Towards a complete resolution of the genetic architecture of disease

Andrew B Singleton[1], **John Hardy**[2], **Bryan J Traynor**[1,3], and **Henry Houlden**[2,4]

[1] Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA [2] Department of Molecular Neuroscience and Reta Lilla Weston Laboratories, Institute of Neurology, London, England [3] Neuromuscular Diseases Research Group, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA [4] MRC Centre for Neuromuscular Diseases at the Institute of Neurology, London, England

Abstract

Following years of linear gains in the genetic dissection of human disease, we are now in a period of exponential discovery. This is particularly apparent for complex disease. Genome-wide association studies have provided myriad associations between common variability and disease, and shown that common genetic variability is unlikely to explain the entire genetic predisposition to disease. Here, we detail how one can expand on this success and systematically identify genetic risks that lead or predispose to disease using next generation sequencing. Geneticists have had for many years a protocol to identify Mendelian disease. Now we have available a similar set of tools for the identification of rare moderate risk loci and common low risk variants. While undoubtedly major challenges remain, particularly with data handling and the functional classification of variants, we suggest that these will be largely practical and not conceptual.

An enviable position

As geneticists we sit in an enviable position, one with a history of considerable success behind us, and a future that holds promise for significant gains. The first successful attempts at identifying genetic mutations in the late 1980's transformed the human genetics field[1–6]. Following this early success, linkage and positional cloning became a mainstay of human genetics; widespread adoption of this approach manifested in a growing body of established genetic links to disease. While this work has been critical to understand the etiology of diseases, improvements to this genetic method have, until recently, been largely evolutionary rather than revolutionary. The advent of highly parallel genotyping and the development of next generation sequencing is now changing the landscape and provides a solution to this gap in genetic technology.

Technological advances in terms of array and DNA sequencing technologies mean that the route towards the complete examination of genetic risk is now largely clear. For any human disorder, we can now plot a rational and systematic approach towards the identification of the underlying genetic factors. This bold statement is not intended to minimize the challenges

ahead, as there will be many, particularly in the identification of the truly pathogenic changes, but rather to lay an outline of this route and to identify the obstacles ahead.

Manolio and colleagues[7] conveniently graphed out the types of genetic risk (modified in Figure 1) and noted three broad categories of risk likely to be tractable by genetic approaches in the near future: high risk–rare alleles which lead to Mendelian (such as *APP* or *PS* mutations in Alzheimer's disease)[8–10] or near Mendelian disease (such as *LRRK2* mutations in Parkinson's disease)[11, 12], moderate risk–low frequency alleles for disease (for example *GBA* mutations in Parkinson's disease)[13, 14], and low risk–common alleles which have modest effects on predisposition to disease (such as those in *SNCA* or *MAPT* in Parkinson's disease and *CLU*, *PICALM* and *CR1* in Alzheimer's disease)[15–18]. Two other categories of allele can be added to this list: those that are neither necessary nor sufficient to cause disease but modify the phenotype (such as modifier genes in cystic fibrosis)[19] and those that are needed to be present with the disease allele for the disease to occur (such as in the Bardet-Biedl syndrome)[20].

While these categories represent points on a spectrum, they help to outline the genotyping and sequencing methods required to solve these genetic problems. Here, we discuss how these categories of pathogenic loci can be systematically dissected and what are the purposes and expectations of this dissection.

High risk–rare alleles (Mendelian disease)

Nearly all high-risk loci display changes that have clear effects on protein coding genes, such as missense, nonsense and splice site changes or gene deletions/insertions (www.ncbi.nlm.nih.gov/Omim/)[21]. Many loci for simple Mendelian disease have been identified by positional cloning during the last 20 years[1, 2, 4]. However this process, while effective, has traditionally been arduous even after the completion of the human genome project lead to a near complete annotation of all genes[22, 23].

Despite the proven power of positional cloning to identify causative loci, there are many reasons why linkage might fail. The two main reasons are: (1) a lack of genetically informative families, particularly in late onset diseases or conditions where high infant mortality or a rapidly progressing condition hamper the collection of a sufficient number of affected individuals within a family, and (2) locus heterogeneity, where pathogenic alleles exist in multiple genes. Furthermore, this approach is expensive, and in the cases of microsatellite linkage and positional candidate sequencing, extremely slow. The development of viable methods for whole exome sequencing[24] provides a powerful alternative to positional cloning, with some notable advantages. First, because potentially causal variants are identified, this method can be applied in families that are too small to provide meaningful information using linkage, effectively allowing small families and even single probands (an individual affected with a disorder who is the first subject from its family to come under study) to be analyzed jointly, irrespective of allelic heterogeneity[25]. Of course, it should also be recognized that the identification of disease alleles by exome sequencing requires families showing statistical evidence for co-segregation or sufficient locus homogeneity and disease frequency for statistical comparison between cases and controls to be possible. Second, this method can be incredibly fast, moving from well-defined trait to mutation within weeks rather than years. Exome sequencing as a method to find mutations has already shown considerable promise, particularly in very rare autosomal recessive diseases. One might expect that until the widespread adoption of whole genome sequencing, this will be the pro forma approach to identify the cause of monogenic diseases and that, coupled with array methods that can rapidly detect gene deletions[26], insertions[27, 28], and autozygosity for recessive loci[29], gene identification for monogenic disease will become quite routine.

Low risk–common alleles

Genome-wide association studies (GWAS) have proven to be an outstanding success in terms of identifying common low risk variants for complex disease[7, 30]. Hundreds of replicated genetic associations have been reported and these are being continually added in the NHGRI catalog of published GWAS (http://www.genome.gov/26525384)[31]. The method, capabilities and limits of GWAS have been well-documented elsewhere [30, 32–34]; however, we have briefly covered these in a text box in order to put the following sections into context (Box 1).

Box 1

Genome-wide association study

GWAS primary aim is to reveal the genomic location of common genetic variants that impart risk for a trait; this method aims to test the common disease, common variant hypothesis[46]. In this scenario several hundred thousand SNPs throughout the genome are typed in a large series of disease cases and disease-free controls. Allele and genotype frequencies at each of these SNPs are then compared between the case group and the control group to detect alleles or genotypes that are over-represented in one group versus the other. A statistically significant association implies that there is a risk variant close to the associated SNP (or plausibly that the associated SNP is the risk variant). Because common variants are tested in GWAS, it is not designed for, and is inefficient at, finding rare genetic variants that contribute to disease. The explicit testing of the common disease common variant hypothesis is an important issue. Criticisms often leveled at GWAS include the notion that this is hypothesis-free research or a fishing expedition, but these experiments often generate important hypotheses.

The power of GWAS over previous candidate gene association studies is the ability of this method to assay common genetic variability throughout the genome. This method is accurate and broad enough that we can use haplotype maps of human populations[47] to predict the genotype of a large number of untyped variants in an individual. This means that genotyping of a few hundred thousand single nucleotide polymorphisms (SNPs) in a subject will allow us to predict (impute) with high confidence the genotype at more than a million additional variants. While many common variants remain untyped and not imputed, linkage disequilibrium between physically close variants often means that association signals at variants surrounding the true risk variant are apparent. This raises two other considerations: First, not all of the genome is covered in a GWAS, most modern arrays capture approximately 90% of the common genetic variability in one way or another. Second, GWAS identify loci not genes. A positive signal from a GWAS is therefore not always within the known functional unit of a gene. Most often the closest gene to an association signal, or a nearby gene that makes biological sense, is nominated as the functional unit of the association. It is important, however, that this nomination remains just that, and does not become dogma fixed in the absence of evidence to support the biologic association.

In relatively few instances do the predisposing loci identified by GWAS have coding changes at their heart[31]. Rather it seems that these loci usually exert their effect either directly or indirectly through modest effects on gene expression. For the sake of clarity in this article, when we discuss the potential effect of genetic variability on gene expression we consider expression to include variability in constitutive, induced and spatial expression, in addition to splicing. Indeed there is a rough relationship between the size effect of a risk locus and the likelihood that a coding change will underpin it, with very few loci that result in moderate increases in risk (less than a doubling in risk for a trait) being explained through protein coding changes. The fact that genetic variability in expression is likely to be the key factor in the

elucidation of the mechanisms of low risk variability raise the importance of understanding the effects of genetic variability on gene expression.

What has GWAS missed in complex disease?

While each complex disease has a unique genetic architecture, some authors have expressed disappointment that GWAS findings often do not explain a large proportion of the heritability of complex disease. As has been previously discussed there are several possibilities for this discrepancy[7]: 1) Heritability calculations are often quite indirect and often involve simplified models of genetic versus non-genetic contributors to disease[35]. 2) Additionally, allelic heterogeneity is a potential confound of this method. Indeed, it as been *a priori* argued that the presence of multiple risk haplotypes at any locus would prevent the identification of single risk alleles through the GWAS approach[36]. *Post hoc,* one would now argue that, at any observed locus, the alleles identified though GWAS might merely reflect a marginal fraction of risk and that one should think of graded rather than dichotomous risk at each locus. 3) Most GWAS are underpowered to detect minor disease associations. There is likely to be a large number of common variants that impart very minor risk for disease outside the limits of what can be reliably detected in GWAS. Collectively these might contribute a substantial amount of risk for a complex disease, but would require enormously large cohorts in the range of hundred thousand to have adequate power for detection. Such large sample sizes might be impractical for many diseases, although there is evidence that this could be feasible for both schizophrenia and bipolar disorder[37]. As we begin to identify minor risk alleles and recognize the presence of multiple risk alleles at the same locus we will undoubtedly revise upwards our estimates of the proportion of heritability we have explained. However, this does not preclude that a substantial proportion of genetic risk in complex disease might result from rare alleles.

Moderate risk–low frequency alleles

Moderate risk and low frequency alleles were previously difficult to identify, both because they do not have a big enough effect on disease risk to display clear Mendelian pattern of inheritance in individual families (and thus are not amenable to linkage mapping), and because individual variants are too rare to be efficiently identified by association strategies (although multiple variants at the same locus might contribute to disease risk). These loci have been described as the "dark matter of disease risk"[7] and as such deserve a more extensive discussion because only now are we beginning to identify them in a systematic fashion using exome or genome sequencing, identification of insertions and deletions on arrays and by imputation based on 1000 genomes (and similar) data (http://www.1000genomes.org/).

What is the likely composition of the unidentified genetic risk for disease?

We hypothesize that a substantial proportion of this undiscovered fraction will be heterozygous loss-of-function alleles. Our argument for this is as follows: We now know that many common low risk variants have modest effects (10%–20%) on gene expression (see below). Loss-of-function variants effectively lead to a 50% loss of expression and would therefore generally be predicted to have larger effects on disease risk. We also know that there are many ways of disrupting gene function: whole or partial gene deletions, missense and splice site changes. Each individual variant will be rare, though the combined burden of all the rare variants at each locus could be substantial. Loss-of-function changes will necessarily have low population frequencies, because homozygotes would suffer from severe, early onset disease consequences, and so the most likely scenario for such loci is that rare and diverse loss of function alleles will occur in any population. Perhaps the best example of a moderate risk, low frequency locus is the occurrence of loss-of-function alleles of glucosecerebrosidase in Parkinson's disease[14]. This was identified through the clinical observation that first and second-degree relatives of Gaucher's disease have increased incidence of Parkinson's disease.

While in fact, the overall strategy for the systematic identification of moderate risk, low allele frequency loci does not depend on the nature of the variant which is sought, its ease of implementation does. Thus while we believe that exome and genome sequencing will be an effective approach to identify a wide variety of low frequency risk alleles, we predict that disease loci where the effect is mediated by loss-of-function alleles will be considerably easier to categorize as disease associated than those mediated by gain-of-function alleles.

Towards the systematic identification of low frequency–moderate risk loci

The primary route for the identification of these loci is likely to be through exome sequencing of large numbers of disease samples. This approach is likely to be augmented through the identification of gene deletions and insertions using arrays (which are often run in conjunction with exome sequencing). Furthermore, as imputation of data from sequencing projects improves, this will facilitate the identification of deleterious variants through haplotype proxies (haplotypes that can predict the presence of individual rare variants) in GWAS (see below). If our hypothesis that many moderate risk loci are likely to be loss-of-function variants is correct, it will make the identification of the moderate risk loci easier because loss-of-function changes are easier to categorize both as a functional group within any one gene and from comparison to mutations and genes known to be associated with recessive forms of disease.

Furthermore, this hypothesis suggests that prospective follow-up of the parents of children with recessive diseases would provide an insight into how their morbidity and mortality differs from the general population.

Some exceptions

The three categories of genetic predisposition detailed above appear to cover most categories of genetic risk. However, some pathogenic loci do not fit within these three categories

High risk–common variants

A few common high-risk variants have been identified: apolipoprotein E and Alzheimer's disease[38] and complement factor H in macular degeneration[39–41]. In both cases, a common, high-risk coding variant has a large effect on disease risk. It is notable that both these diseases are *par excellence* late onset diseases. Perhaps it is to be expected that common high-risk variants which have large effects on disease risk should only be found in diseases which strike after the childbearing age because otherwise there would be a strong selective pressure against them. These loci were easily identifiable by GWAS. However, it is notable that at both loci there is additional risk not accounted for by the major allelic variant[42]. This fact points to the importance of sequencing all loci identified by all approaches to look for graded risk at each locus.

De novo mutations

De novo mutations have always posed a problem for the systematic analysis of disease because they are not inherited from parents and are difficult to find except in genes already associated with disease. Array technologies now offer a rapid and comprehensive way to look for insertions and deletions across the genome and these have proved fruitful especially in those conditions, such as autism, in which there is strong selection against reproductive fitness[43]. Whole exome sequencing will augment this systematic identification of loci in conditions in which there is no family history.

Low risk–low frequency variants

We predict that higher risk variants are likely to alter the coding sequence of genes whereas lower risk variants are likely to exert their effects through mild or moderate effects on gene

expression or splicing. One might therefore expect that parts of the unidentified genetic risk for disease are rare, non-coding alleles that alter risk for disease by a small amount. Of all the genetic risk categories described here, this is probably the one for which it will be most difficult to discern risk, protective, and null alleles. Likely most of these alleles will resist identification by GWAS and by exome sequencing and full genome sequencing will be required. We believe that whole genome sequencing should clearly be the aim because when the method becomes both cheap and efficient it will be, in our opinion, the best way to identify all major risk allele categories, replacing GWAS, positional cloning and exome sequencing. Individually low risk rare variant alleles will be too infrequent and impart too little risk to associate with disease. Thus the main difficulty in using genome sequencing as a method to identify risk alleles might lie in the categorization of alleles at a single locus into risk, protective and benign groups, a necessary intermediate step toward identifying biologic association. It is difficult to estimate how difficult a problem this might be without being able to predict the immediate functional consequences of DNA sequence changes (for e.g. on DNA structure, or dynamics of binding to regulatory proteins). One might suggest that such a barrier will only be overcome by a combination of large numbers of cases and controls, computational prediction of effects of non-coding alleles, and functional assays of sufficient sensitivity to allow determination of variant effects under many different conditions.

Goals of the genetic analysis

The genetic analysis detailed above has two essential goals: first, to get better at genetic prediction of who is at risk for disease and second, to detail the biochemical pathways to pathogenesis. While both are worthy aims, the first has perhaps been overstated. Identification of pathways to disease is therefore a major goal of genetic analysis of disease because through the identification of these pathways we hope to identify targets to intervene in these pathways. This approach is already paying dividends. For example, autophagy has been identified in the analysis of Crohn's disease[44] and the innate immune system in the analysis of Alzheimer's disease[16, 17]. Not only does the analysis of risk yield potential insights into the biochemical pathways underlying a disease, they also lead to predictions as to what other genes might harbor genetic risk for disease. Over time, as the jigsaw puzzle of disease risk gets pieced together, the positioning of each subsequent piece should get easier.

Concluding remarks

Here, we have detailed how disease risk can now be systematically analyzed for all categories of genetic risk. This would seem to imply that all what is now required is that the right samples are collected and that funds are available to ensure the correct analysis. Clearly many practical challenges remain for the promise of systematic analysis to be realized. Perhaps a greater challenge, however, will be to integrate systematic genetic information with largely unsystematic information on the environment. While we can now see how full genetic information can be gathered, it is much more difficult to see how unstructured environmental information can be understood. Hopefully, however, as we identify pathways to disease, this will give us clues as to the crucial environmental exposures we need to consider for each disease[45].

Glossary

Common disease, common variant hypothesis	the idea that genetic risk for common diseases is driven in part by an individual's load of common risk variants. Individually such variants alter risk by a minor amount, but collectively they might increase risk substantially

Exome sequencing	also known as targeted exome capture, this method is an efficient strategy to selectively sequence the coding regions and flanking intronic sections of the human genome to identify novel genes associated with rare and common disorders. An exome accounts for about 180 000 exons and for around 1% of the human genome which translates to about 30 megabases (Mb) in length. It is estimated that the protein-coding regions of the human genome constitute about 85% of the disease-causing mutations
Gain-of-function alleles	classically associated with dominant forms of monogenic disease, gain-of-function alleles are those variants that lead to a new function for the gene product
Genome-wide association studies	also known as whole genome association studies (WGA), it is the examination of genetic variation across a given genome using many hundreds of thousands of single nucleotide polymorphisms (SNPs) on DNA arrays. They are designed to identify genetic associations with observable traits and require large numbers of cases to identify the associated regions
Heritability	the proportion of phenotypic variation in a population that is attributable to genetic variation among individuals as opposed to environmental factors
Imputation	the prediction of missing genotype data based on genotype at physically proximal tested SNPs
Linkage disequilibrium	the non-random association of alleles at two or more loci, not necessarily on the same chromosome. It is not the same as genetic linkage, which describes the tendency of certain loci or alleles to be inherited together. Genetic loci on the same chromosome are physically close to one another and tend to stay together during meiosis, and are thus genetically linked
Loss-of-function alleles	sometimes also called null alleles and most classically associated with recessive forms of disease, loss-of-function alleles are those alleles that result in the gene product having less and sometimes no function. Perhaps the most easily recognized loss-of-function alleles are those that clearly disrupt the production of a protein product, for example protein truncating mutations caused by frameshift or premature stop mutations

References

1. Riordan JR, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989;245:1066–1073. [PubMed: 2475911]

2. Rommens JM, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 1989;245:1059–1065. [PubMed: 2772657]

3. Wallace DC, et al. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. Science 1988;242:1427–1430. [PubMed: 3201231]

4. Kerem B, et al. Identification of the cystic fibrosis gene: genetic analysis. Science 1989;245:1073–1080. [PubMed: 2570460]

5. Hoffman EP, et al. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell 1987;51:919–928. [PubMed: 3319190]

6. Koenig M, et al. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 1987;50:509–517. [PubMed: 3607877]

7. Manolio TA, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–753. [PubMed: 19812666]

8. Goate A, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 1991;349:704–706. [PubMed: 1671712]

9. Rogaev EI, et al. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. Nature 1995;376:775–778. [PubMed: 7651536]

10. Sherrington R, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 1995;375:754–760. [PubMed: 7596406]

11. Paisan-Ruiz C, et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron 2004;44:595–600. [PubMed: 15541308]

12. Zimprich A, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 2004;44:601–607. [PubMed: 15541309]

13. Aharon-Peretz J, et al. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med 2004;351:1972–1977. [PubMed: 15525722]

14. Sidransky E, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med 2009;361:1651–1661. [PubMed: 19846850]

15. Simon-Sanchez J, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet 2009;41:1308–1312. [PubMed: 19915575]

16. Harold D, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 2009;41:1088–1093. [PubMed: 19734902]

17. Lambert JC, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 2009;41:1094–1099. [PubMed: 19734903]

18. Satake W, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet 2009;41:1303–1307. [PubMed: 19915576]

19. Gu Y, et al. Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease. Nature 2009;458:1039–1042. [PubMed: 19242412]

20. Katsanis N, et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science 2001;293:2256–2259. [PubMed: 11567139]

21. den Dunnen JT, Antonarakis SE. Mutation Nomenclature. Current Protocols in Human Genetics 2003;UNIT 7.13

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

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

24. Ng SB, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009;461:272–276. [PubMed: 19684571]

25. Ng SB, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42:30–35. [PubMed: 19915526]

26. van de Leemput J, et al. Deletion at ITPR1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans. PLoS Genet 2007;3:e108. [PubMed: 17590087]

27. Gibbs JR, Singleton A. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. PLoS Genet 2006;2:e150. [PubMed: 17029559]

28. Knight MA, et al. A duplication at chromosome 11q12.2-11q12.3 is associated with spinocerebellar ataxia type 20. Hum Mol Genet 2008;17:3847–3853. [PubMed: 18801880]

29. Camargos S, et al. DYT16, a novel young-onset dystonia-parkinsonism disorder: identification of a segregating mutation in the stress-response protein PRKRA. Lancet Neurol 2008;7:207–215. [PubMed: 18243799]

30. Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med 2009;360:1759–1768. [PubMed: 19369657]

31. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009;106:9362–9367. [PubMed: 19474294]

32. Wang WY, et al. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 2005;6:109–118. [PubMed: 15716907]

33. McCarthy MI, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008;9:356–369. [PubMed: 18398418]

34. Hunter DJ, Kraft P. Drinking from the fire hose--statistical issues in genomewide association studies. N Engl J Med 2007;357:436–439. [PubMed: 17634446]

35. Rose SP. Commentary: heritability estimates--long past their sell-by date. Int J Epidemiol 2006;35:525–527. [PubMed: 16645027]

36. Terwilliger JD, Hiekkalinna T. An utter refutation of the 'fundamental theorem of the HapMap'. Eur J Hum Genet 2006;14:426–437. [PubMed: 16479260]

37. International Schizophrenia C, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748–752. [PubMed: 19571811]

38. Corder EH, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–923. [PubMed: 8346443]

39. Klein RJ, et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385–389. [PubMed: 15761122]

40. Haines JL, et al. Complement factor H variant increases the risk of age-related macular degeneration. Science 2005;308:419–421. [PubMed: 15761120]

41. Edwards AO, et al. Complement factor H polymorphism and age-related macular degeneration. Science 2005;308:421–424. [PubMed: 15761121]

42. Li M, et al. CFH haplotypes without the Y402H coding variant show strong association with susceptibility to age-related macular degeneration. Nat Genet 2006;38:1049–1054. [PubMed: 16936733]

43. Weiss LA, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med 2008;358:667–675. [PubMed: 18184952]

44. Barrett JC, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008;40:955–962. [PubMed: 18587394]

45. Traynor BJ. The era of genomic epidemiology. Neuroepidemiology 2009;33:276–279. [PubMed: 19696518]

46. Lander ES. The new genomics: global views of biology. Science 1996;274:536–539. [PubMed: 8928008]

47. de Silva R, et al. Strong association of the Saitohin gene Q7 variant with progressive supranuclear palsy. Neurology 2003;61:407–409. [PubMed: 12913211]
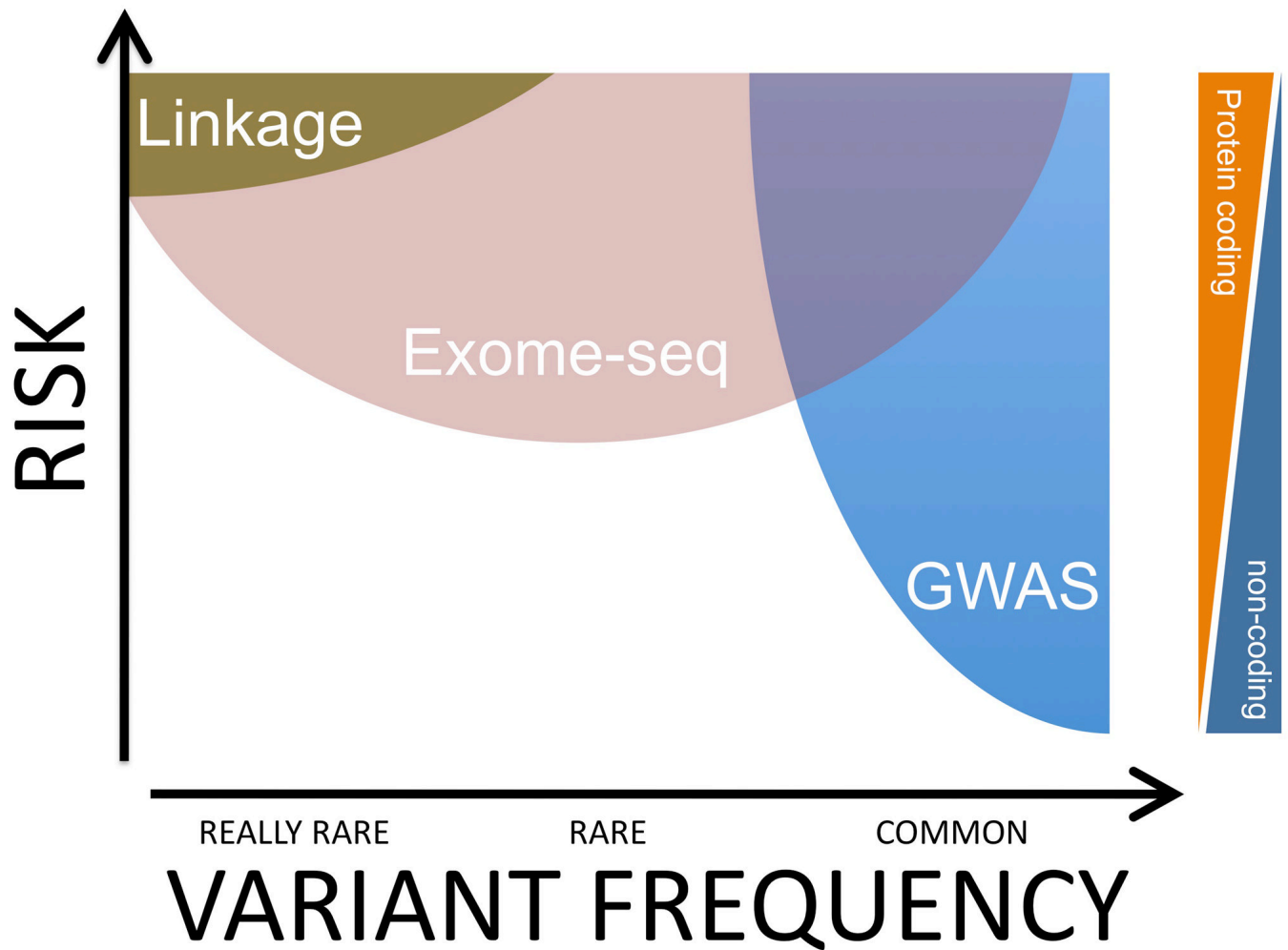
Figure 1.
Schematic of the genetic architecture landscape in disease and the applicability of current methods to finding risk and causative alleles. Modified from the concept presented by Manolio and colleagues[7].