Hum Genet (2013) 132:603–617
DOI 10.1007/s00439-013-1286-3

REVIEW PAPER

Hunting human disease genes: lessons from the past, challenges for the future

Liam R. Brunham · Michael R. Hayden

Received: 27 November 2012 / Accepted: 23 February 2013 / Published online: 17 March 2013

Abstract The concept that a specific alteration in an individual's DNA can result in disease is central to our notion of molecular medicine. The molecular basis of more than 3,500 Mendelian disorders has now been identified. In contrast, the identification of genes for common disease has been much more challenging. We discuss historical and contemporary approaches to disease gene identification, focusing on novel opportunities such as the use of population extremes and the identification of rare variants. While our ability to sequence DNA has advanced dramatically, assigning function to a given sequence change remains a major challenge, highlighting the need for both bioinformatics and functional approaches to appropriately interpret these data. We review progress in mapping and identifying human disease genes and discuss future challenges and opportunities for the field.

L. R. Brunham (✉)
Department of Medicine, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, Canada
e-mail: liam@cmmt.ubc.ca

L. R. Brunham · M. R. Hayden
Translational Laboratory for Genetic Medicine, National University of Singapore and the Association for Science, Technology and Research (A*STAR), Singapore, Singapore

M. R. Hayden
Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, Canada
e-mail: mrh@cmmt.ubc.ca

Introduction

A principal aim of medical genetics is the identification of the specific genes that, when altered, result in human disease. Most of the success in this endeavor has occurred in the context of Mendelian disorders—genetic diseases thought to reflect the action of a single gene-product with major effect. This group of disorders are recognizable in families by their adherence to one of the canonical patterns of inheritance first described by Gregor Mendel in 1865 and re-discovered in the early 1900s: autosomal recessive, autosomal dominant, co-dominant or sex-linked. Mendelian disorders are individually rare and affect less than 5 % of the population but are extraordinarily numerous. More than 7,000 such disorders have been described (Online Mendelian Inheritance in Man 2012), and countless other "private syndromes" may exist in only small numbers of individuals or even single families.

More than 3,600 Mendelian disorders, or ~50 % of all those described, have now been associated with a specific molecular defect (Online Mendelian Inheritance in Man 2012) (Fig. 1), speaking to both the remarkable success and the ongoing opportunity that exists in human disease gene mapping. Pathogenic mutations have been described in slightly more than 5,000 human genes (Human Gene Mutation Database. http://www.hgmd.cf.ac.uk/ac/hahaha.php). However, data from other organisms suggest that a much higher percentage of the ~22,000 human genes may be associated with phenotypes when altered. For example, systematic gene mutation screens in yeast and mice suggest that most of the genes in those organisms are non-lethal and associated with a discernible phenotype when disrupted (Ayadi et al. 2012; Winzeler et al. 1999; Hillenmeyer et al. 2008).

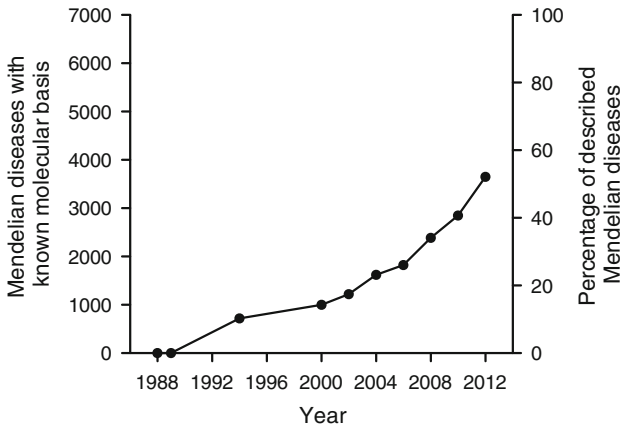

Fig. 1 Mendelian diseases of known molecular basis. The *x*-axis shows the time in years from 1988 to present. The left-hand *y*-axis indicates the cumulative number of Mendelian diseases for which a molecular basis is identified and the right-hand *y*-axis expresses this as a percentage of the approximately 7,000 Mendelian disorders that have been described. The first disease Mendelian disease gene to be cloned was the CFTR transporter involved in cystic fibrosis in 1989. Following the release of the first draft of the human genome sequence in 2001, the rate of discovery of Mendelian disease genes increased dramatically. As of November 2012, the molecular basis of 3,650 Mendelian diseases, or slightly more than 50 % of all Mendelian diseases, is known. Data are adapted from (McKusick 2007; Antonarakis and Beckmann 2006; Hamosh et al. 2005; Pearson et al. 1994; Online Mendelian Inheritance in Man 2012)

Mapping genes by positional cloning

For most of the modern era of human genetics, the principal method for the identification of disease-associated genes was positional cloning using linkage analysis. In this methodology, one or more pedigrees in which the trait of interest is observed to segregate are used for study. DNA from both affected and unaffected individuals are genotyped for polymorphic markers spread throughout the genome. Making use of the recombination that occurs in meiosis, one can then identify a chromosomal region that, based on the presumed mode of inheritance, shows segregation of a disease-associated haplotype in affected individuals, and of a non-disease-associated haplotype in unaffected individuals.

Linkage analysis was first described in fruit flies 100 years ago (Sturtevant 1913) but it was not until the discovery of naturally occurring polymorphic DNA markers in the 1980s that this tool first became available to map human disease genes (Botstein et al. 1980). One of the early successes was mapping of the gene for Huntington disease in 1983 (Gusella et al. 1983). Other examples that soon followed included the genes for cystic fibrosis (Rommens et al. 1989; Knowlton et al. 1985; White et al. 1985), Duchenne muscular dystrophy (Koenig et al. 1987) and many others.

To proceed from an area of linkage to a disease-associated mutation requires knowledge of the specific genes that exist in that chromosomal region. In the pre-human genome era, this involved examination of additional recombinant chromosomes to further refine the linkage interval and a variety of molecular biology techniques to identify expressed genes in the region and search for mutations. The Herculean nature of this task is demonstrated by the fact that it took international research groups a decade to go from identifying the Huntington disease locus to cloning the disease gene and identifying pathogenic mutations, in that case a triplet expansion in the *HTT* gene (MacDonald et al. 1993). With the completion of the initial sequencing of the human genome (Lander et al. 2001; Venter et al. 2001), this task has been tremendously facilitated with a corresponding acceleration in the identification of Mendelian disease genes (Fig. 1). From the 1980s to the year 2000, the molecular basis of approximately 1,000 Mendelian disorders had been discovered. In the first 10 years following the publication of the human genome, this number more than tripled to over 3,000, representing one of the indisputable fruits of the Human Genome Project with direct impact to the patients affected by these disorders and their families.

A form of linkage analysis that can be used in consanguineous families with suspected autosomal recessive traits is homozygosity mapping. This makes use of the principle that DNA markers in the chromosomal region immediately adjacent to the disease locus should be homozygous by descent in such cases (Lander and Botstein 1987). This strategy can efficiently identify genomic regions in which candidate genes can then be tested for the presence of pathogenic mutations. Homozygosity mapping has recently been used in combination with high-density whole genome genotyping to identify disease genes in patients in whom homozygosity by descent is suspected (Molho-Pessach et al. 2012).

The biological and medical importance of these disease-gene discoveries cannot be overstated. Once a disease gene for a Mendelian disorder is identified, an enormous amount of information about the biological function of that gene is provided by the phenotype of individuals in whom it is dysfunctional. Conversely, study of the biological pathways that the gene-product is involved in illuminates disease pathophysiology. From a clinical perspective, identification of a disease gene opens the door to diagnostic and predictive testing where appropriate. For instance, predictive testing for Huntington disease became available in Canada in 1987 using linkage analysis based on localization of the Huntington disease gene (Fox et al. 1989).

Moreover, mutation identification can directly lead to therapeutic insight and already significant advances in targeted therapies have been achieved by understanding

genotype. For instance, cystic fibrosis (CF) is caused by a variety of different mutations in the chloride transporter encoded by the *CFTR* gene. Some mutations, such as the common ΔF508 mutation, result in inability of the encoded protein to reach the plasma membrane. Other mutations, such as G551D which is present in ~5 % of patients, are associated with transport of the protein to the cell surface but failure of the channel to open. Ivacaftor is a novel small molecule that potentiates the *CFTR* channel at the cell surface (Eckford et al. 2012) and leads to significant improvements in lung-function in patients with the G551D mutation (Ramsey et al. 2011). In January 2012, this drug was approved by the Food and Drug Administration for treatment of CF in patients who carry this mutation, making this the first genotype-specific therapy for CF.

A second example involves lipoprotein lipase deficiency, a condition characterized by extremely high levels of triglycerides in plasma and recurrent attacks of painful pancreatitis. Identification of the *LPL* S447X mutation and understanding of its biochemical phenotype as a gain-of-function variant (Ross et al. 2005) ultimately led to the development of the first approved gene therapy product for humans (Yla-Herttuala 2012). Both of these examples illustrate how the understanding of genotype can lead to profound clinical insight, both by guiding appropriate patient selection and by directly leading to the development of new therapeutic products.

Role of next-generation sequencing in disease gene identification

The development of massively paralleled (next generation) sequencing has lead to dramatic acceleration in the pace of genetic discovery. These technologies have enabled two major advances of relevance to the discovery of disease-associated genes. The first is the ability to readily sequence the genome of a single person, thus allowing identification of mutations specific to that individual (previous "human genomes" represented consensus sequences of DNA from several individuals and were, therefore, not suited to the identification of rare mutations). The second major application of next-generation sequencing is the ability to perform whole exome sequencing (WES) in which a targeted capture strategy is used to sub-select the protein-coding exonic portion of the genome and generate sequence data of all known genes (Teer and Mullikin 2010). This results in sequence data covering ~30 MB of the genome, or ~1 % of that examined by whole genome sequencing (WGS), enabling the identification of mutations that result in changes to the amino-acid sequence of encoded proteins while substantially reducing the computational requirements associated with analyzing the resulting data. Examining only the exonic portion of the genome is justified on the basis that the vast majority of Mendelian disease-associated mutations identified by positional cloning strategies result in disruption of the protein-coding sequence (Stenson et al. 2009).

Sequencing of human exomes was first reported in 2009 (Ng et al. 2009) and the use of this technology to discover the genetic cause of a Mendelian disorder, Miller syndrome, followed soon after (Ng et al. 2010). Importantly, this demonstrated that WES makes tractable those conditions that are too rare and in which appropriately sized families are not available for positional cloning strategies, illustrating the power of this approach in situations where only small numbers of affected individuals are available for study.

WES and WGS strategies have now been used to elucidate the molecular etiology of an ever-expanding list of Mendelian disorders, as reviewed elsewhere (Gonzaga-Jauregui et al. 2012; Ku et al. 2011; Bamshad et al. 2011). While this undoubtedly represents a major advance for mapping human disease genes, several challenges remain. Capture methods remain imperfect and can result in unequal depth of coverage at different exonic regions. Our incomplete annotation of all human genes results in a necessarily incomplete view of the human exome. The analytical challenges are also significant. Each sequenced exome results in 20,000–25,000 variants relative to the reference sequence (Bamshad et al. 2011). How do we go from this huge number of variants to a single pathogenic mutation? Typically a number of filtering steps are employed (Bamshad et al. 2011), for instance, by ruling out any variants found in public databases such as dbSNP, 1000 Genomes Project, HapMap or locally available exome databases (Stitziel et al. 2011). This rests on the assumption that the causative variant will be extremely rare and that any individual with the variant will be affected (i.e., the variant is fully penetrant). While this is an efficient method to rule out the vast majority of variants, and is reasonable in scenarios where the disease-causing variant is hypothesized to be novel, it may be overly restrictive. As more genome and exome sequences are deposited in public repositories, the assumption that any variant found in these databases can be ruled out as disease-causing becomes increasingly difficult to justify. This is especially the case for recessive conditions in which the carrier state is relatively common (e.g., *HFE* gene mutations associated with hereditary hemochromatosis) or in situations in which the causative mutation is not fully penetrant. As a result, many groups now set frequency-based tolerances for variants found in dbSNP or other databases, such that variants present in fewer than 1 % of chromosomes are carried forward as candidates for a rare pathogenic mutation (McDonald et al. 2012).

An additional method to reduce the complexity of exome sequence data is through integration with family data or with traditional linkage analysis. For example, if two affected first cousins are used, the number of variants to be considered as candidates is reduced to the roughly 1/8th of the genome shared by two such individuals (Fig. 2). Similarly, the use of two affected second cousins reduces the number of shared variants to 1/32nd of the genome.

Integration of genome or exome sequencing with linkage analysis is also a powerful approach that can narrow the list of candidate genes. This approach has been successfully used to identify mutations associated with metachondromatosis (Sobreira et al. 2010), osteogenesis imperfect (Volodarsky et al. 2013; Cho et al. 2012), Chartcot-Marie-Tooth disease (Kennerson et al. 2013), spinocerebellar ataxia (Lee et al. 2012), opsismodysplasia (Below et al. 2013), distal arthrogryposis (McMillin et al. 2013), craniocervical dystonia (Charlesworth et al. 2012), dyskinesia and facial myokymia (Chen et al. 2012), thoracic aortic aneurysm syndrome (Boileau et al. 2012), distal hereditary motor neuropathy (Beetz et al. 2012), spinal muscular atrophy (Zhou et al. 2012), familial pityriasis rubra pilaris (Fuchs-Telem et al. 2012) and many others (Ku et al. 2011). In many cases, this approach enables gene mapping in only a single affected individual and can substantially narrow the list of candidate mutations generated by WES/WGS. For instance, in a study of metachondromatosis, the inclusion of linkage data reduced the number of candidate genes with disruptive mutations from 109 across the whole genome to only one found in a linkage region (Sobreira et al. 2010).

These tremendous technological breakthroughs raise the tantalizing possibility that in the near future we may discover the molecular etiology of most, or even all, Mendelian diseases. Achieving this remarkable goal will require world-wide collaboration bringing together clinicians caring for patients with rare Mendelian diseases and experts in genome technologies as well as computational biology. Consortia such as the Finding of Rare Disease Genes (FORGE) in Canada, the International Rare Diseases Research Consortium in Europe and the Centers for Mendelian Genomics by the National Institutes of Health in the United States (Bamshad et al. 2012) should have a major impact on applying genomic technologies to unraveling Mendelian disorders.

Unexpected phenotypes as a clue to modifier or suppressor mutations

The widespread availability of sequence data also permits identification of unexpected splits between genotype and phenotype that may suggest the presence of a suppressor or modifier mutation. Examples of this include individuals

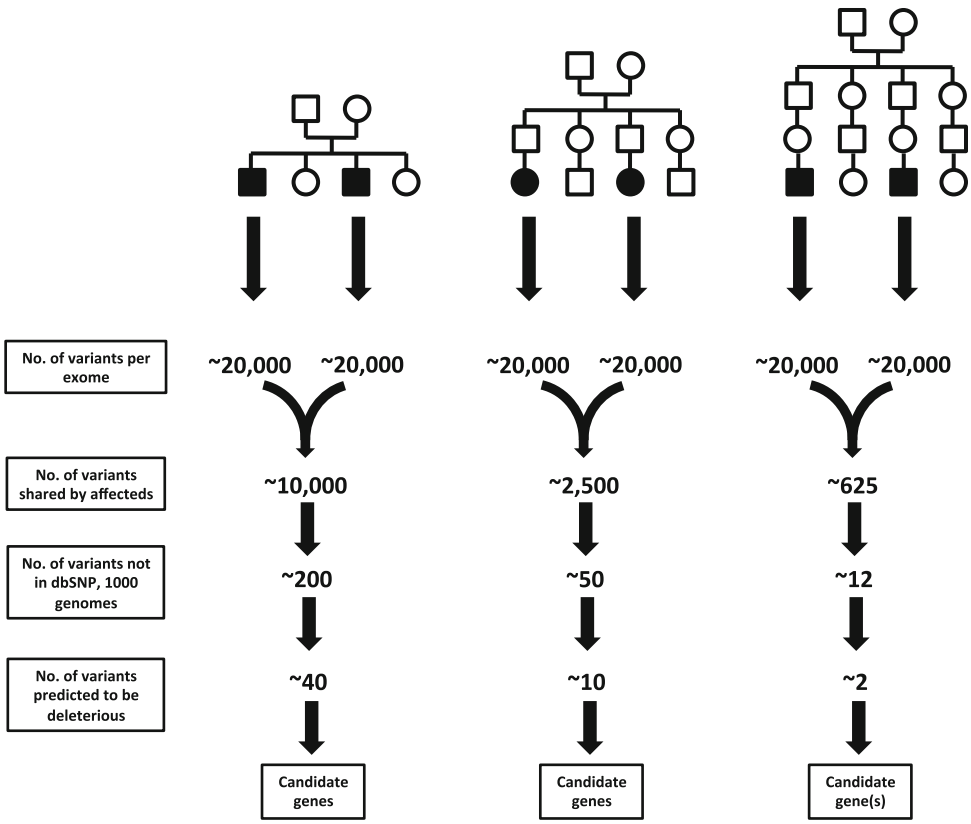

Fig. 2 Theoretical example of filtering steps used to limit number of variants identified by exome sequencing. The three examples indicate two affected individuals who are related to different degrees. Comparing the exomes of siblings, 1st cousins or 2nd cousins will limit candidate variants to the roughly 1/2th, 1/8th or 1/32nd of the genome, respectively, that is shared by two such individuals. Additional discrete filtering and bioinformatics steps can further reduce the number of candidate variants. Examples assume that all but 2 % of variants will be identified in public databases such as dbSNP and 1000 Genomes Project and that approximately 20 % of novel variants will be predicted to be deleterious to the function of the encoded protein

with pathogenic mutations in the LDL receptor gene but normal cholesterol levels (Hobbs et al. 1989), or patients with an expanded CAG repeat in the Huntington disease gene who have not manifest disease by the 95th percentile of age expected for that CAG repeat length, or conversely those that have manifest disease prior to the 5th percentile of age expected (Langbehn et al. 2004; Brinkman et al. 1997). Other examples include individuals homozygous for the ApoE ε4 allele who remain free of Alzheimer's late into life, or individuals homozygous for null alpha-1 antitrypsin alleles (i.e., ZZ genotype) who do not develop obstructive airways disease. All of these examples point to the presence of a mutation in another gene that masks the expected phenotype.

We recently described two families with mutations in the *SCARB1* gene that encodes the HDL receptor, SR-BI (Brunham et al. 2011) and was recently reported to be a cause of high HDL-C levels in humans (Vergeer et al. 2011). While most individuals who carried the *SCARB1* mutations had elevated HDL cholesterol (HDL-C) (>95th percentile for age and gender), one mutation-carrier was observed to have unexpectedly low HDL-C (15th percentile). Sequencing of candidate low HDL-C genes in this individual led to the identification of novel mutation in *ABCA1*, V2091I, that segregated with low HDL-C and may be the cause of this individual's low HDL-C thereby explaining the unexpected phenotype in that patient.

As more genome and exome data become available for a larger and larger number of individuals, we will have increasing opportunity to identify unexpected phenotypes in the presence of a given genotype. The importance of these observations is in pointing to the presence of potential suppressor genes that in many instances may represent novel therapeutic targets. For instance, individuals with familial hypercholesterolemia (FH)-causing mutations but normal levels of LDL-C suggest the presence of a mechanism for lowering LDL-C that is independent of the LDL receptor and would, therefore, be effective in individuals with FH, one of the most common causes of inherited high cholesterol. The identification of such suppressor mutations may lead to novel approaches to modify the course of illness by identifying therapeutic targets that have already been validated in relevant human disease models.

Disease genes for common diseases

In contrast to the remarkable success achieved in the identification of genes for Mendelian diseases, common diseases have proved much more difficult to unravel. Linkage studies for complex disease proved extremely difficult due to a lack of sufficient genomic resolution to identify disease-associated loci using microsatellite markers and inadequate power to detect an association, largely due to the significant locus heterogeneity that characterizes common disease (John et al. 2004; Altmuller et al. 2001; Xiong and Guo 1998). An alternative approach was therefore developed, genetic association studies, in which the frequency of common DNA polymorphisms is compared in unrelated cases versus controls. Association studies provide much greater power to detect variants associated with common disease than does linkage analysis, particularly when the risk conferred by the gene is modest (Risch and Merikangas 1996). The DNA marker studied need not be causal for the disease in question; because of patterns of linkage disequilibrium (LD) in the genome, nearby DNA markers tend to be inherited together. Indeed, the discovery that recombination tends to occur at specific "hot spots" (McVean et al. 2004; Daly et al. 2001) suggested that a single "tag" single nucleotide polymorphism (SNP) could capture most of the common DNA variation in a particular genomic region (Johnson et al. 2001).

Initial association studies were limited to examining candidate genes and therefore were not suited to the identification of novel genetic risk loci (Tabor et al. 2002). Multiple technical and conceptual advances changed this, including the development of high-throughput genotyping technologies capable of genotyping thousands of SNPs in large numbers of individuals, large catalogues of SNPs (Sachidanandam et al. 2001), knowledge of patterns of LD in the genome (International HapMap Consortium 2003) and analytic frameworks to analyze enormous datasets. By 2006, genome-wide association studies (GWAS) became a reality (Hirschhorn and Daly 2005). GWAS involves genotyping high-density panels of polymorphisms in several thousand cases and controls to identify common variants (>1 % allele frequency) that are hypothesized to underlie some portion of the heritability of common diseases. The basis for these studies is the common disease–common variant hypothesis which posits that genetic risk for common diseases will be due at least in part to a small number of disease-predisposing alleles per locus that exist at high frequency (Lander 1996; Chakravarti 1999).

More than 1,000 GWAS have now been reported with evidence of association of thousands of SNPs for dozens of common conditions (Hindorff et al. 2012) and has led to substantial advances in our understanding of the role of common variation in common disease (Altshuler et al. 2008). In general, GWAS has been successful in identifying numerous variants that are associated with common disease at high levels of statistical significance; however, most of these alleles confer only small effects sizes (Odds ratios <1.5). As such, GWAS has largely not succeeded in identifying disease genes with large effects, raising the concept of the "missing heritability" of common disease

(Manolio et al. 2009). Rare variants or structural variants not represented on GWAS genotyping panels may be a source of this missing heritability (Manolio et al. 2009). Alternatively, estimates of heritability may themselves be inaccurate, for instance, due to the effect of gene–gene interactions which are generally not incorporated into heritability estimates and would have the effect of increasing the heritability apparently left to account for (Zuk et al. 2012).

How will we move forward to uncover the remaining heritability of these common conditions? One exciting approach is the use of genotyping and next-generation sequencing technologies to identify rare and private variants. The alternative to the common disease–common variant hypothesis is that most variants underlying disease in humans will be individually rare because any variant with a major effect on fitness will tend to be removed from the population by the actions of natural selection, thus keeping the frequency of such variants very low. Indeed, rare variants are 4 times more likely to be functional (based on bioinformatics predictions) than are more common variants (minor allele frequency greater than 0.5 %) (Tennessen et al. 2012). Most of the variation in the human genome consist of common variants that arose prior to the migration of modern humans out of Africa and are shared between continental populations—these are the variants represented in HapMap (International HapMap Consortium 2003) that form the basis for GWAS. However, the vast majority of the coding, polymorphic alleles in humans are individually rare and represent evolutionarily recent mutations that have occurred with the rapid population expansion over the past 10,000 years (Tennessen et al. 2012). If common human diseases are characterized genetically by marked allelic and locus heterogeneity, as has been postulated (McClellan and King 2010), then new methods will be required to identify the rare alleles that contribute to human disease.

A study of sick sinus syndrome (SSS) in the Icelandic population is illustrative of how these rare variants can be identified. These investigators initially performed GWAS in a population of ~800 cases and nearly 40,000 population-based controls to define a region on chromosome 14q11 that contained 3 SNPs associated with SSS at genome-wide levels of significance. Subsequent WGS in a selection of 87 individuals identified 11 million variants that were then imputed into the GWAS cases and controls using long-range haplotype phasing (Kong et al. 2008). This approach identified a rare coding variant (minor allele frequency 0.38 % in Icelandic population) in the *MYH6* gene, located at the 14q11 locus, that was associated with SSS with an Odds ratio of 12.5 and p value of 1.5×10^{-29} (Holm et al. 2011). This association was further validated by direct genotyping in an additional set of SSS cases and controls. This study provides compelling evidence of a rare variant with large effect size that impacts a complex trait and provides impetus for efforts to identify such variants for common diseases broadly. This study also suggests that combining GWAS, WGS and imputation in a population isolate serves as one method by which this class of variant can be identified.

Imputation of rare variants identified by WGS into patients who have been genotyped for GWAS panels provides a cost-effective method for performing association studies of a large number of rare, potentially functional variants. This approach was used, also in the Icelandic population, to identify a missense substitution (R47H) in the *TREM2* gene, with an allele frequency of 0.12–0.63 % in the various populations studied, that confers a threefold increased risk of Alzheimer's disease (Jonsson et al. 2012). The same *TREM2* variant was simultaneously identified by a candidate gene sequencing approach and was found to occur significantly more frequently in individuals with Alzheimer's compared to controls (Guerreiro et al. 2012). The imputation of rare variants provides one attractive means to extend the unbiased nature of GWAS to low-frequency variants selected for potential functionality. Similarly, "exome-chip" approaches will soon enable direct genotyping of several hundred thousand putatively functional low-frequency coding variants identified by WGS or WES (Kathiresan and Srivastava 2012).

Ultimately, the most comprehensive approach for searching for rare variants that influence common disease will be WGS or WES of large numbers of cases and controls to identify variants associated with these phenotypes. Though this strategy remains both technically and economically prohibitive at present, it will no doubt be deployed in the near future. A key consideration of such studies would be the statistical power to detect an association. Because power decreases as a function of allele frequency, very large sample sizes will likely be necessary to identify disease-association of rare variants. With several hundred cases and controls, very few genes have adequate power to identify a rare variant that confers an Odds ratio of 5 or more (Tennessen et al. 2012). For smaller effect sizes, such as 1.5, several thousand cases and controls will be required (Raychaudhuri 2011). Accepted levels of statistical significance that take into account the burden of multiple testing for rare variants across the genome or exome need to be established and will be more stringent than levels used in GWAS reflecting the far greater number of rare than common variants in the genome; a p value of 10^{-11} has been proposed (Raychaudhuri 2011). Issues of incomplete penetrance, genetic heterogeneity, and gene–gene and gene-environment interactions will be additional complications requiring close attention.

Using population extremes to identify disease genes

An additional strategy to identify rare variants that confer disease risk involves studying individuals representing the ends of a quantitative phenotype, a strategy known as "sequencing the extremes" (Fig. 3). The prototypical study to establish this approach involved sequencing 3 candidate genes for low levels of HDL-C, *ABCA1*, *APOA1* and *LCAT*, in individuals with the lowest and highest 5 % of HDL-C levels in a population (Cohen et al. 2004). The number of non-synonymous sequence variants in these genes that were unique to either the low or high HDL-C groups was compared and revealed a statistically significant excess of rare mutations among individuals with low HDL-C (20 versus 3 in the low versus high HDL-C groups). Most of these mutations were in the *ABCA1* gene (Cohen et al. 2004).

This method has subsequently been used to provide evidence that rare variants in specific genes influence levels of LDL-C (Cohen et al. 2005) and intestinal sterol absorption (Cohen et al. 2006; Fahmi et al. 2008), triglycerides (Hobbs et al. 1989; Romeo et al. 2007), body mass (Ahituv et al. 2007) and high levels of HDL-C (Brunham et al. 2011; Tietjen et al. 2012). A re-sequencing study of four genes implicated by GWAS to be associated with hypertriglyceridemia demonstrated that these same genes harbor an excess of rare variants, with 28.1 % of hypertriglyceridemic individuals carrying a rare variant in one of these genes compared to 15.3 % of individuals with normal triglyceride levels (Johansen et al. 2010). Importantly, this demonstrates that some portion of the variability of common disease is likely due to rare variants in the same genes in which common variation has been associated with these traits.

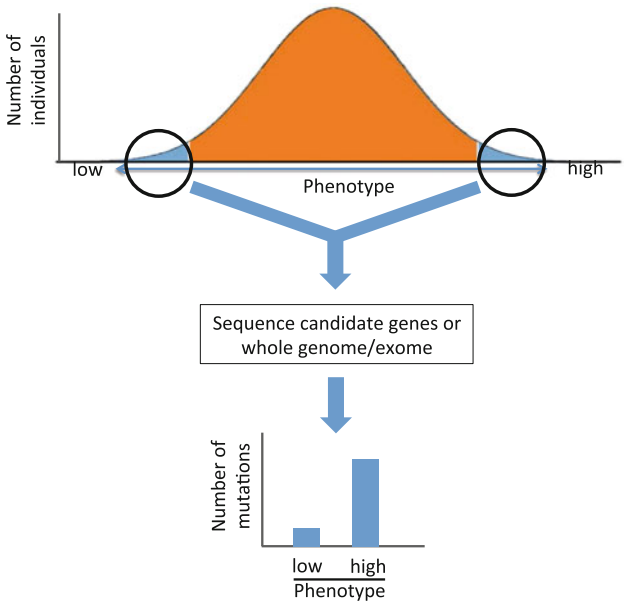

Fig. 3 Sequencing the extremes to identify rare variants involved in common disease. The individuals from the high and low ends of the population distribution of a trait are chosen for study. Either candidate genes or whole exomes or genomes are sequenced and the number of rare variants in a given gene is compared between the high and low groups. An excess of rare sequence variants in a given gene in the high versus low (or vice versa) group provides evidence for a role of that gene in the phenotype under study. See text for examples of where this has been used

Sequencing the extremes has emerged as a powerful tool to provide evidence for the role of rare mutations in specific genes impacting complex traits. However, a limitation of this approach has been that it is restricted to candidate genes. For instance, in the study of low levels of HDL-C described above, the *ABCA1*, *APOA1* and *LCAT* genes were chosen because mutations in each of these genes are known to cause rare, recessive Mendelian conditions of low HDL-C. To identify novel disease genes, this approach would need to be extended to perform WGS or WES in the population extremes of a trait.

Indeed, such a strategy has recently been reported to discover a gene that impacts *Pseudomonas* infection in individuals with cystic fibrosis (Emond et al. 2012). Most patients with CF become colonized with *Pseudomonas* early in life with a median of 1 year (Li et al. 2005), but substantial variability exists in this phenotype. Emond et al. (2012) identified two groups of patients representing the extremes of *Pseudomonas* colonization: those who became chronically infected prior to the 10th percentile age of onset and those who remained culture negative beyond 14 years of age. Exome sequencing of these individuals identified a single gene, *DCTN4*, that showed statistically significant evidence of association with time to *Pseudomonas* infection after correcting for multiple testing. This study suggests that the extreme phenotype approach can be successfully combined with exome sequencing to identify novel genes involved in complex traits. More studies are clearly needed to determine if this will prove to be a successful approach in general and new analytic methods are required to assess the statistical significance of rare variants identified through such approaches.

De Novo mutations and copy number variants in human disease

The availability of WGS and WES also affords the opportunity to delineate DNA variants that are present in the genome or exome of an offspring but absent in that individual's parents. This trio-sequencing approach not only enables accurate determination of the mutation rate but also importantly allows for precise identification of de novo mutations at single nucleotide resolution (Veltman

and Brunner 2012). De novo mutations are, in a sense, the most extreme form of rare mutations, in that they may be private and have not been subject to selective pressure in previous generations. These variants are therefore prime suspects for playing a role in human disease.

Several recent studies have examined the contribution of de novo mutations to common disease. Four recent studies reported exome sequencing of 'trios' (affected child and parents) or 'quads' (affected child, unaffected sibling and parents) in families with autism spectrum disorder (ASD) (Neale et al. 2012; O'Roak et al. 2012; Sanders et al. 2012; Iossifov et al. 2012). These studies found a small increase in de novo mutations in ASD cases compared to unaffected controls that tended to occur in genes that are biologically related to each other or to previously known ASD genes. These mutations were predominantly paternal in origin, and the number of mutations increased with increasing paternal age, consistent with the known increased risk of ASD to children of older fathers.

While the ability to accurately detect such mutations opens the door to many exciting lines of discovery that may account for some of the missing heritability of common disease, these studies also highlight many challenges. For instance, how do we interpret the biological significance of these mutations and link them with disease? In particular, how do we distinguish a single de novo mutation, even an apparently deleterious one, from the substantial background of mutational events that these studies have demonstrated? These studies also highlight the extreme genetic heterogeneity of these disorders. For example, of the more than 120 genes implicated in the studies of autism, only six genes had disruptive mutations in more than one individual (Muers 2012). While this may suggest that ASD represents hundreds of genetically distinct entities, it remains unclear to what extent this sub-classification will yield meaningful clinical insight.

One class of genetic variation with a large contribution from de novo events is structural variation that affects the number of copies of a particular chromosomal region and can thereby impact phenotypes via a gene-dosage effect of the implicated gene or genes. Our understanding of the diversity of copy number variants (CNVs) has advanced dramatically in recent years and we now know that CNVs are remarkably common in the genomes of normal individuals (Sebat et al. 2004; Iafrate et al. 2004). While some CNVs are polymorphic in human populations, most are rare or private. Most CNVs are thought to arise due to non-allelic homologous recombination between highly identical regions of the genome (Girirajan et al. 2011). CNVs affect an order of magnitude more base pairs in the genome than do SNPs (Lupski 2007), and more than one-third of all human genes may be partially or totally in the region of a CNV (Gonzaga-Jauregui et al. 2012). The Database of Genomic Variants currently listed more than 15,000 CNV loci in the human genome (The Database of Genomic Variants 2013).

CNVs have long been associated with 'genomic disorders'—rare highly penetrant syndromes typically associated with neurodevelopmental delay (Girirajan et al. 2011). Rare CNVs have also been documented to play an important role in psychiatric and behavioral conditions such as schizophrenia and ASD (Xu et al. 2008; Walsh et al. 2008; Sebat et al. 2007). More recently, a CNV at 17p was shown to influence obesity traits in mice, with deletion of this region leading to obesity and metabolic syndrome, and duplication leading to protection from obesity (Lacaria et al. 2012). CNVs have also been reported to play a role in congenital heart disease (Soemedi et al. 2012; Hitz et al. 2012), epilepsy (Helbig et al. 2009) and many other conditions (Zhang et al. 2009; Stankiewicz and Lupski 2010; Girirajan et al. 2011).

Comprehensive detection of CNVs in the human genome remains a challenge but recent technological advances have accelerated our ability to do so. In particular, array comparative genome hybridization and paired-end mapping, in which the presence of a CNV is suggested by a size difference between the fragment length and the corresponding region of the reference sequence (Korbel et al. 2007). Notably, these methods allow for accurate detection of smaller CNVs with more precise resolution of breakpoints. Widespread application of these techniques to an expanding list of common diseases should provide further insight into the role of CNVs in common genetic diseases.

What constitutes proof that a mutation causes a disease?

How do we prove that a mutation in a gene is causative of disease? Linkage analysis can provide evidence of DNA variants in a gene that segregate with disease in a family. Finding mutations in the same gene in unrelated families provide additional evidence of causality. For highly characteristic rare phenotypes that segregate in a family, the accepted standard is replication across three independent families. For variants identified in extremely rare or private phenotypes, the burden of proof is less well defined and new analytic methods are required. This underscores an important principle relevant to both traditional and contemporary gene mapping. Namely, that finding different, rare, pathogenic mutations in the same gene or in the same biological pathway, in unrelated individuals with the same phenotype provides important support for that gene being involved in a given biological process (McClellan and King 2010).

Bioinformatics algorithms such as Polyphen (Adzhubei et al. 2010), SIFT (Kumar et al. 2009), SNAP (Johnson et al. 2008) or PANTHER (Mi et al. 2010) are frequently used as a

first approximation for determining if a given DNA or amino-acid sequence variant is likely to impair the function of the encoded protein. Most of these programs rely on some combination of conservation of the site in question in related proteins within and across species as well as knowledge of structure of the protein. Functional data for a limited number of genes suggest that these methods have reasonable accuracy for the prediction of deleterious effects on specific proteins (Ng and Henikoff 2006). For example, PANTHER correctly identified as deleterious or benign ~95 % of rare variants in the *ABCA1* gene (Brunham et al. 2005). However, none of these methods are perfectly sensitive or specific. Indeed, in the exome sequencing study that identified the molecular cause of Miller syndrome, the use of Polyphen as a filter for functional variants would have excluded the mutation in *DHODH*, ultimately found to be causative of that disease (Ng et al. 2010). Moreover, agreement among the various methods are poor: in a deeply sequenced set of exome data, about half of all detected SNPs were predicted to be functional by at least one of seven different bioinformatics programs, but the various programs were in full agreement on only 1 % of variants (Tennessen et al. 2012).

Functional studies are, therefore, crucial to establish which DNA variants truly impact protein function and are causal in disease. This remains a major rate-limiting step, because, for most gene-products, a readily availably functional assay does not exist. For many genes, we simply do not know enough about the function of the gene-product and therefore lack methods for testing the consequence of sequence variation, especially in high-throughput fashion. Full functional annotation of all human genes remains a lofty and distant goal.

In GWAS, proof generally takes the form of a level of statistical significance. In particular, a p value threshold of 5×10^{-8} is felt to reflect the nominal significance threshold of a finding being due to chance 1 in 20 times corrected for the approximately 1 million independent tests incurred in a genome-wide scan (Pe'er et al. 2008). However, it is important to note that this provides evidence only of association and not of causation. Despite these high levels of statistical stringency, most variants identified by GWAS are in non-coding regions of the genome and the causal variant is often unknown. This may be explained by linkage disequilibrium between the associated variant and a true functional variant, or by a regulatory effect of a non-coding variant. Differentiating between these possibilities can be extremely challenging.

For example, a locus at 1p13 was known to be strongly associated with LDL cholesterol in a meta-analysis of >100,000 individuals (Teslovich et al. 2010). Fine-mapping of this region identified 6 SNPs that showed the greatest degree of association in the genomic region between the genes *CELSR2* and *PSRC1* (Musunuru et al. 2010) but the causative variant was not obvious. Sequential testing of these variants identified a single SNP that resulted in increased gene expression by creating a transcription factor binding site for *SORT1*, a gene shown to regulate hepatic VLDL secretion.

A second example involves variants in a gene desert at chromosome region 9p21 associated with coronary artery disease (McPherson et al. 2007). Through a series of computational and experimental approaches using immortalized cell lines, Harismendy et al. 2011 demonstrated that a single SNP in this region interrupts a binding site for *STAT1*, a transcription factor implicated in inflammatory responses, and that this locus participates in a long-range physical interaction with the *CDKN2A/B* locus. These two examples highlight how challenging it is to move from genetic association to mechanistic insight, but provide important models for how this may be achieved.

Conclusions

The past decade has witnessed tremendous progress in the identification of the molecular bases of human disease. Indeed, for Mendelian disorders, it is conceivable that in the future we may possess a complete catalogue of the genetic basis for each of these conditions. Common diseases continue to be much more challenging to address, but recent advances in sequencing and genotyping methods are yielding exciting results. Next-generation sequencing has dramatically altered our ability to identify human disease genes, but many challenges remain. In particular, our ability to ascribe functional significance to a given DNA variant remains limited. Ultimately, the identification of the molecular causes of disease will continue to illuminate the pathophysiology of both common and rare conditions and offer opportunities for improvements in the diagnosis, treatment and prevention of disease.

Acknowledgments This work was supported by the Biomedical Research Council (BMRC) of the Agency for Science, Technology and Research (A*STAR), Singapore and by the National University of Singapore. MRH is a Canada Research Chair in Human Genetics. LRB was supported by the Clinician Investigator Program at the University of British Columbia.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and

server for predicting damaging missense mutations. Nat Methods 7(4):248–249. doi:10.1038/nmeth0410-248

Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S, Doelle H, Ersoy B, Kryukov G, Schmidt S, Yosef N, Ruppin E, Sharan R, Vaisse C, Sunyaev S, Dent R, Cohen J, McPherson R, Pennacchio LA (2007) Medical sequencing at the extremes of human body mass. Am J Hum Genet 80(4):779–791. doi:10.1086/513471

Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M (2001) Genomewide scans of complex human diseases: true linkage is hard to find. Am J Hum Genet 69(5):936–950. doi:10.1086/324069

Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. Science 322(5903):881–888. doi:10.1126/science.1156409

Antonarakis SE, Beckmann JS (2006) Mendelian disorders deserve more attention. Nat Rev Genet 7(4):277–282

Ayadi A, Birling MC, Bottomley J, Bussell J, Fuchs H, Fray M, Gailus-Durner V, Greenaway S, Houghton R, Karp N, Leblanc S, Lengger C, Maier H, Mallon AM, Marschall S, Melvin D, Morgan H, Pavlovic G, Ryder E, Skarnes WC, Selloum M, Ramirez-Solis R, Sorg T, Teboul L, Vasseur L, Walling A, Weaver T, Wells S, White JK, Bradley A, Adams DJ, Steel KP, Hrabe de Angelis M, Brown SD, Herault Y (2012) Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. Mamm Genome 23(9–10):600–610. doi:10.1007/s00335-012-9418-y

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J (2011) Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12(11):745–755. doi:10.1038/nrg3031

Bamshad MJ, Shendure JA, Valle D, Hamosh A, Lupski JR, Gibbs RA, Boerwinkle E, Lifton RP, Gerstein M, Gunel M, Mane S, Nickerson DA (2012) The Centers for Mendelian Genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions. Am J Med Genet Part A 158A(7):1523–1525. doi:10.1002/ajmg.a.35470

Beetz C, Pieber TR, Hertel N, Schabhuttl M, Fischer C, Trajanoski S, Graf E, Keiner S, Kurth I, Wieland T, Varga RE, Timmerman V, Reilly MM, Strom TM, Auer-Grumbach M (2012) Exome sequencing identifies a REEP1 mutation involved in distal hereditary motor neuropathy type V. Am J Hum Genet 91(1):139–145. doi:10.1016/j.ajhg.2012.05.007

Below JE, Earl DL, Shively KM, McMillin MJ, Smith JD, Turner EH, Stephan MJ, Al-Gazali LI, Hertecant JL, Chitayat D, Unger S, Cohn DH, Krakow D, Swanson JM, Faustman EM, Shendure J, Nickerson DA, Bamshad MJ (2013) Whole-genome analysis reveals that mutations in inositol polyphosphate phosphatase-like 1 cause opsismodysplasia. Am J Hum Genet 92(1):137–143. doi:10.1016/j.ajhg.2012.11.011

Boileau C, Guo DC, Hanna N, Regalado ES, Detaint D, Gong L, Varret M, Prakash SK, Li AH, d'Indy H, Braverman AC, Grandchamp B, Kwartler CS, Gouya L, Santos-Cortez RL, Abifadel M, Leal SM, Muti C, Shendure J, Gross MS, Rieder MJ, Vahanian A, Nickerson DA, Michel JB, Jondeau G, Milewicz DM (2012) TGFB2 mutations cause familial thoracic aortic aneurysms and dissections associated with mild systemic features of Marfan syndrome. Nat Genet 44(8):916–921. doi:10.1038/ng.2348

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32(3):314–331

Brinkman RR, Mezei MM, Theilmann J, Almqvist E, Hayden MR (1997) The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. Am J Hum Genet 60(5):1202–1210

Brunham LR, Singaraja RR, Pape TD, Kejariwal A, Thomas PD, Hayden MR (2005) Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the *ABCA1* gene. PLoS Genet 1(6):e83. doi:10.1371/journal.pgen.0010083

Brunham LR, Tietjen I, Bochem AE, Singaraja RR, Franchini PL, Radomski C, Mattice M, Legendre A, Hovingh GK, Kastelein JJ, Hayden MR (2011) Novel mutations in scavenger receptor BI associated with high HDL cholesterol in humans. Clin Genet 79(6):575–581. doi:10.1111/j.1399-0004.2011.01682.x

Chakravarti A (1999) Population genetics—making sense out of sequence. Nat Genet 21(1 Suppl):56–60. doi:10.1038/4482

Charlesworth G, Plagnol V, Holmstrom KM, Bras J, Sheerin UM, Preza E, Rubio-Agusti I, Ryten M, Schneider SA, Stamelou M, Trabzuni D, Abramov AY, Bhatia KP, Wood NW (2012) Mutations in ANO3 cause dominant craniocervical dystonia: ion Channel implicated in pathogenesis. Am J Hum Genet 91(6):1041–1050. doi:10.1016/j.ajhg.2012.10.024

Chen YZ, Matsushita MM, Robertson P, Rieder M, Girirajan S, Antonacci F, Lipe H, Eichler EE, Nickerson DA, Bird TD, Raskind WH (2012) Autosomal dominant familial dyskinesia and facial myokymia: single exome sequencing identifies a mutation in adenylyl cyclase 5. Arch Neurol 69(5):630–635. doi:10.1001/archneurol.2012.54

Cho TJ, Lee KE, Lee SK, Song SJ, Kim KJ, Jeon D, Lee G, Kim HN, Lee HR, Eom HH, Lee ZH, Kim OH, Park WY, Park SS, Ikegawa S, Yoo WJ, Choi IH, Kim JW (2012) A single recurrent mutation in the 5′-UTR of IFITM5 causes osteogenesis imperfecta type V. Am J Hum Genet 91(2):343–348. doi:10.1016/j.ajhg.2012.06.005

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305(5685):869–872. doi:10.1126/science.1099870

Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet 37(2):161–165. doi:10.1038/ng1509

Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, Hobbs HH (2006) Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc Natl Acad Sci USA 103(6):1810–1815. doi:10.1073/pnas.0508483103

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29(2):229–232. doi:10.1038/ng1001-229

Eckford PD, Li C, Ramjeesingh M, Bear CE (2012) Cystic fibrosis transmembrane conductance regulator (CFTR) potentiator VX-770 (ivacaftor) opens the defective channel gate of mutant CFTR in a phosphorylation-dependent but ATP-independent manner. J Biol Chem 287(44):36639–36649. doi:10.1074/jbc.M112.393637

Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, Wright FA, Rieder MJ, Tabor HK, Nickerson DA, Barnes KC, Gibson RL, Bamshad MJ (2012) Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat Genet 44(8):886–889. doi:10.1038/ng.2344

Fahmi S, Yang C, Esmail S, Hobbs HH, Cohen JC (2008) Functional characterization of genetic variants in NPC1L1 supports the sequencing extremes strategy to identify complex trait genes. Hum Mol Genet 17(14):2101–2107. doi:10.1093/hmg/ddn108

Fox S, Bloch M, Fahy M, Hayden MR (1989) Predictive testing for Huntington disease: I. Description of a pilot project in British Columbia. Am J Med Genet 32(2):211–216. doi:10.1002/ajmg.1320320214

Fuchs-Telem D, Sarig O, van Steensel MA, Isakov O, Israeli S, Nousbeck J, Richard K, Winnepenninckx V, Vernooij M, Shomron N, Uitto J, Fleckman P, Richard G, Sprecher E (2012) Familial pityriasis rubra pilaris is caused by mutations in CARD14. Am J Hum Genet 91(1):163–170. doi:10.1016/j.ajhg.2012.05.010

Girirajan S, Campbell CD, Eichler EE (2011) Human copy number variation and complex genetic disease. Annu Rev Genet 45:203–226. doi:10.1146/annurev-genet-102209-163544

Gonzaga-Jauregui C, Lupski JR, Gibbs RA (2012) Human genome sequencing in health and disease. Annu Rev Med 63:35–61. doi:10.1146/annurev-med-051010-162644

Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, Cruchaga C, Sassi C, Kauwe JS, Younkin S, Hazrati L, Collinge J, Pocock J, Lashley T, Williams J, Lambert JC, Amouyel P, Goate A, Rademakers R, Morgan K, Powell J, St George-Hyslop P, Singleton A, Hardy J (2012) TREM2 variants in Alzheimer's disease. N Engl J Med. doi:10.1056/NEJMoa1211851

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY et al (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306(5940):234–238

Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledge base of human genes and genetic disorders. Nucleic Acids Res 33 (Database issue):D514–D517

Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature 470(7333):264–268. doi:10.1038/nature09753

Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, de Kovel C, Baker C, von Spiczak S, Kron KL, Steinich I, Kleefuss-Lie AA, Leu C, Gaus V, Schmitz B, Klein KM, Reif PS, Rosenow F, Weber Y, Lerche H, Zimprich F, Urak L, Fuchs K, Feucht M, Genton P, Thomas P, Visscher F, de Haan GJ, Moller RS, Hjalgrim H, Luciano D, Wittig M, Nothnagel M, Elger CE, Nurnberg P, Romano C, Malafosse A, Koeleman BP, Lindhout D, Stephani U, Schreiber S, Eichler EE, Sander T (2009) 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet 41(2):160–162. doi:10.1038/ng.292

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 320(5874):362–365. doi:10.1126/science.1150021

Hindorff LA, MacArthur J, Wise A, Junkins HA, P.N. H, Klemm AK, Manolio TA (2012) A Catalog of Published Genome-Wide Association Studies. http://www.genome.gov/gwastudies. Accessed 18 Sep 2012

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6(2):95–108. doi:10.1038/nrg1521

Hitz MP, Lemieux-Perreault LP, Marshall C, Feroz-Zada Y, Davies R, Yang SW, Lionel AC, D'Amours G, Lemyre E, Cullum R, Bigras JL, Thibeault M, Chetaille P, Montpetit A, Khairy P, Overduin B, Klaassen S, Hoodless P, Nemer M, Stewart AF, Boerkoel C, Scherer SW, Richter A, Dube MP, Andelfinger G (2012) Rare copy number variants contribute to congenital left-sided heart disease. PLoS Genet 8(9):e1002903. doi:10.1371/journal.pgen.1002903

Hobbs HH, Leitersdorf E, Leffert CC, Cryer DR, Brown MS, Goldstein JL (1989) Evidence for a dominant gene that suppresses hypercholesterolemia in a family with defective low density lipoprotein receptors. J Clin Investig 84(2):656–664. doi:10.1172/JCI114212

Holm H, Gudbjartsson DF, Sulem P, Masson G, Helgadottir HT, Zanon C, Magnusson OT, Helgason A, Saemundsdottir J, Gylfason A, Stefansdottir H, Gretarsdottir S, Matthiasson SE, Thorgeirsson GM, Jonasdottir A, Sigurdsson A, Stefansson H, Werge T, Rafnar T, Kiemeney LA, Parvez B, Muhammad R, Roden DM, Darbar D, Thorleifsson G, Walters GB, Kong A, Thorsteinsdottir U, Arnar DO, Stefansson K (2011) A rare variant in MYH6 is associated with high risk of sick sinus syndrome. Nat Genet 43(4):316–320. doi:10.1038/ng.781

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36(9):949–951. doi:10.1038/ng1416

International HapMap Consortium (2003) The International HapMap Project. Nature 426(6968):789–796. doi:10.1038/nature02168

Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekritsky M, Pradhan K, Ghiban E, Kramer M, Parla J, Demeter R, Fulton LL, Fulton RS, Magrini VJ, Ye K, Darnell JC, Darnell RB, Mardis ER, Wilson RK, Schatz MC, McCombie WR, Wigler M (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74(2):285–299. doi:10.1016/j.neuron.2012.04.009

Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA (2010) Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet 42(8):684–687. doi:10.1038/ng.628

John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, Eyre S, Jones KW, Ollier W, Silman A, Gibson N, Worthington J, Kennedy GC (2004) Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. Am J Hum Genet 75(1):54–64. doi:10.1086/422195

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29(2):233–237. doi:10.1038/ng1001-233

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24(24):2938–2939. doi:10.1093/bioinformatics/btn564

Jonsson T, Stefansson H, Ph DS, Jonsdottir I, Jonsson PV, Snaedal J, Bjornsson S, Huttenlocher J, Levey AI, Lah JJ, Rujescu D, Hampel H, Giegling I, Andreassen OA, Engedal K, Ulstein I, Djurovic S, Ibrahim-Verbaas C, Hofman A, Ikram MA, van Duijn CM, Thorsteinsdottir U, Kong A, Stefansson K (2012) Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med. doi:10.1056/NEJMoa1211103

Kathiresan S, Srivastava D (2012) Genetics of human cardiovascular disease. Cell 148(6):1242–1257. doi:10.1016/j.cell.2012.03.001

Kennerson ML, Yiu EM, Chuang DT, Kidambi A, Tso SC, Ly C, Chaudhry R, Drew AP, Rance G, Delatycki MB, Zuchner S, Ryan MM, Nicholson GA (2013) A new locus for X-linked dominant Charcot Marie Tooth Disease (CMTX6) is caused by mutations in the pyruvate dehydrogenase kinase isoenzyme 3 (*PDK3*) gene. Hum Mol Genet. doi:10.1093/hmg/dds557

Knowlton RG, Cohen-Haguenauer O, Van Cong N, Frezal J, Brown VA, Barker D, Braman JC, Schumm JW, Tsui LC, Buchwald M

et al (1985) A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. Nature 318(6044):380–382

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50(3):509–517

Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, Olason PI, Ingason A, Steinberg S, Rafnar T, Sulem P, Mouy M, Jonsson F, Thorsteinsdottir U, Gudbjartsson DF, Stefansson H, Stefansson K (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet 40(9):1068–1075. doi:10.1038/ng.216

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318(5849):420–426. doi:10.1126/science.1149504

Ku CS, Naidoo N, Pawitan Y (2011) Revisiting Mendelian disorders through exome sequencing. Hum Genet 129(4):351–370. doi:10.1007/s00439-011-0964-2

Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4(7):1073–1081. doi:10.1038/nprot.2009.86

Lacaria M, Saha P, Potocki L, Bi W, Yan J, Girirajan S, Burns B, Elsea S, Walz K, Chan L, Lupski JR, Gu W (2012) A duplication CNV that conveys traits reciprocal to metabolic syndrome and protects against diet-induced obesity in mice and men. PLoS Genet 8(5):e1002713. doi:10.1371/journal.pgen.1002713

Lander ES (1996) The new genomics: global views of biology. Science 274(5287):536–539

Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236(4808):1567–1570

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921. doi:10.1038/35057062

Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR (2004) A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin Genet 65(4):267–277. doi:10.1111/j.1399-0004.2004.00241.x

Lee YC, Durr A, Majczenko K, Huang YH, Liu YC, Lien CC, Tsai PC, Ichikawa Y, Goto J, Monin ML, Li JZ, Chung MY, Mundwiller E, Shakkottai V, Liu TT, Tesson C, Lu YC, Brice A, Tsuji S, Burmeister M, Stevanin G, Soong BW (2012) Mutations in KCND3 cause spinocerebellar ataxia type 22. Ann Neurol 72(6):859–869. doi:10.1002/ana.23701

Li Z, Kosorok MR, Farrell PM, Laxova A, West SE, Green CG, Collins J, Rock MJ, Splaingard ML (2005) Longitudinal development of mucoid Pseudomonas aeruginosa infection and lung disease progression in children with cystic fibrosis. JAMA 293(5):581–588. doi:10.1001/jama.293.5.581

Lupski JR (2007) Genomic rearrangements and sporadic disease. Nat Genet 39(7 Suppl):S43–S47. doi:10.1038/ng2084

MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, Barnes G, Taylor SA, James M, Groot N, MacFarlane H, Jenkins B, Anderson MA, Wexler NS, Gusella JF, Bates GP, Baxendale S, Hummerich H, Kirby S, North M, Youngman S, Mott R, Zehetner G, Sedlacek Z, Poustka A, Frischauf A-M, Lehrach H, Buckler AJ, Church D, Doucette-Stamm L, O'Donovan MC, Riba-Ramirez L, Shah M, Stanton VP, Strobel SA, Draths KM, Wales JL, Dervan P, Housman DE, Altherr M, Shiang R, Thompson L, Fielder T, Wasmuth JJ, Tagle D, Valdes J, Elmer L, Allard M, Castilla L, Swaroop M, Blanchard K, Collins FS, Snell R, Holloway T, Gillespie K, Datson N, Shaw D, Harper PS (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72(6):971–983

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. Nature 461(7265):747–753. doi:10.1038/nature08494

McClellan J, King MC (2010) Genetic heterogeneity in human disease. Cell 141(2):210–217. doi:10.1016/j.cell.2010.03.032

McDonald KK, Stajich J, Blach C, Ashley-Koch AE, Hauser MA (2012) Exome analysis of two limb-girdle muscular dystrophy families: mutations identified and challenges encountered. PLoS ONE 7(11):e48864. doi:10.1371/journal.pone.0048864

McKusick VA (2007) Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet 80(4):588–604

McMillin MJ, Below JE, Shively KM, Beck AE, Gildersleeve HI, Pinner J, Gogola GR, Hecht JT, Grange DK, Harris DJ, Earl DL, Jagadeesh S, Mehta SG, Robertson SP, Swanson JM, Faustman EM, Mefford HC, Shendure J, Nickerson DA, Bamshad MJ (2013) Mutations in ECEL1 cause distal arthrogryposis type 5D. Am J Hum Genet 92(1):150–156. doi:10.1016/j.ajhg.2012.11.014

McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC (2007) A common allele on chromosome 9 associated with coronary heart disease. Science 316(5830):1488–1491. doi:10.1126/science.1142447

McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P (2004) The fine-scale structure of recombination rate variation in the human genome. Science 304(5670):581–584. doi:10.1126/science.1092500

Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD (2010) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res 38((Database issue)):D204–D210. doi:10.1093/nar/gkp1019

Molho-Pessach V, Rios JJ, Xing C, Setchell KD, Cohen JC, Hobbs HH (2012) Homozygosity mapping identifies a bile acid biosynthetic defect in an adult with cirrhosis of unknown etiology. Hepatology 55(4):1139–1145. doi:10.1002/hep.24781

Muers M (2012) Human genetics: fruits of exome sequencing for autism. Nat Rev Genet 13(6):377. doi:10.1038/nrg3248

Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Koteliansky V, Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S, Rader DJ (2010) From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature 466(7307):714–719. doi:10.1038/nature09266

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, Cai G, Lihm J, Dannenfelser R, Jabado O, Peralta Z, Nagaswamy U, Muzny D, Reid JG, Newsham I, Wu Y, Lewis L, Han Y, Voight BF, Lim E, Rossin E, Kirby A, Flannick J, Fromer M, Shakir K, Fennell T, Garimella K, Banks E, Poplin R, Gabriel S, DePristo M, Wimbish JR, Boone BE, Levy SE, Betancur C, Sunyaev S, Boerwinkle E, Buxbaum JD, Cook EH Jr, Devlin B, Gibbs RA, Roeder K, Schellenberg GD, Sutcliffe JS, Daly MJ (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485(7397):242–245. doi:10.1038/nature11011

Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 7:61–80. doi:10.1146/annurev.genom.7.080505.115630

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461(7261):272–276. doi:10.1038/nature08250

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ (2010) Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet 42(1):30–35. doi:10.1038/ng.499

Online Mendelian Inheritance in Man (2012) http://www.ncbi.nlm.nih.gov/Omim/mimstats.html. Accessed October 1, 2012

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485(7397):246–250. doi:10.1038/nature10989

Pearson P, Francomano C, Foster P, Bocchini C, Li P, McKusick V (1994) The status of online Mendelian inheritance in man (OMIM) medio 1994. Nucleic Acids Res 22(17):3470–3473

Pe'er I, Yelensky R, Altshuler D, Daly MJ (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 32(4):381–385. doi:10.1002/gepi.20303

Ramsey BW, Davies J, McElvaney NG, Tullis E, Bell SC, Drevinek P, Griese M, McKone EF, Wainwright CE, Konstan MW, Moss R, Ratjen F, Sermet-Gaudelus I, Rowe SM, Dong Q, Rodriguez S, Yen K, Ordonez C, Elborn JS (2011) A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. N Engl J Med 365(18):1663–1672. doi:10.1056/NEJMoa1105185

Raychaudhuri S (2011) Mapping rare and common causal alleles for complex human diseases. Cell 147(1):57–69. doi:10.1016/j.cell.2011.09.011

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273(5281):1516–1517

Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC (2007) Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat Genet 39(4):513–516. doi:10.1038/ng1984

Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N et al (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245(4922):1059–1065

Ross CJ, Liu G, Kuivenhoven JA, Twisk J, Rip J, van Dop W, Excoffon KJ, Lewis SM, Kastelein JJ, Hayden MR (2005) Complete rescue of lipoprotein lipase-deficient mice by somatic gene transfer of the naturally occurring LPLS447X beneficial mutation. Arterioscler Thromb Vasc Biol 25(10):2143–2150. doi:10.1161/01.ATV.0000176971.27302.b0

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409(6822):928–933. doi:10.1038/35057149

Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Gunel M, Roeder K, Geschwind DH, Devlin B, State MW (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485(7397):237–241. doi:10.1038/nature10945

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305(5683):525–528. doi:10.1126/science.1098918

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K,

Lehtimaki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M (2007) Strong association of de novo copy number mutations with autism. Science 316(5823):445–449. doi:10.1126/science.1138659

Sobreira NL, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, Gumbs CE, Pevsner J, Thomas G, Valle D, Hoover-Fong JE, Goldstein DB (2010) Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet 6(6):e1000991. doi:10.1371/journal.pgen.1000991

Soemedi R, Wilson IJ, Bentham J, Darlay R, Topf A, Zelenika D, Cosgrove C, Setchfield K, Thornborough C, Granados-Riveron J, Blue GM, Breckpot J, Hellens S, Zwolinkski S, Glen E, Mamasoula C, Rahman TJ, Hall D, Rauch A, Devriendt K, Gewillig M, Os J, Winlaw DS, Bu'Lock F, Brook JD, Bhattacharya S, Lathrop M, Santibanez-Koref M, Cordell HJ, Goodship JA, Keavney BD (2012) Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. Am J Hum Genet 91(3):489–501. doi:10.1016/j.ajhg.2012.08.003

Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61:437–455. doi:10.1146/annurev-med-100708-204735

Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN (2009) The Human Gene Mutation Database: 2008 update. Genome Med 1(1):13. doi:10.1186/gm13

Stitziel NO, Kiezun A, Sunyaev S (2011) Computational and statistical approaches to analyzing variants identified by exome sequencing. Genome Biol 12(9):227. doi:10.1186/gb-2011-12-9-227

Sturtevant AH (1913) The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J Exp Zool 14:43–59

Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 3(5):391–397. doi:10.1038/nrg796

Teer JK, Mullikin JC (2010) Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet 19(R2):R145–R151. doi:10.1093/hmg/ddq333

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337(6090):64–69. doi:10.1126/science.1219240

Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, Konig IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Doring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA Jr, Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466(7307):707–713. doi:10.1038/nature09270

The Database of Genomic Variants (2013) http://projects.tcag.ca/variation/. Accessed January 21, 2013

Tietjen I, Hovingh GK, Singaraja RR, Radomski C, Barhdadi A, McEwen J, Chan E, Mattice M, Legendre A, Franchini PL, Dube MP, Kastelein JJ, Hayden MR (2012) Segregation of LIPG, CETP, and GALNT2 mutations in Caucasian families with extremely high HDL cholesterol. PLoS ONE 7(8):e37437. doi:10.1371/journal.pone.0037437

Veltman JA, Brunner HG (2012) De novo mutations in human genetic disease. Nat Rev Genet 13(8):565–575. doi:10.1038/nrg3241

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R,

Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. Science 291(5507):1304–1351. doi:10.1126/science.1058040

Vergeer M, Korporaal SJ, Franssen R, Meurs I, Out R, Hovingh GK, Hoekstra M, Sierts JA, Dallinga-Thie GM, Motazacker MM, Holleboom AG, Van Berkel TJ, Kastelein JJ, Van Eck M, Kuivenhoven JA (2011) Genetic variant of the scavenger receptor BI in humans. N Engl J Med 364(2):136–145. doi:10.1056/NEJMoa0907687

Volodarsky M, Markus B, Cohen I, Staretz-Chacham O, Flusser H, Landau D, Shelef I, Langer Y, Birk OS (2013) A deletion mutation in TMEM38B associated with autosomal recessive osteogenesis imperfecta. Hum Mutat. doi:10.1002/humu.22274

Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Roccanova P, Makarov V, Lakshmi B, Findling RL, Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gochman P, Long R, Chen Z, Davis S, Baker C, Eichler EE, Meltzer PS, Nelson SF, Singleton AB, Lee MK, Rapoport JL, King MC, Sebat J (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320(5875):539–543. doi:10.1126/science.1155174

White R, Woodward S, Leppert M, O'Connell P, Hoff M, Herbst J, Lalouel JM, Dean M, Vande Woude G (1985) A closely linked genetic marker for cystic fibrosis. Nature 318(6044):382–384

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 285(5429):901–906

Xiong M, Guo SW (1998) The power of linkage detection by the transmission/disequilibrium tests. Hum Hered 48(6):295–312

Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet 40(7):880–885. doi:10.1038/ng.162

Yla-Herttuala S (2012) Endgame: glybera finally recommended for approval as the first gene therapy drug in the European union. Mol Ther 20(10):1831–1832. doi:10.1038/mt.2012.194

Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10:451–481. doi:10.1146/annurev.genom.9.081307.164217

Zhou J, Tawk M, Tiziano FD, Veillet J, Bayes M, Nolent F, Garcia V, Servidei S, Bertini E, Castro-Giner F, Renda Y, Carpentier S, Andrieu-Abadie N, Gut I, Levade T, Topaloglu H, Melki J (2012) Spinal muscular atrophy associated with progressive myoclonic epilepsy is caused by mutations in ASAH1. Am J Hum Genet 91(1):5–14. doi:10.1016/j.ajhg.2012.05.001

Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci USA 109(4):1193–1198. doi:10.1073/pnas.1119675109