Short Communication

Open Access CrossMark

Relative influence of submicroscopic genomic variants in the etiology of diseases and personalized medicine: time has come to move forward!

Abstract

For five decades, the discovery, mapping and analysis of genomic markers in non-coding, coding-, and/or intergenic regions have provided precious information which can be used for translational and personalized medicine. Indeed, submicroscopic genomic variations (e.g. point mutations, microsatellites, single nucleotide polymorphisms (*SNPs*), copy number variations (CNVs), microsatellites) have been associated with changes in gene expression and clinical phenotypes (e.g. pathologies, population diversity, genetic adaptation and/or evolution). Emerging findings, including my works, highlighted the important role of *SNPs* and CNVs in sickle cell anemia (SCA) patients. In particular, I reported that many of them often occurred in genes involved inflammation, auto-immunity, lipid metabolism, and cell adhesion when adult SCA patients with stroke complication were compared to stratified controls (e.g. groups of SCA patients without stroke). The dynamism of the genome with possible combined role of sub-microscopic genomic alterations in complex diseases such SCA, strongly suggest a need for elaborated multi-disciplinary approaches to treat patients in a personalized fashion. In this manuscript, I critically provide a short cut for personalized medicine by first describing major genomic variants before focusing on the role of *SNPs* and CNVs in human pathology using SCA, the first reported genetic disease, as a key example.

Keywords: genomics, structural submicroscopic variants, theranostics, personalized medicine, translational medicine, sickle cell anemia

Farid Menaa

Department of Oncology, Stem cells and Nanomedicine, Fluorotronics Inc., USA

Correspondence: Farid Menaa, Department of Oncology, Stem cells and Nanomedicine, Fluorotronics Inc., 2453 Cades Way, Vista, CA 92081, USA, Email dr.fmenaa@gmail.com

Abbreviations: NHEJ, non-homologous end joining; MAS, marker-assisted selection; SSRs, simple sequence repeats; STRs, short tandem repeats; *SNPs*, single nucleotide polymorphisms; CNVs, copy number variations; SCA, sickle cell anemia; MAF, minor allele frequency; LCRs, low-copy repeats; NGS, next-generation sequencing; GWAS, genomic-wide association studies; gDNA, genomic deoxyribonucleotide acid; HbF, fetal hemoglobin; MMBIR, microhomology- mediated break- induced replication

Sub-microscopic genomic variants

Mutations are defined as a change of the nucleotide sequence resulting from:[1–4] (i) unrepaired damage(s) to DNA (e.g. errors in non-homologous end joining (NHEJ); (ii) errors in the process of replication (e.g. error-prone translesion synthesis); (iii) a point mutation (e.g. nonsensense-, missense-); (iv) a frameshift mutation such as insertion (e.g. duplications) or deletion of DNA segments by mobile genetic elements (e.g. transposons/"jumping genes") through genetic recombination. Mutations (e.g. loss-of-function-, gain-of-function-, dominant negative/antimorphic-, lethal, back/reversion, conditional, neutral or silent) in somatic and/or germinal cells, can be spontaneous (e.g. tautomerism, depurination, deamination, slipped strand mispairing) and naturally occurring (e.g. during inheritance) or induced (e.g. by mutagens such as radiation, alkylating agents or by experimental mutagenesis). Mutations are essential for specie evolution,[5] and so for the genetic diversity (natural selection). They can be harmful/deleterious, beneficial/advantageous or neutral/nearly neutral.[6] *Microsatellites* represent simple, short or variable sequence repeats of DNA (e.g. $(CA)_n$), and so they are often named as SSRs (i.e. simple sequence repeats), STRs (short tandem repeats) or VNTRs

(variable number tandem repeats).[7] A large part of microsatellites are present in transposons.[8] Interestingly, these microsatellites are contributing to studies that investigate frame shift mutations (i.e. insertion or deletion), for marker-assisted selection (MAS), finger-printing and/or to better understand regulation of gene expression. Importantly, microsatellites are considered to be good genomic markers when the number of sequence repetitions is at least greater than 10 due to the fact that the level of inter- or intra-specific polymorphisms becomes higher.[9] Such length changes usually occur when potential for replication slippage during meiosis is relatively high.[10,11]

SNPs are common structural genomic variations (e.g. C>T) within a population or between populations (i.e. about 63millions *SNPs* in humans, according to NCBI). Interestingly, almost all *SNPs* are bi-allelic facilitating research investigations (e.g. minor allele frequency (MAF) at a locus).[12] *SNPs*, which affect only one single nucleotide base, are more frequently present in non-coding regions when compared to coding regions where two types of *SNPs* are found (i.e. synonymous and non-synonymous).[13] Unlike synonymous *SNPs*, non-synonymous *SNPs* (i.e. missense- or nonsense-) affect the protein sequence, and *SNPs* distributed in non-coding genomic regions are able to alter several cellular processes (e.g. pre-mRNA splicing, mRNA stability, gene expression). Genetic recombination, mutation rate, and/or AT microsatellites can determine SNP density (e.g. high $(AT)_n$ is linked to low SNP density).[14,15] *SNPs* are often associated with the susceptibility of certain diseases (e.g. SCA, cancers, and neurodegenerative disorders). CNVs represent a group of structural rearrangements of the genome from 1Kb to several mega-bases (e.g. deletions, duplications, inversions, translocations) that may contribute to the phenotypic

Relative influence of submicroscopic genomic variants in the etiology of diseases and personalized medicine: time has come to move forward!

Copyright:
©2014 Menaa   **34**

diversity in humans as well as to the etiology of complex pathologies such as cardio- and neurovascular diseases.[16–19] CNVs originate from inheritance, *de novo* mutation, low-copy repeats (LCRs) (http://en.wikipedia.org/wiki/Low_copy_repeats), segmental duplicate(s), alterations in the replication process (e.g. microhomology- mediated break- induced replication (MMBIR)).[19–22]

Role of SNPs and CNVs in etiology of diseases and personalized medicine

The evoked sub-microscopic genomic alterations, in particular mutations, *SNPs* and CNVs, are known to deeply contribute in the etiology of most human pathologies, depending on their respective frequency (i.e. relative rate), function (e.g. gain or loss of), molecular interaction (possible additive or synergic effects), genomic location (e.g. exon, intron, promoter, junction), microenvironment (e.g. epigenetic considerations such methylation, acetylation) (see "Bayesian Network" model.[23,24] Therefore, one should keep in mind that CNVs, *SNPs* and/or mutations (e.g. silent) are not always associated directly with a disease/a clinical phenotype at a determined time point or period.

Tremendous advances in the development of state-of-the-art genomic technologies (e.g. microarrays, next-generation sequencing (NGS) platforms)[25] and in the constitution of large databases derived from many international research investigations (e.g. Human genome sequencing Project, international HapMap project involving genomic-wide association studies (GWAS)) that aimed to genotype and map millions of variants[25,26] for better understanding of the complexity of particular diseases have been performed. Nevertheless, many gaps still need to be filled in order to obtain a reliable big picture of the functional gene/genomic dynamic in a spatio-temporal and clinical context (e.g. chemotherapy and/or radiotherapy, which can induce new genomic alterations).[27–29] Further, for personalized theranostic and prognosis medicine, the genomic deoxyribonucleotide acid (gDNA) used for sequencing the human genome, the current meta-analyses at the different biosystem levels (i.e. RNA and proteins) and using different technological platforms induced a relative number of bias (i.e. intra- and inter-errors).[27,30,31] To minimize such effects, deep analyses, interpretations and validations are requested. In this context, system biology much matters when OMICS are involved. Indeed and interestingly, one study on genetic variations between different species of Drosophila suggests that, if a mutation changes a protein produced by a gene, the result is likely to be harmful, with an estimated 70% of amino acid polymorphisms that have damaging effects, and the remainder being either neutral or weakly beneficial.[6] Nowadays, several databases describe the characteristics of variants in humans (e.g. frequency, location, their association between them and diseases) are available online (e.g. NCBI, OMIM, SNPedia, Human Gene Mutation, GWAS Central, Genebank).[25,26,32–36]

Until recently, GWAS have been mainly focused on associating *SNPs* with a particular clinical phenotype, which undoubtedly help a lot for personalized medicine.[37] Indeed, the identification of significant genetic variants of major effect or "modifiers" in complex diseases, can be used as markers for a specific disease such as age-related macular degeneration, diabetes, obesity, cancers, cardio- and neurovascular diseases (i.e. stroke).[28,38] For instance, a point mutation in the *APOE* (apolipoprotein E) gene was associated with a high susceptibility of developing Alzheimer disease.[39] In more complex diseases, *SNPs* rather work in coordination as seen in the case of osteoporosis (i.e. SNP-SNP interaction within *APOE* gene).[40] Besides, *SNPs* are relevant pharmacogenomic targets for drug therapy,[41] and represent stable inherited markers which are useful for specie evolution or adaptation studies independently of an observable phenotypic impact.[42]

In addition to *SNPs*, rare or common CNVs are distributed in the human genome and each CNV ranged from about 1Kb to several megabases in size.[34–36,43] Indeed, the HapMap data analysis estimated *SNPs* frequency to 83.6% while CNVs represented as low as 17.7%, with a little overlap (1.3%) between *SNPs* and CNVs signals.[34] This roughly confirms a more recent study which reported that CNV variation accounted for about 12% of the human genome.[43] Interestingly, since about 0.4% of the genome of unrelated people typically differs with respect to CNVs[44] and *de novo* CNVs,[45] CNVs can be used as markers for population studies and for twins/individual differentiation. Remarkable studies showed that the patterns of both *SNPs* and CNVs together combined to environmental factors are required to produce the disease phenotype.[16,36,46] Further, likewise *SNPs* and possibly in combination to them, CNVs can affect the individual's drug response individual and so, the subsequent susceptibility to health complications (e.g. disease resistance, adverse effects).[16] Thereby, CNVs has been associated with several complex health conditions (e.g. cancers, infections, auto-immune diseases, autism, schizophrenia, idiopathic learning disability).[47–54] Indeed, higher EGFR copy number than normal has been associated with non-small cell lung cancer.[47] However and importantly, a higher copy number of a particular gene (e.g. *CCL3L1*) is not always associated with a poor prognosis (i.e. HIV infection),[48] while a low copy number of certain other genes (e.g. *CD16*) can increase the risk of developing a complex disease (i.e. systemic lupus erythematosus).[49] Further, rare CNVs play a crucial role in causing disease, which is rarely observed with most common CNVs.[55] It is even worth to note that common CNVs of certain genes (e.g. *AMY1*) can even be beneficial, and this could be explained by their favored frequency during evolution for adaptation.[56,57] These observations are in line with my ongoing experiments in adult SCA patients which confirmed that rare CNVs can be reliable and targetable causative markers of diseases/disease complications.[58,59]

Etiology of sickle cell anemia: important influence of SNPs and CNVs

The discovery of a pathological hemoglobin S (HbS) by Pauling and colleagues in 1949 was the first demonstration that the production of an abnormal protein could be the cause of a genetic disorder.[60] Thereby, the SCA is a quite interesting example because:[28,38,58–72] (i) it is the first diagnosed molecular disease, which is caused by an unique mutation (i.e. single nucleotide substitution (β6Glu (GAG)→Val (GTG)) in the normal β-globin gene (HBB) and inherited following an autosomal recessive Mendelian pattern; (ii) it is the most common hemoglobinopathy, which induces a structural transformation of the normal (i.e. "donut-like shape") red blood cells (RBCs) into intravascular sickle RBCs (i.e. "croissant-like shape"); (iii) the homozygous form of SCA (HbSS), which is the symptomatic form, is associated with numerous complications, including stroke, a major health public concern worldwide manifested by vaso-occlusive events and episodic hemolysis; (iv) its large panel of subsequent complications were found to be associated with *SNPs* and/or CNVs (e.g. *SNPs* in the *CYBA, ANG1, TGFBR3, SELP, IL4R, ADRB2, VCAM1, LDL-R, AGT, ANXA2* or *TEK* genes; CNVs in the *UGT2B28* gene). Conversely, it

Relative influence of submicroscopic genomic variants in the etiology of diseases and personalized medicine: time has come to move forward!

Copyright:
©2014 Menaa
35

is important to keep in mind that certain genomic variants are not always associated with SCA seriousness. Thereby, *SNPs* in *ADCY9* or *BCL11A* genes were associated with decreased stroke risk in SCA patients due to their participation in up-regulating fetal hemoglobin (HbF) production. Overall, these findings suggest a need for multi-disciplinary approaches to manage SCA complexity with more confidence (e.g. over-expression of HbF but also for a large panel of genomic variants which can be used as reliable biomarkers of SCA disease).[28] The great advancement of the OMICS era shall provide soon more insights regarding the etiology of SCA complications and contribute to the implementation of the HapMap data in order to develop efficient and safer theranostics.

Conclusions and perspectives

Although GWAS are fast emerging, it is clear that the set of markers analyzed up-to-date do not cover the variability of the entire genome yet. Therefore, continuous data implementation is still needed. Owing to this consideration, CNVs can implement *SNPs* for the diagnosis and prevention of a given disease (i.e. monogenic or polygenic). Eventually, variants can be considered as valuable biomarkers of disease, and should permit precocious diagnosis and more efficient treatment of a given patient.

Acknowledgements

None.

Conflict of interest

Author declares that there is no conflict of interest.

References

1. Aminetzach YT, Macpherson JM, Petrov DA. Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. *Science*. 2005;309(5735):764–767.

2. Bertram J. The molecular biology of cancer. *Mol Aspects Med*. 2000;21(6):167–223.

3. Burrus V, Waldor M. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol*. 2004;155(5):376–386.

4. Hastings PJ, Lupski JR, Rosenberg SM, et al. Mechanisms of change in gene copy number. *Nat Rev Genet*. 2009;10(8):551–564.

5. Hurst GD, Werren JH. The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet*. 2001;2(8):597–606.

6. Sawyer SA, Parsch J, Zhang Z, et al. Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. *Proc Natl Acad Sci USA*. 2007;104(16):6504–6510.

7. Turnpenny P, Ellard S. *Emery's Elements of Medical Genetics*. 12th ed. London: Elsevier; 2005.

8. Scherer S. *A Short Guide to the Human Genome*. New York: Cold Spring Harbor University Press, Cold Spring; 2008.

9. Queller DC, Strassman JE, Hughes CR. Microsatellites and Kinship. *Trends Ecol Evol*. 1993;8(8):285–288.

10. Tautz D, Schlotterer. Simple sequences. *Curr Opin Genet Dev*. 1994;4(6):832–837.

11. Weber JL, Wong C. Mutation of human short tandem repeats. *Hum Mol Genet*. 1993;2(8):1123–1128.

12. Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001;409(6822):928–933.

13. Barreiro LB, Laval G, Quach H, et al. Natural selection has driven population differentiation in modern humans. *Nat Genet*. 2008;40(3):340–345.

14. Nachman MW. Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet*. 2001;17(9):481–485.

15. Varela MA, Amos W. Heterogeneous distribution of *SNPs* in the human genome: microsatellites as predictors of nucleotide diversity and divergence. *Genomics*. 2010;95(3):151–159.

16. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic variability. *Nat Rev Genet*. 2007;8(8):639–646.

17. Pollex RL, Hegele RA. Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation*. 2007;115(24):3130–3138.

18. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437–455.

19. Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *s*2006;52(1):103–121.

20. Cheng Z, Ventura M, She X, et al. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature*. 2005;437(7055):88–93.

21. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*. 2007;131(7):1235–1247.

22. Hastings PJ, Lupski JR, Rosenberg SM, et al. Mechanisms of change in gene copy number. *Nat Rev Genet*. 2009;10(8):551–564.

23. Pearl J. Bayesian Networks: A model of self-activated memory for evidential reasoning (UCLA Technical Report CSD-850017). *Proceedings of the 7th Conference of the Cognitive Science Society*. Irvine, California: University of California; 1985.

24. Pearl J. *Probabilistic Reasoning in Intelligent Systems*. San Francisco, California: Morgan Kaufmann; 1988.

25. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52–58.

26. International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437(7063):1299–1320.

27. Menaa F. Next-generation sequencing or the dilemma of large-scale data analysis: opportunities, insights, and challenges to translational, preventive and personalized medicine. *J Investig Genomics*. 2014;1(1):1–3.

28. Menaa F. Stroke in sickle cell anemia patients: a need for multidisciplinary approaches. *Atherosclerosis*. 2013;229(2):496–503.

29. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9(5):356–369.

30. Zhou B, Shi J, Whittemore AS. Optimal methods for meta-analysis of genome-wide association studies. *Genet Epidemiol*. 2011;35(7):581–591.

31. Craig DW, Goor RM, Wang Z, et al. Assessing and managing risk when sharing aggregate genetic variant data. *Nat Rev Genet*. 2011;12(10):730–736.

32. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004;305(5683):525–528.

33. Iafrate A, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet*. 2004;36(9):949–951.

Relative influence of submicroscopic genomic variants in the etiology of diseases and personalized medicine: time has come to move forward!

Copyright:
©2014 Menaa **36**

34. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007;315(5813):848–853.

35. Carlson B. *SNPs - A shortcut to personalized medicine. Genetic Engin Biotechn News*. 2008:28(12).

36. Sebastiani P, Ramoni MF, Nolan V, et al. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nat Genet*. 2005;37(4):435–440.

37. Wolf AB, Caselli RJ, Reiman EM, et al. *APOE* and neuroenergetics: an emerging paradigm in Alzheimer's disease. *Neurobiol Aging*. 2013;34(4):1007–1017.

38. Singh M, Singh P, Juneja PK, et al. SNP-SNP interactions within *APOE* gene influence plasma lipids in postmenopausal osteoporosis. *Rheumatol Int*. 2010;31(3):421–423.

39. Mohd Fareed, Mohammad Afzal. Single nucleotide polymorphism in genome-wide association of human population: A tool for broad spectrum service. *Egyptian J Med Hum Genet*. 2013;14(2):123–134.

40. Thomas PE, Klinger R, Furlong LI, et al. Challenges in the association of human single nucleotide polymorphism mentions with unique database identifiers. *BMC Bioinformatics*. 2011;12(Suppl 4):S4.

41. Human Genome Structural Variation Working Group, Eichler EE, Nickerson DA, et al. Completing the map of human genetic variation. *Nature*. 2007;447(7141):161–165.

42. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444–454.

43. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Ann Rev Med*. 2010;61:437–455.

44. Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature*. 2008;453(7191):56–64.

45. Human Genetic Variation Fact Sheet. *National Institute of General Medical Sciences (NIH)*. 2008.

46. Estivill X, Armengol L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet*. 2007;3(10):1787–1799.

47. Cappuzzo F, Hirsch FR, Rossi E, et al. Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. *J Natl Cancer Inst*. 2005;97(9):643–655.

48. Gonzalez E, Kulkarni H, Bolivar H, et al. The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*. 2005;307(5714):1434–1440.

49. Aitman TJ, Dong R, Vyse TJ, et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature*. 2006;439(7078):851–855.

50. Cook EH, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature*. 2008;455(7215):919–923.

51. Pinto J, Pagnamenta AT, Klei L, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010;466(7304):368–372.

52. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of *de novo* copy number mutations with autism. *Science*. 2007;316(5823):445–449.

53. St Clair D. Copy number variation and schizophrenia. *Schizophr Bull*. 2008;35(1):9–12.

54. Knight S, Regan R, Nicod A, et al. Subtle chromosomal rearrangements in children with unexplained mental retardation. *Lancet*. 1999;354(9191):1676–1681.

55. Welcome Trust Case Consortium, Craddock N, Hurles ME, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*. 2010;464:713–720.

56. Zhang F, Gu W, Hurles ME, et al. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 2009;10:451–481.

57. Perry GH, Dominy NJ, Claw KG, et al. Diet and evolution of human amylase gene copy number variation. *Nat Genet*. 2007;39(10):1256–1260.

58. http://www.ichg2011.org/ and http://www.ashg.org/

59. http://events.hct.ac.ae/2012/02/12/744/ and http://www.icddt.com/

60. Pauling L, Itano H, Singer SJ, et al. Sickle cell anemia: a molecular disease. *Science*. 1949;110(2865):543–548.

61. Bender MA, Hobbs W. *Sickle Cell Disease*. In: Pagon RA, Adam MP, Ardinger HH, editors. GeneReviews™. Seattle, WA: University of Washington; 2003.

62. Ingram VM. A specific chemical difference between the globins of normal human and sickle-cell anemia haemoglobin. *Nature*. 1956;178(4537):792–794.

63. Frenette PS, Atweh GF. Sickle cell disease: old discoveries, new concepts, and future promise. *J Clin Invest*. 2007;117(4):850–858.

64. Bunn HF. Pathogenesis and treatment of sickle cell disease. *N Engl J Med*. 1997;337(11):762–769.

65. Steinberg MH. Pathophysiology of sickle cell disease. *Baillieres Clin Haematol*. 1998;11(1):163–184.

66. Taylor JG 6th, Nolan VG, Mendelsohn L, et al. Chronic hyper-hemolysis in sickle cell anemia: association of vascular complications and mortality with less frequent vasoocclusive pain. *PLoS One*. 2008;3(5):e2095.

67. http://www.ichg2011.org/ and http://www.ashg.org/

68. http://www.eshg.org/eshg2011

69. Tang DC, Prauner R, Liu W, et al. Polymorphisms within the angiotensinogen gene (GT-repeat) and the risk of stroke in pediatric patients with sickle cell disease: a case-control study. *Am J Hematol*. 2001;68(3):164–169.

70. Flanagan JM, Frohlich DM, Howard TA, et al. Genetic predictors for stroke in children with sickle cell anemia. *Blood*. 2011;117(24):6681–6684.

71. Hoppe C, Klitz W, Cheng S, et al. Gene interactions and stroke Gene interactions and stroke risk in children with sickle cell anemia. *Blood*. 2004;103(6):2391–2396.

72. Uda M, Galanello R, Sanna S, 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 U S A*. 2008;105(5):1620–1625.