

PERSPECTIVES

GENOME-WIDE ASSOCIATION STUDIES — VIEWPOINT

Mendelian disorders and multifactorial traits: the big divide or one for all?

Stylianos E. Antonarakis, Aravinda Chakravarti, Jonathan C. Cohen and John Hardy

Abstract | For the past century, Mendelian and multifactorial traits have existed at opposite ends of the disease spectrum in humans. Furthermore, the recent emphasis on genome-wide association studies for uncovering variants that underlie common diseases has risked deepening the divide — or has it? Four experienced human geneticists express their views on the changing landscape of human disease studies and the impact of new technologies and study designs on the age-old aim of connecting a genomic variant with its phenotypic consequences.

Q *How has recent interest in human multifactorial traits affected research into Mendelian disorders?*

Stylianos E. Antonarakis. For about a century, the argument for focusing on rare Mendelian disorders was as follows: Mendelian phenotypes will reveal genes involved in common multifactorial traits, metabolic pathways involved in these disorders and networks of functionally related proteins that may be involved in the development of multifactorial traits. Many examples supported these predictions, and a whole generation of human geneticists had a happy rollercoaster ride in the discovery of pathogenic variation for Mendelian traits. It is true, however, that such Mendelian disorders provided little understanding of the genetic contribution to multifactorial traits. Over the past five years, advances in genomics-related technologies have lured many scientists back to tackling multifactorial traits, this time with more powerful tools. During this period, Mendelian disorders have been somewhat neglected, partly because they are rare and partly because of the perception that, in terms of public health impact, discovering the causes of multifactorial traits would be more rewarding. The

overall lesson from genome-wide association (GWA) studies has been that common genomic variation explains no more than 5% of the phenotypic variation. This has logically led to the current interest in the impact of rare variants on common multifactorial traits. But studies of rare variation take us back to the need to investigate rare traits — this is the only means we have to molecularly understand the one issue that I think is fundamental in all genetic traits, namely the issue of penetrance. The understanding of penetrance — the fraction of phenotypic expression given the genotypic variant — will reveal why a variant (common or rare) results in a phenotypic trait. I argue that the best way to understand penetrance is by focusing again on Mendelian traits because of the strong link that exists between the phenotype and the pathogenic variant in these diseases.

Aravinda Chakravarti. Accurate genetic and physical maps of the human genome, the early products of the genome projects, have revolutionized our understanding of Mendelian diseases by elucidating the specific genes and mutations involved. Today, we know the molecular basis of over 4,000 Mendelian disorders and are using the information to contemplate treatment strategies

for some; however, over 2,000 entities remain unsolved. There are likely to be many more monogenic disorders that fail to 'Mendelize' and so have been beyond the reach of simple analysis; these include disorders caused by maternal or paternal effect genes (a bearer's phenotype being determined by the mother's or father's genotype, respectively), and those whose expression is contingent on maternal-fetal or environmental interactions or on other, as yet unknown genetic mechanisms. There is no doubt that the long-neglected common multifactorial traits and disorders are now of great interest because new high-resolution genomic tools have made it possible to genetically dissect them — as a consequence, attention on rare single-gene disorders has diminished. Nevertheless, the molecular mechanisms of complex diseases are still elusive, and progress in understanding them has been halting. As we turn from mapping loci to understanding mechanisms, we will return to the analysis of monogenic disorders, for which the answers may be more accessible.

Jonathan C. Cohen. The recent focus on multifactorial traits has diverted attention from Mendelian disorders, but the primary obstacle to discovering additional genes that cause Mendelian disorders has been technical. Linkage analysis successfully resolved most of the more common single-gene disorders, but until the development of whole-genome sequencing, no good alternative approaches were available for conditions that were not tractable to linkage-based methods. The advent of whole-genome sequencing is transforming our ability to identify gene defects that cause human diseases.

John Hardy. Yes it has, in two ways: one good and one bad. The good effect is that the technologies developed for the assessment of complex disorders are also incredibly powerful for elucidating Mendelian disorders: whole-genome SNP arrays are extraordinarily useful for homozygosity mapping and for finding insertions and deletions that can lead to Mendelian disease¹. If you are lucky, you can now 'instantly clone' a gene for a disease: a student in our laboratory cloned a gene for a dominant ataxia in a week, using arrays².

The contributors*

Stylanos E. Antonarakis is Professor of Genetic Medicine and Chairman of the Department of Genetic Medicine and Development of the Medical Faculty of the University of Geneva, Switzerland. He received his medical training from the University of Athens, Greece, and then trained as a geneticist at Johns Hopkins University in Baltimore, USA; he joined the faculty of John Hopkins in 1982 and rose quickly to become Professor of Genetics. In 1993, he moved to the University of Geneva to take up the chair of genetics. His research interests span Mendelian disorders (including thalassaemias and haemophilia), aneuploidy syndromes (including trisomy 21) and multifactorial disorders (including schizophrenia). He has published extensively on genetic disorders and genome analysis, and he is a highly cited physician–scientist who has also served as President of the European Society of Human Genetics.

Aravinda Chakravarti is Director of the Center for Complex Disease Genomics and Professor of Medicine, Pediatrics, Molecular Biology & Genetics at the Johns Hopkins University School of Medicine, Baltimore, USA, and Professor of Biostatistics at the Bloomberg School of Public Health, Baltimore, USA. He is a member of the US National Academy's Institute of Medicine and an honorary fellow of the Indian Academy of Sciences, and he served as president of the American Society of Human Genetics in 2008. He has been a key participant and architect of the Human Genome, HapMap and 1000 Genomes projects, and he is one of the founding editors of *Genome Research* and *Annual Reviews of Genomics & Human Genetics*. His research is aimed at genome-scale analysis of gene variation and function to understand the molecular genetic basis of human disease.

Jonathan C. Cohen is a professor of internal medicine at the University of Texas Southwestern Medical Center, Dallas, USA. His laboratory focuses on the genetics of metabolic disorders with a particular emphasis on lipids and lipoproteins.

John Hardy is Professor of Neuroscience at the University College London Institute of Neurology, UK, where he has worked since 2007. Previously he was Chief of the Laboratory of Neurogenetics at the National Institute on Aging at the National Institutes of Health in Bethesda, Maryland, USA, and before that he was Chair of Neuroscience at the Mayo Clinic in Jacksonville, Florida, USA. Although he trained as a neuropharmacologist, he has become a geneticist and has worked on the genetics of neurodegenerative diseases for over 25 years.

*Listed in alphabetical order.

And of course, whole-exome sequencing is exceptionally valuable for finding genes for rare Mendelian diseases for which there is not adequate family material³, even though it will be used more extensively for complex diseases. The bad effect is that there is a creeping belief that sorting out Mendelian diseases is boring and that, for some reason, it is not worth funding. I find this a shocking and irresponsible view: we need to find these genes for the patients and their families and for the mechanistic insights they provide. To dismiss applications that seek to find these genes on the grounds that the science is not innovative is not acceptable. In my view, too much effort is being placed on trying to get people to think outside the box, when the answer is sometimes clearly inside the box.

Q How do studies of multifactorial and Mendelian traits contribute to a mechanistic understanding of disease?

S.E.A. Mendelian traits reveal highly penetrant variants, whereas the study of multifactorial traits will probably uncover an interacting group of low- to medium-penetrance variants. An understanding of the underlying molecular mechanisms that

mediate the entire spectrum of penetrance (in each cell type) is, in my view, the key to the mechanistic understanding of disease in each individual. The study of genetic and environmental influences on the phenotype, which is being made possible by using current genome analysis tools, holds promise not only for furthering our scientific understanding but also for improving diagnostic prediction and for introducing individualized therapeutic interventions.

A.C. There are two definitions of 'mechanism' in genetics, one concerning inheritance and the other concerning the encoded biology that is disrupted in disease. Mendelian traits are, by definition, understood at the inheritance level and many are also understood at the biological level, as most mutations probably disrupt a single, critical, rate-limiting step (for example, an enzyme). However, the molecular mechanisms of many Mendelian diseases still require much work: it is one matter to show *fibrillin 1* deficiency in *Marfan syndrome* and quite another to show that the phenotypic effects (and the therapeutic target) arise from downstream transforming growth factor- β activation. By contrast, 'complex' traits are simply not understood

at either level. To further this understanding, we first of all need to solve the mystery of how a trait or disease is 'synthesized' from the underlying genetic variation: which genes and why them? Which mutations and why them? What is the biological principle that necessitates multiple mutations? Therefore, intensive mapping of the loci associated with multifactorial traits and diseases — the majority of the treasury of human inheritance — is a continuing priority. Second, our understanding of the suite of genetic programs that are compromised in these diseases is in its very early stages but is exceedingly important, as these programs will provide unique views of how 'biology' is encoded in our genomes. My prediction is that our understanding of mechanisms in multifactorial disease will not be put on hold until the complete mapping of a trait — instead, the knowledge accrued will contribute to a gradual unfurling of both inheritance and biological mechanisms by continued reiteration of a mapping–mechanism–mapping cycle.

J.C.C. Genetic studies of multifactorial and Mendelian traits provide molecular points of entry to disease mechanisms by revealing the genes that contribute to a disorder. The two categories of studies usually provide different types of information. The identification of mutations responsible for Mendelian disorders reveals crucial components of the disease process. Conversely, most variants associated with multifactorial disorders have modest phenotypic effects and therefore provide little indication of the importance of the gene in the disease. Alleles with large phenotypic effects that are sufficiently common in the general population are particularly informative for complex disorders. For example, low-frequency alleles of proprotein convertase subtilisin/kexin type 9 (*PCSK9*) provided direct evidence that lifelong reduction of cholesterol is sufficient to protect against coronary heart disease. We have referred to these alleles as 'Goldilocks' alleles: they are well suited for gaining insights into genotype–phenotype relationships because they have a large effect size, but they are also common enough to be useful in population-based studies. Goldilocks alleles account for only a minor fraction of population attributable risk but nonetheless provide direct insights into the mechanistic underpinnings of complex diseases.

J.H. We have to expect, or at least hope, that the loci identified for any particular disease will map onto related pathways. We don't yet know enough to be sure that this will

generally be the case. But for the neurodegenerative diseases I know about, the early signs are hopeful. In many of the neurodegenerative diseases characterized by protein deposition, Mendelian disease analysis often finds the gene encoding the deposited protein, and normal genetic variability at the same locus contributes to the risk of the sporadic disease⁴. Perhaps the best example of this is frontotemporal dementia: here, mutations in the microtubule-associated protein tau (*MAPT*) gene cause the autosomal-dominant disease⁵, and normal genetic variation in *MAPT*, leading to overexpression of the tau mRNA⁶, is one of the predisposing factors for the sporadic disease (called *progressive supranuclear palsy*). More generally, it is encouraging to see that the loci implicated in Alzheimer's disease^{6,7} are mapping onto pathways that had been previously implicated in the disease^{8,9}.

Q Are there points of convergence between the two approaches?

S.E.A. There is a natural convergence between the two approaches in that both ends of the disease spectrum (and the disorders in between — that is, oligogenic traits) have their molecular basis in genomic (as well as epigenetic) variation. Therefore the focus on genome variability — common, near rare and rare — is the 'common currency' for understanding all genetic disorders. All genetic disorders and traits teach us something about the penetrance of pathogenic variants given the total (and unique) variation that exists in each individual genome. This statement also emphasizes the need to study human populations to understand human genetic disorders.

Model organisms could help tremendously in deciphering the basic biological mechanisms and the connection between functional 'units'; however, because they lack the genomic variability that currently exists in humans, they cannot fully explain human multifactorial traits.

A.C. It is currently popular to think of human (and other) inheritance as either Mendelian or multifactorial, although these are the extremes and many models of inheritance (involving varying numbers of genes) lie in between. A more apt classification of inheritance would be Mendelian versus non-Mendelian. However, there are many reasons for non-Mendelian inheritance, and they can involve as few as one gene to many genes. Unfortunately, the segregational complexity is not a good indicator of gene number, as the inheritance of 2 or 20 genes can be similarly complex. Nevertheless, the genetics literature routinely emphasizes the Mendelian-multifactorial dichotomy, and this is reflected in the methods chosen for gene mapping: linkage analysis for Marfan syndrome and GWA studies for type 2 diabetes. In our own studies of *Hirschsprung disease*, which is decidedly non-Mendelian, we have found it useful to study the entire gamut of samples (outbred cases and controls, trios, nuclear families and large kindreds in Mennonites) using a broad range of genetic technologies (karyotyping, linkage, association, sequencing, cellular assays and functional studies in model systems) to identify new genes and find their roles in the disease mechanism. This is equally true for hypertension, for which no one sample type or technology has led to overwhelming success. This 'holistic' approach can provide deeper insights into

inheritance and disease mechanisms than any single tactic, suggesting that 'convergence' is essential to good genetic analysis of any trait or disease.

J.C.C. The two approaches offer complementary pathways towards the same overall goal (*viz.* identifying genes for diseases and traits), and there are several points at which they converge. Genetic studies of blood lipid levels provide an example: many genes identified by GWA studies also underlie Mendelian disorders of lipid metabolism. These studies also provide examples of potential complementarity: HMG-CoA reductase (*HMGCR*), a crucial gene in the cholesterol biosynthesis pathway, was associated with blood levels of low-density lipoprotein cholesterol in population studies but does not seem to have a Mendelian form, presumably because severe mutations in this gene would be lethal. Similarly, not all genes associated with Mendelian forms of dyslipidaemia have been consistently associated with blood lipid levels in population-based studies.

J.H. A recent review by Manolio and colleagues¹⁰ had a beautiful figure (FIG. 1) that made the point that the involvement of a particular gene variant (at a certain population frequency and of a certain strength) in a particular disease was on a continuum and that Mendelian diseases are at one end of this continuum. Of course, this reflects what we have known for many years, namely that some Mendelian diseases are 'not fully penetrant'. I think we are at a really exciting time now in seeing this continuum clearly and having systematic ways of attacking risk loci across this continuum. The current practice for identifying variants across the spectrum goes as follows (although of course this is a simplification): for Mendelian disease we use positional cloning; for infrequent, high-risk rare variants we use exome sequencing; and for common, low-risk variants we use GWA studies. This means that, in principle, we have the tools to map anything!

Q Which design priorities would maximize the potential of disease-mapping efforts?

S.E.A. I would list three research priorities. First, there is a need to reduce the cost of sequencing and the false-positive and false-negative rate of high-throughput sequencing. This will lead to the sequencing of hundreds of thousands of human genomes (both germline and somatic), and this new reality will provide the 'final lap' in the genetic analysis of disorders and traits.

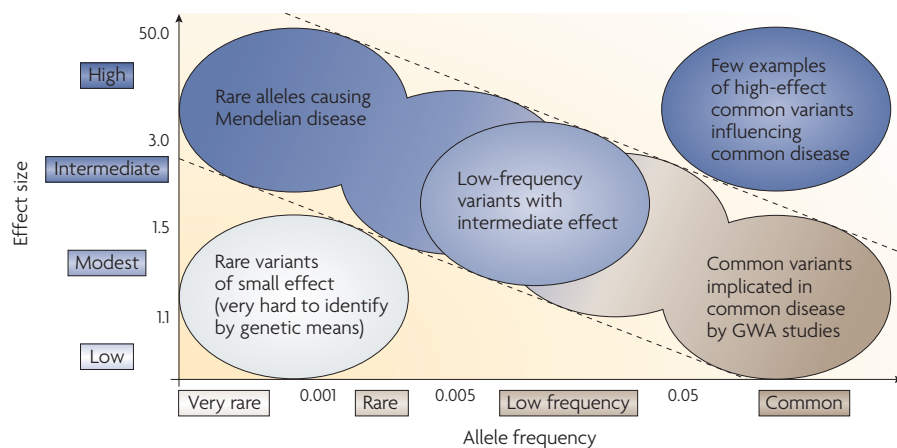


Figure 1 | Feasibility of identifying genetic variants by risk-allele frequency and strength of genetic effect (odds ratio). Reproduced, with permission, from *Nature* REF. 10 © (2009) Macmillan Publishers Ltd. All rights reserved. GWA, genome-wide association.

Second, we ought to identify all functional elements in the human genome, and the functional consequences of the variants in each cell type and at each development stage. This will lead to an understanding of the functional rules for all of the functional elements in a similar way to our current understanding of the functional rules of protein-coding genes.

Third, we should identify the epigenetic modifications of individual genomes and decipher the levels of genetic control that exist in each genome, including those that underlie gene-expression variation. Full knowledge of the connectivity of the functional genomic elements and the network of regulatory elements will identify the fraction of the genome (let's call it the 'medical genome') that needs to be investigated first in relation to human disorders.

A.C. It is clear that we have powerful methods for identifying genes involved in Mendelian traits. Nevertheless, there are numerous unsolved rare phenotypes that cluster in families, are presumed to be monogenic and are amenable to similar analysis. The emergence of 'exome' sequencing should spur this area along nicely. For more common traits, genetic analysis cannot and should not rely on a single design, although such prescriptions are quite popular based on strong untested assumptions about the genetic architecture of multifactorial traits. A major unsolved question in genetics is the inherited basis of complex traits: these are often viewed as either composed of many common alleles of small effect or composed of fewer rare alleles of large effect. It is, of course, most likely that both types of allele exist, as do numerous other interesting genetic scenarios. Consequently, examining the primary phenotype with and without associated phenotypes, in different types of families and using different genetic designs may be the best route to understanding the full spectrum of mutations. In addition, genetic studies with family members that come from different geographies (environments) and clinical challenge tests will become more important for genetic dissection. We will need to design new linkage studies with higher-resolution genetic maps to understand how and why traits show family resemblance. We will also need to engage in a new generation of GWA studies to probe the contribution of common and rare alleles across the frequency spectrum. Finally, everyone expects that DNA sequencing will have a large role in gene discovery, but we have much work to accomplish before whole-genome sequencing on large numbers of samples becomes both routine and affordable.

J.C.C. Three design elements can facilitate the identification of genes for multifactorial and Mendelian diseases. The first is to maximize the use of phenotypic information. For complex traits, selecting individuals with extreme phenotypes is one example. A second approach is to sequence a large sample and use *post hoc* analysis to identify subsets of individuals who share a common genetic aetiology. This strategy takes into account the fact that for many conditions, no single gene accounts for most cases. The third strategy is to use family information, particularly when evaluating the effects of rare mutations identified by resequencing unrelated individuals.

J.H. I am reminded of the great Warren Zevon song 'Lawyers, Guns and Money'. Send samples, kit and money. Samples are needed because to get to the bottom of diseases we need to collect a lot of well-phenotyped samples. We also need lots of expensive kit: array readers and sequencers obviously, but less obviously an enormous increase in our computing infrastructure needs. As for money, these experiments are still extremely expensive; this is a point that is often difficult to comprehend for grant committees, which have a tradition of allocating ~US\$15,000 per labworker per year. When our laboratory started to do whole-genome associations, we were spending US\$25,000 a day on consumables. So to identify the genetic basis of an important disease, I think you need a consortium approach in which all of the groups working on a disease get together and form a network, possibly involving the disease charity, and start working together. Given the expense of such an endeavour, the funders may also need to collaborate. This consortium should then decide which samples within the group should be analysed first and which would make good replication samples. Perhaps most importantly, there should be an insistence on (ethically appropriate) data release, so that the data are not held at a single site but rather are freely available to all who are working or wish to work on the disease.

This mode of research poses severe challenges to the traditional ways of working. In this model it is very difficult to give credit to the people who do the legwork: the Ph.D. student who gets all of the DNA together, the researcher who organizes the laboratory side of the study, the clinician who makes sure all of the ethics and institutional review board forms are in order, and so on. This big science way of working is foreign to most biology laboratories, but it is clearly the way we need to move. This mode of working does have its benefits too: a good consortium

serves as a forum for exciting and data-driven discussions about the mechanistic basis of disease pathogenesis.

Q *How are technologies changing the way in which we study the genetics of disease?*

S.E.A. The evolution of the approaches for identifying the pathogenic variation that causes monogenic or monolocus disorders has been amazing to witness. When I started my genetics career in 1980, the only gene to work on was the β -globin (*HBB*) gene, in which pathogenic mutations were known to cause thalassaemias, sickle cell disease and methemoglobinaemias¹¹. To identify pathogenic variants, it was necessary to work on well-characterized families and to painstakingly clone the genomic region in λ vectors. The discovery of a few polymorphisms (genotyped by restriction analysis and Southern blotting) led Botstein and others¹² to predict the possibility of linkage mapping of monogenic phenotypes. The incredible good luck of mapping a genomic region that is involved in Huntington's disease¹³ using a few polymorphic markers provided proof of principle that Mendelian disorders could be mapped to a narrow genomic interval amenable to further analysis and positional cloning. The introduction of PCR¹⁴ made both the genotyping and the search for genomic variants easier. International projects aimed at improving the infrastructure for genome analysis (such as the Human Genome Project)^{15,16}, the sequencing of the genomes of other species¹⁷, the discovery of all common genetic variation¹⁸, and the development of genotyping and sequencing platforms have led to the discovery of the genetic basis of more than 2,400 Mendelian disorders and traits (see the [OMIM](#) database and the [Human Gene Mutation Database](#)). High-throughput sequencing is now changing the routine again^{3,19}. So far, the discovery chain has been to map first and then to sequence the pathogenic variant²⁰; some now propose that mapping may no longer be necessary, and that the game boils down to using database searches to discover the pathogenic variant among the sea of other variants. If we ignore copy-number variation, which can also contribute to disease, the problem 'simply' becomes a 1 in ~10,000 lottery; given that there are approximately 10,000 non-synonymous, nonsense or splice variants in each exome, which is the one that causes the phenotype? Of course, if one considers that variants in conserved non-coding regions²¹ and other functional genomic

regions could cause Mendelian disorders, the search for pathogenic variants becomes more complicated. But database searches after high-throughput sequencing will not be particularly productive if there is no hypothesis behind them based on an understanding of genetics. We should not forget that conquering genetic disorders requires genetics thinking: mapping and genome sharing among affected individuals continue to be the *sine qua non* tools for establishing a link between a genomic variant and its phenotypic consequences. These two simple tools, which are related to meiosis, usually drastically reduce the search space and aid the identification of the pathogenic variant.

A.C. Model systems gained their genetic power from having induced or single-gene mutations propagated in crosses — in other words, through studies of 'known' variants that were present at high frequencies in the laboratory. However, the study of human genetics depends on natural variation in our populations. Consequently, the mapping of this unknown variation in many humans is central to the practice of human genetics, both in our research and in the clinic. No wonder then that any technology that increases the resolution, speed and accuracy of detecting sequence variation and its effects is crucial: DNA sequencing is the exemplar today. However, DNA sequencing is necessary but not sufficient for understanding the genetics of disease (or a trait) because the biological implications of most sequence alterations are unknown. The genome projects were conceived to deduce the genomic structure so that the biochemical functions encoded in our genomes could be unravelled. Therefore, any technology that will allow us to obtain the genome sequence, to obtain the modifications to this sequence and to read the functions associated with a specific genome will be as crucial, if not more so, than DNA sequencing. In the near future, DNA sequencing is surely going to improve the mapping of traits and diseases. However, we will soon want to know what all the sequence alterations mean. What is still missing in human genetics is a facile experimental system. Although cultured cells (fibroblasts and lymphoblasts) have played a substantial part in aiding our understanding of genetic disease, I expect that the revolution in human stem-cell isolation, differentiation and manipulation (by genes or synthesized molecules) will be more important for understanding the genomic changes and the molecular basis of human disease.

J.C.C. Next-generation sequencing technology has the greatest potential to advance genetics because it allows the identification of rare sequence variations with large phenotypic effects. For Mendelian disorders, whole-genome (or whole-exome) sequencing provides the opportunity to identify genes in pedigrees that are not tractable to linkage analysis. For complex disorders, genome-wide sequencing of large case-control series will pinpoint mutations that have a significant impact on disease susceptibility in subsets of individuals.

J.H. It is clear how technology changes are driving our genetic research: geneticists are not getting cleverer, they are just getting better tools. But perhaps less obvious, and in the long term more important, is the way that '-ome' technologies are driving all of biology. Hypothesis-driven biology has dominated the way we have done biology for the past 50 years, but it was not always so. Linnaeus was an -ome biologist: he wanted to catalogue all of life. I think we are entering a period in which more and more aspects of biology will benefit from the application of the unbiased screening technologies that genetics has made so familiar.

*Stylianios E. Antonarakis is at the Department of Genetic Medicine and Development, University of Geneva Medical School, 1 Rue Michel-Servet, 1211 Geneva, Switzerland; and the University Hospitals of Geneva, 4 Rue Gabrielle-Perret-Gentil, 1211 Geneva, Switzerland.
e-mail: Stylianios.Antonarakis@unige.ch*

*Aravinda Chakravarti is at the Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Broadway Research Building, Suite 579, 733 North Broadway, Baltimore, Maryland 21205, USA.
e-mail: aravinda@jhmi.edu*

*Jonathan C. Cohen is at the University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9052, USA.
e-mail: jonathan.cohen@utsouthwestern.edu*

*John Hardy is at the Department of Molecular Neuroscience and Reta Lilla Weston Laboratories, Institute of Neurology, Queen Square, London WC1N 3BG, UK.
e-mail: j.hardy@ion.ucl.ac.uk*

doi:10.1038/nrg2793

- Gibbs, J. R. & Singleton, A. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. *PLoS Genet.* **2**, e150 (2006).
- van de Leemput, J. & Chandran, J. *et al.* Deletion at *ITPR1* underlies ataxia in mice and spinocerebellar ataxia 15 in humans. *PLoS Genet.* **3**, e108 (2007).
- Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
- Singleton, A., Myers, A. & Hardy, J. The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex diseases. *Hum. Mol. Genet.* **13**, R123–R126 (2004).

- Hutton, M. *et al.* Association of missense and 5'-splice-site mutations in *tau* with the inherited dementia FTDP-17. *Nature* **393**, 702–705 (1998).
- Harold, D., Abraham, R. & Hollingworth, P. *et al.* Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. *Nature Genet.* **41**, 1088–1093 (2009).
- Lambert, J.-C. *et al.* Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. *Nature Genet.* **41**, 1094–1099 (2009).
- McGeer, E. G., Klegeris, A. & McGeer, P. L. Inflammation, the complement system and the diseases of aging. *Neurobiol. Aging* **26** (Suppl. 1), 94–97 (2005).
- Alexander, J. J., Anderson, A. J., Barnum, S. R., Stevens, B. & Tenner, A. J. The complement cascade: Yin–Yang in neuroinflammation — neuro-protection and -degeneration. *J. Neurochem.* **107**, 1169–1187 (2008).
- Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Weatherall, D. J. Phenotype–genotype relationships in monogenic disease: lessons from the thalassaemias. *Nature Rev. Genet.* **2**, 245–255 (2001).
- Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331 (1980).
- Gusella, J. F. *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234–236 (1983).
- Saiki, R. K. *et al.* Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354 (1985).
- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
- Ng, S. B. *et al.* Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genet.* **42**, 30–35 (2010).
- Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
- Dermitzakis, E. T., Reymond, A. & Antonarakis, S. E. Conserved non-genic sequences — an unexpected feature of mammalian genomes. *Nature Rev. Genet.* **6**, 151–157 (2005).

Competing interests statement

A.C. and J.H. declare competing financial interests: see Web version for details.

DATABASES

Entrez Gene: <http://www.ncbi.nlm.nih.gov/gene>
HBB | HMGCR | MAPT | PCSK9

Human Gene Mutation Database:

<http://www.hgmd.cf.ac.uk/ac/index.php>

OMIM: <http://www.ncbi.nlm.nih.gov/omim>

Hirschsprung disease | Marfan syndrome | progressive supranuclear palsy

UniProtKB: <http://www.uniprot.org>
fibrillin 1

FURTHER INFORMATION

Stylianios E. Antonarakis's homepage: http://www.medicine.unige.ch/recherche/groupe/b_donnees/sujet_248_1.html

Aravinda Chakravarti's homepage:

<http://chakravarti.igmm.jhmi.edu>

Jonathan C. Cohen's homepage: <http://www.utsouthwestern.edu/utsw/cda/dept/108278/files/110591.html>

John Hardy's homepage:

<http://www.researcherid.com/rid/C-2451-2009>

AnEuploidy: <http://www.aneuploidy.eu>

Frontiers in Genetics: <http://www.frontiers-in-genetics.org>

'Lawyers, Guns and Money' by Warren Zevon:

<http://www.youtube.com/watch?v=S5puAN1PGQw>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF