. Menzel S, Thein SL. Genetic architecture of hemoglobin F control. *Curr Opin Hematol*. 2009;16(3):179-86.
 19. Badens C, Joly P, Agouti I, Thuret I, Gonnet K, Fattoum S, et al. Variants in genetic modifiers of (beta)-thalassemia can help to predict the major or intermedia type of the disease. *Haematologica*. 2011 Jul 26. [Epub ahead of print]
 20. Farrell JJ, Sherva RM, Chen Z-Y, Luo H-Y, Chu BF, Ha SY, et al. A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression. *Blood*. 2011;117(18):4935-45.

The *JAK2* 46/1 haplotype: a marker of inappropriate myelomonocytic response to cytokine stimulation, leading to increased risk of inflammation, myeloid neoplasm, and impaired defense against infection?

Sylvie Hermouet^{1,2} and Mathias Vilaine¹

¹Inserm UMR 892, Institut de Recherche Thérapeutique, Université de Nantes, Nantes; ²Laboratoire d'Hématologie, Centre Hospitalier Universitaire, Nantes, France

E-mail: sylvie.hermouet@univ-nantes.fr doi:10.3324/haematol.2011.055392

(Related Original Article on page 1613)

The 46/1 haplotype of chromosome 9p, present in about 45% of the general population, is associated with a predisposition to mutations in the Janus Kinase 2 (*JAK2*) gene on the same allele and to chronic myeloproliferative neoplasms (MPN): polycythemia vera, essential thrombocythemia and primary myelofibrosis.^{1,4} The 46/1 haplotype is also associated with a predisposition to MPN with no mutation of *JAK2*, and with MPN with mutation in *MPL*, a gene located on a different chromosome (1p).^{5,6} An increased frequency of the 46/1 haplotype in patients with splanchnic vein thrombosis has also been reported but these findings remain controversial.^{7,8} The 46/1 haplotype has been studied in chronic myelogenous leukemia and in chronic myelomonocytic leukemia; no significant increase in frequency was noted (Table 1). Last year Andrikovics *et al.* reported that patients with acute myeloid leukemia (AML) with the 46/1 haplotype had a higher frequency of normal karyotype (NK).⁹ In the present issue of *Haematologica*, the same group found that the *JAK2* 46/1 haplotype is associated with an increased frequency of acute myelomonocytic leukemia and a tendency to reduced survival because of death from infection in patients with NK-AML.¹⁰ The latter findings need confirmation by other studies, for it is of great interest to establish whether the *JAK2* 46/1 haplotype is in fact a marker of myelomonocytic dysfunction and subsequently an unfavorable risk factor in NK-AML, as Nahajevszky *et al.* suggest, and/or in other disease categories.¹⁰ In fact, although the majority of studies have failed to detect any association of the 46/1 haplotype with hematologic and clinical parameters (Table 1), Tefferi *et al.* found that the 46/1 haplotype status influenced survival in primary myelofibrosis, and evolution toward myelofibrosis in polycythemia vera.⁵ Moreover, it has been reported that the frequency of the *JAK2* 46/1 haplotype is increased in the context of severe inflammation, for instance Crohn's disease.^{11,12} Altogether, the literature (reviewed in Table 1) shows that the common 46/1 haplotype is associated with predisposition to several types of disorders of variable severity: rare myeloid malignancies with or without *JAK2/MPL* mutation, including MPN and perhaps also acute myelomonocytic leukemia and NK-AML; inflamma-

tory diseases; and possibly, reduced defense against infection. The mechanisms that underlie the increased risk of acquisition of MPN and *JAK2/MPL* mutation in carriers of the 46/1 haplotype are not understood. Additionally, the mechanisms that make the V617F mutation occur preferentially in the *JAK2* gene of the 46/1 allele remain largely unknown.

A haplotype (contraction of "haploid genotype") is a set of closely linked genetic markers present on the same chromosome, which are not easily separable by recombination and thus tend to be inherited together, and can be identified by patterns of single nucleotide polymorphisms. The 46/1 haplotype, represented in Figure 1, is a 280 Kb-long region of chromosome 9p that includes three genes in their entirety: *JAK2*, *INSL6* (Insulin-like 6) and *INSL4*. Of the three genes, only *JAK2* is expressed in hematopoietic cells. The so-called "GGCC" part of the 46/1 haplotype begins in intron 10 and finishes in intron 15 of the *JAK2* gene and is characterized by four single nucleotide polymorphisms located in introns 10, 12, 14 and 15. The four single nucleotide polymorphisms are in complete linkage disequilibrium, meaning that they are always inherited together. The four single nucleotide polymorphisms replace three thymidines (T) and one cytosine (C) by two guanines (G) and two cytosines, resulting in a G G C C pattern, hence the phrase *JAK2* "GGCC" haplotype (not to be confused with "GC-rich"). The "GGCC" part of the 46/1 haplotype is not "GC-rich" but it includes sequences frequently mutated in MPN: *JAK2* exons 14 and 12, and to a lesser degree, exons 13 and 15. A plausible explanation for the high frequency of mutations in *JAK2* exons 14 and 12 and in *MPL* exon 10 in association with MPN combines the presence of flanking DNA repeat elements (represented in Figure 1) and the fact that the three exons encode for protein domains critical for the function of the *Jak2/Mpl* couple. Activating mutations in these exons are likely to be detected as they confer growth advantages to the mutated clone, whereas silent mutations typically remain below the detection level of most diagnostic assays.¹³ DNA repeat elements include GC- or AT-rich sequences; they can be mutation "hot spots", cause mismatching during DNA replication, or form fragile chromosomal break