Published in final edited form as:
Curr Opin Cardiol. 2011 May ; 26(3): 175–180. doi:10.1097/HCO.0b013e3283459857.

Medical DNA Sequencing

AJ Marian, M.D.
The Brown Foundation Institute of Molecular Medicine, Texas Heart Institute at St. Luke's Episcopal Hospital The University of Texas Health Science Center Houston, TX

Abstract

Purpose of review—To discuss implications of information garnered through whole genome and exome sequencing in the practice of cardiovascular medicine

Recent findings—Whole-genome and exome sequencing unveils medical information embedded in individual genomes and exomes, which could be incorporated into the practice of medicine for diagnostic and therapeutic gains. The mankind, however, has considerable genetic diversity, as each genome encompasses about 4 million DNA sequence variants (DSVs). The challenging task is to identify the variants that have clinical implications.

DSVs exert a continuum of effect sizes on the phenotype that ranges from negligible to large. From clinical perspective, selected categories, in order of their significance, are disease-causing; likely disease-causing; disease-associated; biologically functional but unknown clinical significance; and unknown functional and clinical significance variants. The frequency of DSVs in the genome also follows a gradient from rare for the disease-causing variants to common for variants with unknown clinical and biological significance. A subset of DSVs might have implications in accurate and preclinical diagnosis, prognostication and individualization of therapy. Clinical phenotypes, however, are too complex to be determined solely by a single DSV. Even in the case of disease-causing variants, the severity of the disease is determined by multiple additional genetic and non-genetic factors.

Summary—Medical DNA sequencing is expected to retool the clinicians with the information content of DSVs. DSVs with large effect sizes are likely to offer clinical utility in early and pre-clinical diagnosis, prognostication and individualization of therapy.

Keywords

Genetics; Next-Generation Sequencing; Polymorphism; Genetic Testing

INTRODUCTION

Advances in science are typically incremental, albeit some increments are larger than the others and a few are immense. Among the gigantic leaps forward in molecular genetics was the development of DNA sequencing by the chain termination technique by Dr. Frederick Sanger in 1977 [1]. Dr. Sanger and Dr. Walter Gilbert were awarded the Nobel Prize in Chemistry in 1980 for the development of DNA sequencing methods. The Sanger

Address for correspondence Director, Center for Cardiovascular Genetics Institute of Molecular Medicine University of Texas Health Sciences Center Texas Heart Institute at St. Luke's Episcopal Hospital 6770 Bertner Street, Suite C900A Houston, TX 77030 Phone: 713 500 2350 Fax: 713 500 2320 Ali.J.Marian@uth.tmc.edu.

sequencing, as it is commonly known, revolutionized the field of molecular genetics and biology. It had an enormous impact in deciphering molecular genetic basis of various diseases, particularly single gene disorders. Automation of the technique led to completion of the draft sequence of the human genome about 10 years ago, the finished sequence in 2004 and sequencing of the first diploid human genome in 2007 [2–4]. Generation of high throughput instruments led to a precipitous drop in the cost of DNA sequencing by Sanger method from the initial price of ~ $30 per base to the current price of ~ 1 penny per base. Large-scale or genome or exome sequencing by Sanger sequencing, therefore, remained costly and time consuming. Sequencing of the human genome in its draft form by Sanger sequencing took approximately 11 years and cost approximately $3 billion. Today, sequencing of a genome by the Sanger technique would cost several months and several million dollars. Therefore, large-scale applications of DNA sequencing to elucidate the molecular genetic basis of common disorders and garner information embedded in the nucleotide sequence for medical use had to await the development of faster and cheaper DNA sequencing technologies.

Dr. Sydney Brenner, Nobel Laureate in Physiology and Medicine (2002), usher in a new era in nucleic acid sequencing by introducing the massively parallel signature sequencing (MPSS) technique in 2000 [5]. The technique introduced the concept of sequencing millions of copies of the DNA fragments simultaneously. It soon followed by the alternative approaches of massively parallel sequencing based on multiplex polony sequencing [6], pyrosequencing (454 Roche, Inc) [7], reversible dye-termination (Genome Analyzer by Illumina, Inc.) and sequencing by ligation (SOLiD by Applied Biosystems, Inc). Collectively, these new technologies are referred to as the “Next-Generation Sequencing” platforms, as opposed to the automated Sanger Sequencing method, which is considered as a `first-generation' technology. The next-generation sequencing instruments have increased the DNA sequencing output and have reduced the cost of DNA sequencing by several orders of magnitude. Some of the newer instruments can generate up to 200 gigabases (Gb) per sequencing run or 20 to 30 Gb of DNA sequence per day. The output provides excellent read coverage for a couple of human genomes in toto (each genome is about 3.2 billion base pairs) and several exomes (each exome is about 30 million base pairs). Thus, the next generation sequencing instruments have made it possible to sequence the entire human genome in weeks if not days at a cost of several thousands dollars [8]. The sequencing technology continues to improve and along with it the bioinformatics tools continue to advance our ability to analyze the massive amount of the sequence data that is generated by these machines. These advances are making it possible to bring the power of DNA sequence to the bedside and apply it to the practice of medicine.

TEXT OF REVIEW

There is no phenotype that does not have a genetic component as a determinant, whether it is susceptibility to a specific disease or a response to an explicit intervention. Single gene disorders are clear and well-appreciated examples of the effects of genetic mutations on the clinical phenotypes. Likewise, familial aggregation of complex traits, such as coronary atherosclerosis, even in the absence of a clear Mendelian inheritance pattern, as well as twin studies point to the contributions of the genetic factor to the common complex phenotype. Perhaps, less clear are the contributions of the genetics to susceptibility to the so-called acquired diseases, for example viral myocarditis. Here, genetic factors are important determinants of susceptibility to viruses, severity of the inflammatory responses, response to therapeutic interventions and the clinical outcome.

Conventionally inheritance is considered vertical transmission of the genetic information through DNA nucleotides. Therefore, the search for the genetic basis of human diseases has

primarily focused on identification of the variations in the nucleotide sequence rather than other modifications to the genome. Epigenetic modifications of the genome which lead to stable propagation of gene expression from one generation to another are expected to be also important determinants of heritability of the phenotype [9]. Epigenetics might turn out to be as important as the DSVs, as the heritable determinant of the phenotype. The specific epigenetic changes that might be responsible for heritable cardiovascular diseases are currently less well defined.

Medical sequencing in single gene disorders

Genetic studies are most powerful in familial diseases. Single gene diseases are typically familial diseases with a clear pattern of inheritance. On the spectrum of the nature's genetic and phenotypic gradients, single gene disorders are caused by variants that exert large effect sizes [10]. Typically DSVs with large effect sizes are rare and perhaps, for that reason are easier to identify and establish as the causal variants. Consequently, familial single gene disorders are most suitable for identifying the causal mutations. It is not surprising that molecular genetic techniques, namely positional cloning and DNA sequencing have been successfully applied to delineate the genetic causes of several dozen cardiovascular diseases.

Genotype-phenotype correlation studies have further emphasized the significance of genetic background, namely modifier sequence variants, as important determinants of the phenotype in single gene disorders [11]. It is evident that the phenotype, even in single-gene disorders, results from the cumulative effects of multiple alleles with the causal allele contributing the most and others to variable degrees. Next-generation sequencing in familial single gene disorders affords the opportunity not only to identify the causal but also the modifier variants. Advanced bioinformatics and mathematical modeling are necessary to identify the modifier variants among the large number of alleles that are typically shared among the family members [11].

We recently sequenced the exomes of 7 members of a family with dilated cardiomyopathy (DCM) of whom 4 were clinically affected. Approximately 2/3 of the exons were adequately covered to provide reliable sequence reads. As would be expected every family member had about 5,000 non-synonymous single nucleotide polymorphisms (nsSNP) including about 200 novel nsSNPs. Moreover, 28 nsSNPs including 4 novel nsSNPs co-segregated with the phenotype, i.e., were present in 4 individuals with DCM but were absent in 3 phenotypically normal individuals. Genes containing the 4 novel missense variants are expressed in the heart but not exclusively. Among the shared variants is a known variant in *JPH2*, which is a known gene for cardiomyopathies [12, 13]. The findings of several novel and known variants co-segregating with the phenotype illustrate the challenges one faces in identifying the causal variants in small and mid size families with the relatively common autosomal dominant diseases. This is contrast to autosomal recessive diseases or rare autosomal dominant diseases, such as Kabuki or Miller syndromes or large families wherein deep sequencing has led to identification of the causal variants [14–16]. The presence of multiple rare or uncommon variants suggest the power of next-generation sequencing in delineating the genetic determinants of the phenotype as a whole and the potential difficulties in specifically identifying the specific causal mutations in the relatively common autosomal dominant diseases.

Medical sequencing in common complex diseases

Contributions of genetic factors to susceptibility to complex phenotype is supported by the familial aggregation and twin studies [17–19]. The proportion of a complex phenotype that is genetically determined, i.e., heritability, varies from 20% to 80%, depending on the phenotype and the specifics of the studies. While there is a genetic component to all

complex phenotype, unlike the single gene disorders, typically there is no principal allele to impart a major effect. The complex phenotype results from the cumulative effects of a large number of alleles, each making a small contribution to expression of the phenotype. The focus of genetic studies of complex traits, because of technical limitations, has been on the common alleles. Genome-Wide Association Studies (GWAS), wherein cases and controls are typed for several thousands to a million alleles, offer the most unbiased approach to identification of genetic variants that are associated with a complex phenotype. GWAS compare the frequencies of the alleles in cases with the phenotype of interest and controls without. The current generation of GWAS, by design, is restricted to genotyping of the known and typically the common SNPs. The common alleles, however, typically exert modest and often indiscernible effects on the phenotype. Therefore, several thousands cases and controls are typed to detect modest differences in the allele frequencies. Consequently, clinical utility of the results of GWAS – in terms of diagnosis and risk stratification – has been quite limited and practically nil. For instance, the effect sizes of SNPs that have been associated with systemic blood pressure or plasma levels of low-density or high-density lipoprotein cholesterol in GWAS have been quite modest [20–23]. This is also the case for electrocardiographic phenotypes such as the QRS duration or the QT interval [24, 25]. Nevertheless, it is important to clarify and emphasize that despite the small effect sizes of the associated alleles with the complex phenotype, GWAS could lead to identification of novel molecular pathways and hence, shed significant insights into the molecular pathogenesis of the phenotype, when complemented with functional studies [26, 27]. A partial list of GWA studies including GWAS of cardiovascular phenotypes can be found at http://www.genome.gov/26525384.

Despite the success of GWAS in findings alleles that are associated with complex phenotypes, it has become evident that the common variants account only for small fraction of heritability of the complex phenotypes and a significant fraction of heritability of a complex phenotype remains unaccounted for. This has led to the notion of “missing heritability” or “the dark matter of heritability”. The missing heritability in part may be explained by the epigenetic factors [28]. Alternatively rare DSVs that might impact large effect sizes on the complex phenotypes might explain the “missing heritability”. Deep DNA sequencing using next-generation sequencing instruments is being applied to identify the rare alleles with large effect sizes as determinants of complex phenotypes.

The existing DNA sequencing technology affords the opportunity to sequence the entire exome and genome at a reasonable cost (currently ~ $10,000.00 but shifting toward < $1,000 per genome). The daunting challenge, however, is to distinguish the clinically important alleles. Each human genome contains approximately 3.5 million SNPs and about 10,000 nsSNPs, including a couple of hundred novel nsSNPs (reviewed in [10]). These nsSNPs are not totally innocuous but do exert several biological functions. The functional and biological significance of these variants follows a gradient that ranges from negligible to profound [10]. About 2/3rd of the nsSNPs could be predicted to be pathogenic. In addition, each genome has several hundred thousands structural variations, some of which encompass several thousand and some times million base pairs of DNA [29, 30]. These structural variations could lead to duplication or deletion of a gene or multiple genes and hence, might have significant clinical implications. The clinical phenotype is the consequence of interactions of a number of alleles, one or two exerting a relatively large effect sizes, a handful exerting moderate and a huge number exerting modest effect sizes. Collectively, these findings illustrates the complexity of the human genome and the difficulty in deciphering the contributions of the variants, whether SNPs or copy number variants (CNVs), to the phenotype.

Clinical applications of DNA sequencing

With the continued drop in the price of whole genome and whole exome sequencing, it is likely that genome and exome sequence data will be part of every patient's medical chart in the near future. Physicians will have access to about 4 million DSVs in every patient's genome. Hence, they will face the intimidating task of identifying individuals who are at risk of a disease prior to development of the clinical phenotype, determining responsiveness to a therapy, and predicting the clinical outcomes. The clinical significance of DSVs will follow a gradient from being highly significant, such as the causal variants to those that have no clear biological significance. For simplicity and from clinical perspective selected regions in the gradient of phenotypic effects of DSVs merits highlighting (Table 1):

Disease-causing variants

This category entails variants that have been already established to cause diseases in humans, typically in families with single gene disorders. Evidence for causality in this category is strong and supported by several lines of data. The phenotype co-segregates with inheritance of these variant in relatively large families or in multiple families. These variants also have several characteristics (Table 2) that support their causal role in the pathogenesis of the phenotype and imply their clinical, biological and functional significance.

Disease-causing variants are expected to be rare in each genome/exome. They are typically identified in familial setting and sometimes in the pre-clinical stage as part of the genetic screening of the at risk family members. Identification of disease-causing variants could enable the physicians to make pre-clinical diagnosis and intervene – whenever applicable – to prevent the disease and predict the clinical outcomes or slow evolution of the phenotype. In addition, utilizing the functional and biological evidence, physicians might be able to target the therapies toward specific pathways involved in the pathogenesis of the phenotype. Finally, detection of these variants will enable the physicians to establish a firm and robust diagnosis and exclude the potential phenocopy conditions. Nevertheless, physicians must be aware that the clinical phenotypes are too remote from the DNA nucleotides and too complex to be determined entirely by the nucleotide changes, regardless of the magnitude of the effect size. A clinical phenotype, even in single gene disorders, results from the cumulative effects of multiple alleles, rare and common, and various other factors including epigenetics, microRNAs, post-translational modifications and the environmental factors among the others. As would be expected disease-causing variants will have the highest position in the gradient of phenotypic effects and clinical implications.

Likely disease-causing variants

These variants are typically present exclusively in the family members who exhibit the phenotype or in the cases with the phenotype but not in normal family members or independent normal individuals. However, unlike the disease-causing variants, the strengths of the genetic data is somewhat limited, typically because of the relatively small size of the families that hinders establishing robust genetic linkage or co-segregation. Likewise, low penetrance further compounds conclusiveness of the causal role of these variants. Therefore, the variant may be present in a minority of family members or sporadic cases who do not exhibit a phenotype, at least at a young age. Robust statistical data, nonetheless, is necessary to consider the variants as the likely-disease causing variants. Likewise, the significance of these variants is partly determined by their functional and biological significance per criteria listed in Table 2. Clinical implications of the likely disease-causing variants are similar to those described for the disease-causing variants except that the level of certainty is less.

Disease-associated variants

Evidence of an association is typically based on the observation of higher frequencies of these variants in cases with the disease than in controls. The evidence is preferably based on the results of large GWAS. A subset that are functional or impart major effects on protein structure are expected to impart greater clinical significance that those without. Accordingly, a non-synonymous SNP (nsSNP) that affects enzymatic function of a protein is more likely to offer valuable information than an SNP that bis located in an intergene region. Nevertheless, even though synonymous, intronic and intergenic SNPs might have considerable biological functions. The clinical significance of this category of DSVs largely reflects the effect size of the alleles but is typically low. The effect size of the variant should not be equated with the statistical significance level, typically reflected by the p values. The p values in this type of studies from clinical point of view are not much valuable. To clarify, an increase in the minor allele frequency from 0.45 to 0.50 in a very large study population could lead to exceedingly low p values but a 5% increase in the frequency of an allele in a population does not offer much clinical utility in a single individual. Likewise, the relative risks or Odds ratios should be interpreted in the context of pre-test likelihood of the clinical event. For example a 2-fold increase in the risk of heart failure is not much informative if the *a priori* risk of heart failure is 1:100,000 in the population. It is also important to keep in mind that the disease-associated allele might be in linkage disequilibrium (LD), defined as co-segregation of two alleles, often adjacent, more than by chance alone, with another allele that is responsible for the phenotypic effect. The extent of LD in the genome varies but could extent to several million base pairs [31]. Thus, disease-associated DSVs have limited utility in prognostication and risk stratification, as the clinical phenotypes are the consequences of entwined, and dynamic interactions among multiple alleles, other genetic and genomic factors and the environmental factors.

DSVs that are associated with the inter-individual variations in response to drugs, both efficacy and toxicity typically exert modest to moderate effect sizes and have some clinical utility. Variants of *CYP2C9* and *VKORC1*, which encode cytochrome P450 isoform 2C9 and vitamin K epoxide reductase are responsible for approximately 30% of the inter-individual variations in Coumadin dose. Likewise, DSVs in genes coding for P-450 enzymes CYP3A4, CYP3A5, and CYP2C19 are associated with responsiveness to clopidogrel, an inactive pro-drug that is converted to an active metabolite in the liver [32, 33]. Similarly, DSVs in α2C- and β1-adrenergic receptors can influence the response of patients with systolic heart failure to treatment with β blockers [34, 35]. Moreover, DSVs in *APOE,PCSK9*, and *HMGCR* have been implicated in response to statins [36]. In terms of drug toxicity, DSVs in *SLCO1B1*, encoding solute carrier organic anion transporter 1B1 are associated with statin-induced myopathy [37, 38]. Likewise, DSVs in genes causing congenital long QT syndrome are associated with drug-induced cardiac arrhythmias [39]. Given the clinical implications and financial burden of DSVs in drug efficacy and toxicity, medical DNA sequencing could have considerable pharmacogenetics implications. Finally, it merit noting that scientific value of the disease-associated variants, despite the relatively low clinical utility of the vast majority of them, could be tremendous, as they can point out to novel pathways and mechanisms [26, 40].

Functional variants not associated with a disease

The human genome is quite complex and contains various forms of genetic polymorphisms including insertions, deletions, non-sense variants, splice junction variants, CNVs, etc. A significant number of these variants are potentially functional or are known to exert biological functions. However, these variants are not associated with any clinical phenotype or a disease. For example, each human in average contains about 10,000 nsSNP, a significant number of which are expected to be biologically functional or pathogenic, based

on bioinformatics predictions but are not known to cause any disease. These variants, at the present time, have minimal clinical utility or application.

Variants with unknown significance

The vast majority of ~ 4 million DNA sequence variants in the genome probably fit into this category. With the exception of few, all variants located in inter-genic regions and introns are not known to convey biological functions. These variants have no clinical utility at the present time.

CONCLUSION

Advances in DNA sequencing technologies have afforded the opportunity to delineate the molecular genetic basis of common as well as rare diseases. Collectively, these advances have ushered in the era of medical DNA sequencing. Humans, however, have considerable genetic diversity, as each genome contains approximately 4 million DSVs. A small number of these DSVs exert large effect sizes. Hence, medical sequencing could afford the opportunity to harness the information content of these DSVs for pre-clinical and accurate diagnosis, individualization of therapy and even prognostication. The highest impact of whole genome/exome sequencing is in providing insights into the pathogenesis of human diseases, which in turn could lead to identification of new diagnostic and prognostic markers and drug targets. Clinicians are astute to recognize that the clinical phenotypes are complex and result from intertwined, non-linear, and often stochastic interactions among multiple constituents including non-genetic factors in addition to DSVs. Despite some limitations, medical sequencing is expected to play a prominent role in the practice of medicine in the 21st century.

Acknowledgments

Supported by grants from the NHLBI, Clinical Scientist Award in Translational Research from the Burroughs Wellcome Fund, and The TexGen Fund from the Greater Houston Community Foundation.

REFERENCES

[1]. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc.Natl.Acad.Sci.U.S.A. 1977; 74:5463–5467. [PubMed: 271968]

[2]. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science. 2001; 291:1304–1351. [PubMed: 11181995]

[3]. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]

[4]. Levy S, Sutton G, Ng PC, et al. The Diploid Genome Sequence of an Individual Human. PLoS.Biol. 2007; 5:e254. [PubMed: 17803354]

[5]. Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol. 2000; 18:630–4. [PubMed: 10835600]

[6]. Shendure J, Porreca GJ, Reppas NB, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005; 309:1728–32. [PubMed: 16081699]

[7]. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437:376–80. [PubMed: 16056220]

[8]. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010; 327:78–81. [PubMed: 19892942]

[9]. Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. Science. 330:612–6. [PubMed: 21030644]

[10]. Marian AJ. Nature's genetic gradients and the clinical phenotype. Circ Cardiovasc Genet. 2009; 2:537–9. [PubMed: 20031631]

[11]. Daw EW, Chen SN, Czernuszewicz G, et al. Genome-wide mapping of modifier chromosomal loci for human hypertrophic cardiomyopathy. Hum Mol Genet. 2007; 16:2463–71. [PubMed: 17652099]

[12]. Landstrom AP, Weisleder N, Batalden KB, et al. Mutations in JPH2-encoded junctophilin-2 associated with hypertrophic cardiomyopathy in humans. J Mol Cell Cardiol. 2007; 42:1026–35. [PubMed: 17509612]

[13]. Matsushita Y, Furukawa T, Kasanuki H, et al. Mutation of junctophilin type 2 associated with hypertrophic cardiomyopathy. J Hum Genet. 2007; 52:543–8. [PubMed: 17476457]

[14]. Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010; 42:790–3. [PubMed: 20711175]

[15]. Roach JC, Glusman G, Smit AF, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science. 2010; 328:636–9. [PubMed: 20220176]

[16]. Musunuru K, Pirruccello JP, Do R, et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. N Engl J Med. 2010; 363:2220–7. [PubMed: 20942659]

[17]. Marenberg ME, Risch N, Berkman LF, et al. Genetic susceptibility to death from coronary heart disease in a study of twins. N.Engl.J.Med. 1994; 330:1041–1046. [PubMed: 8127331]

[18]. Post WS, Larson MG, Myers RH, et al. Heritability of left ventricular mass: the Framingham Heart Study. Hypertension. 1997; 30:1025–1028. [PubMed: 9369250]

[19]. Adams TD, Yanowitz FG, Fisher AG, et al. Heritability of cardiac size: an echocardiographic and electrocardiographic study of monozygotic and dizygotic twins. Circulation. 1985; 71:39–44. [PubMed: 4038369]

[20]. Arora P, Newton-Cheh C. Blood pressure and human genetic variation in the general population. Curr Opin Cardiol. 2010

[21]. Kathiresan S, Manning AK, Demissie S, et al. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. BMC Med Genet. 2007; 8(Suppl 1):S17. [PubMed: 17903299]

[22]. Pirruccello J, Kathiresan S. Genetics of lipid disorders. Curr Opin Cardiol. 2010

[23]. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet. 2008; 40:189–197. [PubMed: 18193044]

[24]. Sotoodehnia N, Isaacs A, de Bakker PI, et al. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. Nat Genet. 2010; 42:1068–76. [PubMed: 21076409]

[25]. Pfeufer A, Sanna S, Arking DE, et al. Common variants at ten loci modulate the QT interval duration in the QTSCD Study. Nat Genet. 2009; 41:407–14. [PubMed: 19305409]

[26]. Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010; 466:714–9. [PubMed: 20686566]

[27]. Wahe A, Kasmapour B, Schmaderer C, et al. Golgi-to-phagosome transport of acid sphingomyelinase and prosaposin is mediated by sortilin. J Cell Sci. 2010; 123:2502–11. [PubMed: 20571055]

[28]. Jiang YH, Bressler J, Beaudet AL. Epigenetics and human disease. Annu.Rev Genomics Hum Genet. 2004; 5:479–510. 479-510. [PubMed: 15485357]

[29]. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010; 11:446–50. [PubMed: 20479774]

[30]. Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. Nature. 2008; 453:56–64. [PubMed: 18451855]

[31]. Dawson E, Abecasis GR, Bumpstead S, et al. A first-generation linkage disequilibrium map of human chromosome 22. Nature. 2002; 418:544–548. [PubMed: 12110843]

[32]. Simon T, Verstuyft C, Mary-Krause M, et al. Genetic determinants of response to clopidogrel and cardiovascular events. N Engl J Med. 2009; 360:363–75. [PubMed: 19106083]

[33]. Mega JL, Close SL, Wiviott SD, et al. Cytochrome p-450 polymorphisms and response to clopidogrel. N Engl J Med. 2009; 360:354–62. [PubMed: 19106084]

[34]. Liggett SB, Mialet-Perez J, Thaneemit-Chen S, et al. A polymorphism within a conserved beta(1)-adrenergic receptor motif alters cardiac function and beta-blocker response in human heart failure. Proc Natl Acad Sci U S A. 2006; 103:11288–93. [PubMed: 16844790]

[35]. Small KM, Wagoner LE, Levin AM, et al. Synergistic polymorphisms of beta1- and alpha2C-adrenergic receptors and the risk of congestive heart failure. N Engl J Med. 2002; 347:1135–42. [PubMed: 12374873]

[36]. Thompson JF, Hyde CL, Wood LS, et al. Comprehensive whole-genome and candidate gene analysis for response to statin therapy in the Treating to New Targets (TNT) cohort. Circ Cardiovasc Genet. 2009; 2:173–81. [PubMed: 20031582]

[37]. Voora D, Shah SH, Spasojevic I, et al. The SLCO1B1*5 genetic variant is associated with statin-induced side effects. J Am Coll Cardiol. 2009; 54:1609–16. [PubMed: 19833260]

[38]. Link E, Parish S, Armitage J, et al. SLCO1B1 variants and statin-induced myopathy--a genomewide study. N Engl J Med. 2008; 359:789–99. [PubMed: 18650507]

[39]. Roden DM, Viswanathan PC. Genetics of acquired long QT syndrome. J Clin Invest. 2005; 115:2025–32. [PubMed: 16075043]

[40]. Kjolby M, Andersen OM, Breiderhoff T, et al. Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. Cell Metab. 2010; 12:213–23. [PubMed: 20816088]

Table 1

Selected Sections in The Gradient of Effect Sizes of DNA Sequence Variants

Type	Frequency	Effect size	Clinical implications
Disease-causing variants	Rare	Large	Accurate diagnosis, Pre-clinical diagnosis Risk-stratification? Prognostication?
Likely-disease causing variants	Very uncommon	Moderate	Accurate diagnosis, Pre-clinical diagnosis
Disease-Associated Variants	Uncommon	Small to moderate	Overall small but varies according to the effect size
Functional variants not associated with a disease	Common	Small	Minimal
Variants with no known function	Very common	Minimal to ?	Minimal to ?

Table 2

Indicators of Biological and Clinical Significance of the DNA Sequence Variants

Indicator	Class of Variants
Evidence of genetic linkage Co-segregation with the phenotype in large families (considering compound mutations and incomplete penetrance)	Causal variants Likely causal variants
Absent in phenotypically normal individuals matched for ethnicity	Causal variants Likely causal variants
Absent in dbSNP (unless MAF less than the prevalence of the phenotype)	Causal variants Likely causal variants
Known mutations in mutation databases	Causal variants Likely causal variants
Recurring mutations in independent families	Causal variants Likely causal variants
Multiple mutations in the same gene, known to cause the phenotype of interest	Causal variants Likely causal variants Disease-associated variants
Evolutionary conservation of the affected amino acid	Causal variants Likely causal variants Disease-associated variants Functional variants
Predicted biochemical and biological functions	Causal variants Likely causal variants Disease-associated variants Functional variants