PERSPECTIVES

OPINION

Mendelian disorders deserve more attention

Stylianos E. Antonarakis and Jacques S. Beckmann

Abstract | The study of inherited monogenic diseases has contributed greatly to our mechanistic understanding of pathogenic mutations and gene regulation, and to the development of effective diagnostic tools. But interest has gradually shifted away from monogenic diseases, which collectively affect only a small fraction of the world's population, towards multifactorial, common diseases. The quest for the genetic variability associated with common traits should not be done at the expense of Mendelian disorders, because the latter could still contribute greatly to understanding the aetiology of complex traits.

According to latest estimates, there are barely 25,000 protein-coding genes in the human genome[1] (encoding perhaps an order of magnitude more distinct gene products). Although this number will probably be revised many times in the near future, an important and challenging question currently occupies the minds of biomedical researchers, funding agencies, patient groups and biotechnology and drug companies: what is the function(s) of each of these genes, and how does genetic variation in each gene contribute to health and disease? Indeed, understanding the interactions that occur between gene products, and the complex and hierarchical network of biochemical, signal-transduction and developmental pathways in which these products participate, is necessary to design effective treatments for the myriad of rare and common inherited disease phenotypes.

But what is the best way to begin to understand gene function? There is no single answer to this question. Human genetics, biochemistry, bioinformatics and the study of model organisms each provide complementary ways to functionally annotate the human genome. Large efforts are currently under way — using all the above approaches — to identify the functional elements in the human genome[2].

We argue that a timely and low-risk option for elucidating gene function is to link naturally occurring pathogenic mutations with monogenic disorders (monolocus disorders). The process is well known to medical geneticists: in the past 20 years, pathogenic allelic variants of a total of 1,822 genes — less than 10% of our gene repertoire — have been found to cause highly penetrant, monogenic Mendelian diseases (as of 6 February 2006 in OMIM, a comprehensive knowledge-based database of human genes and Mendelian phenotypes[3,4], see Further information).

However, over the past few years there has been a steady, in our view unjustified, move away from Mendelian disorders and towards the more common multifactorial diseases. Here, we focus on current trends in medical genetic research, highlighting lessons learned from the study of Mendelian disorders. We emphasize that the latter represent not only an indispensable treasure trove for elucidating gene function and for reconstructing normal and pathological pathways, but that they could also contribute significantly to our understanding of the molecular genetic basis of common, complex diseases. Therefore, research on Mendelian traits should be encouraged.

Drifting away from Mendelian disorders

Many of the genes that underlie human genetic disease were discovered before the sequencing of the euchromatic portion of the human genome was completed, before the introduction of methods for genome-wide genotyping and gene-expression analyses, and before the discovery of the extraordinary variability among human genomes. It is therefore odd that after the complete sequencing of the human and other genomes, interest in monogenic disorders declined in favour of the strong current research focus on complex and common phenotypes. This trend is illustrated in FIG. 1, which shows the number of novel disease–gene matches that have been recorded on OMIM in the past 15 years (FIG. 1a) and the number of pathogenic mutations recorded in the Human Gene Mutation Database[5] (HGMD) (FIG. 1b). Paradoxically, and contrary to expectations, after the sequencing of the euchromatic portion of the human genome was completed, the number of newly discovered gene–disease matches and additional pathogenic mutations has declined compared with the previous decade, in which only more primitive tools for gene discovery were available. A comparison of the number of disease-gene identifications that were published in *Nature Genetics* during the first six months of 2000 (n=37) and 2005 (n=14 and mostly as brief communications) is but another illustration of this tendency. This example alone cannot, of course, be taken as an indication of the type of research that is being carried out: because gene mapping and cloning are considered to have become easier, such studies are less likely to be published in top-tier journals. Furthermore, the availability of funding resources, which favour the study of common (and therefore commercially attractive) targets over that of rare orphan diseases, has further encouraged this trend.

Besides genuine interest and editorial and funding policies, additional factors have contributed to the shift in emphasis in human genetics research. Most highly prevalent, highly penetrant and easily diagnosed monogenic entities were elucidated first. Consequently, we are now faced with a proportionately larger number of rare Mendelian disorders; the number of these disorders is hard to estimate (we predict, based on knockout gene experiments in yeast and worms[6,7], that there are more

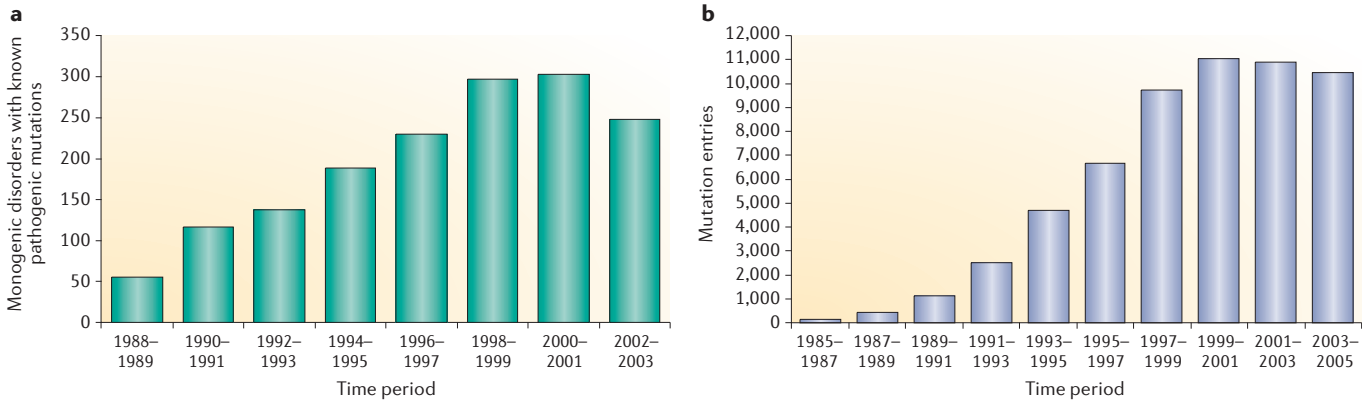

Figure 1 | **Progress in disease-gene identification. a** | Graph depicting the number of allelic variants (pathogenic mutations that cause monogenic disorders) recorded in OMIM during the indicated time intervals. **b** | Graph depicting the number of disease-associated mutation entries in the Human Gene Mutation Database published in the indicated time intervals.

than several thousand; perhaps at least one per gene), and the conventional strategies to map and identify a pathogenic locus or variant might not apply. Elucidation of these loci calls for the development of alternative powerful strategies and tools.

Although studies on the genetic predisposition to complex phenotypes should be an important focus of activities in the coming years, we wish to emphasize the tremendous power of monogenic phenotypes to aid the classification and understanding of human diseases. Studies on Mendelian traits have provided unprecedented insights into mutation processes and molecular pathophysiology: they have revealed the existence of genetic phenomena such as uniparental disomy, parental imprinting and epistatic interactions. However, despite the impressive progress made, a fundamental understanding of the biological cause of most of these diseases is still lacking, warranting further research in this area. We argue that, with the development of the genomic infrastructure and the advancement of technological platforms, we now have the best opportunity to link, directly or indirectly, most genes with aberrant phenotypes. Some of the relevant technologies are discussed in BOX 1.

Are we through with Mendelian disorders?

Some investigators might argue that there are not enough monogenic disorders left to study. This is probably incorrect: there are currently more than 1,500 known monogenic phenotypes for which the defective gene remains elusive (see OMIM database). Many of these are rare and occur only in a few families. Adequate clinical and technological resources need to be established to allow these precious samples to be shared as part of collaborative efforts.

In addition, there are large, poorly studied populations that have high levels of consanguinity, and that might segregate many unknown recessive monogenic phenotypes. As many as 20–50% of marriages are consanguineous worldwide, involving approximately 1 billion people[8]. Numerous but as-yet uncharted isolated populations also harbour a high frequency of recessive alleles due to founder effects and inbreeding.

Also, most monogenic causes of foetal wastage and subfertility are currently unexplored[9]. To investigate the causes of foetal wastage, there have been calls to collect and preserve, whenever possible, the affected tissues, which would then allow genetic analyses to be complemented by RNA- or protein-expression profiles[10]. Because many of these foetal defects are also characterized by sub-microscopic chromosomal abnormalities (detectable by the new high-resolution tiling path CGH arrays[11]), we would recommend that the existing large sequencing capacity also be used to systematically sequence around these rearrangements. This approach could uncover new pathological mutations as well as provide the basis for a molecular explanation for dysmorphic features. A successful application of this method is the discovery that mutations in the *CHD7* (chromodomain helicase DNA-binding protein 7) gene are one cause of CHARGE syndrome[12].

Besides recessive traits, an unknown number of dominant and X-chromosome-linked monogenic phenotypes might exist in African and other populations that are not adequately studied. A concerted and well-organized effort needs to be initiated (preferably led by international organizations or local interested groups) to collect samples from accurately diagnosed individuals and their families from selected communities. This is in practice the most difficult, yet essential, task in the discovery process, as the experimental aspects are now almost off-the-shelf endeavours.

Human or mouse genetic models?

Some investigators argue that the gene-knockout project of all ORFs in mice and the precise assessment of the resulting phenotypes[13,14] will provide similar information in a much more controlled, systematic and comprehensive way to human molecular genetic studies. There are several shortcomings of this approach; first, not all human genes have a mouse orthologue and vice versa, and second, several clinically relevant phenotypes have human-specific manifestations (such as mental retardation or cognitive behaviour), which might be difficult to assay or reproduce in mice. Even when phenotypes are easily scorable and the corresponding genes are present in both species, it is important to emphasize that an understanding of gene dysfunction is better accomplished by analysing an allelic series of (naturally occurring) mutations than simply by looking at the consequences of sizable gene deletions or other inactivating mutations.

Furthermore, in many instances, similar phenotypes can be caused by mutations in any of several distinct genes (a phenomenon known as genetic heterogeneity) and, in other cases, different mutations within the same gene can give rise to distinct phenotypes; the lamin A/C gene, in which an allelic series of mutations results in up to six different clinical phenotypes[15], constitutes a particularly illustrative example. In other instances, the same pathologic mutation, depending on the genetic background in which it occurs, might display different

expressivity, penetrance or even distinct diseases (for example, in limb girdle muscular dystrophy type 2B and Miyoshi myopathy[16,17]). Yet in other diseases, such as in hereditary prion disease, the phenotype is determined by the combined effect of pathogenic mutations and polymorphic variants[18]. These subtleties might not be picked up by using mouse genetic models alone.

Also, phenotypic assessment in mice is not as complete and well-developed and sophisticated as it is in humans. One essential advantage in human biology is the large number of patient–physician contacts and the many more laboratory and imaging tests done in humans that uncover genetically related phenotypes.

Finally, there are considerable differences in the phenotypic spectra between mice and humans. Mice have a shorter lifespan and some of the late-onset human phenotypes are not apparent in them. Foetal development also differs between man and mice. Mutations that have severe consequences in humans could be mild in mice, such as the correlation between *APP* (amyloid precursor protein) mutations and Alzheimer disease[19], and *HPRT* (hypoxanthine guanine phosphoribosyltransferase) mutations and Lesch–Nyhan syndrome[20]. Environmental challenges, including infections, affect mice in a different way to humans and are also less frequent. Furthermore, there are important differences in brain structure and function between these species.

Lessons from monogenic disorders

Novel genetic mechanisms. Our present knowledge of the different types of pathogenic mutations, and their consequences and pathogenic mechanisms, has been discovered in the course of the ongoing quest for Mendelian-disease loci[4,5]. For example, changes in triplet-repeat copy number are now an established genetic mutation mechanism[21], but they were totally unexpected when first reported.

Phenotypic complexity and modifier genes. Another important contribution of monogenic disorders to understanding gene function derives from their genetic and allelic heterogeneity, as well as phenotypic variability. With over one hundred deafness-related loci recognized so far, mutations in the same gene (or even a single mutant allele) can result in different individuals in either syndromic, isolated deafness or asymptomatic carriers. Mutations in one of the most prevalent deafness alleles in the connexin 26 gene (*CX26*), can be inherited in either a dominant or recessive mode[22]. In cystic fibrosis, subtle variations in the number of simple nucleotide repeats in introns can result in phenotypic and clinical variability[23]. There are many more examples that illustrate the variability in the penetrance, expressivity and severity of disease among mutation carriers, even within single families, pointing towards the involvement of additional genetic factors or modifier genes[24,25]. These examples raise the possibility that monogenic entities, like complex traits, could also reflect a continuous phenotypic distribution (as illustrated by Bardet–Biedl syndrome and Meckel syndrome[26]).

The nature and influence of most of the modifier loci that modulate the clinical phenotypes of Mendelian traits remain unknown. Such loci could be among the first beneficiaries of the HapMap project[27–29] (see Further information). Cystic fibrosis and sickle-cell anaemia, for example, represent primary targets for an exhaustive search for modifiers. The knowledge and statistical tools developed in the course of this endeavour will in return benefit the exploration of complex traits.

To add to the phenotypic complexity, the study of discordant monozygotic twins carrying the same mutation embedded in identical genomes could also be informative on the nature of the multiple epigenetic, environmental and stochastic factors or processes contributing to disease manifestation. An example is the case of two pairs of sisters carrying, respectively, a *DMD* (Duchenne muscular dystrophy) mutation[30] or a Gly2019Ser mutation in the *LRRK2* gene that is associated with Parkinson disease[31]. Older monozygous twins exhibit remarkable differences in the overall content and genomic distribution of

Box 1 | **From phenotype to gene**

New genotyping methods. The use of genome-wide linkage analysis to map the critical genomic intervals that contain a disease-causing allele has become much easier; this is due to the development of novel genotyping methods (for example, the platforms developed by Affymetrix, Sequenom and Illumina) and to the results of the HapMap project, which provide information on common SNP variants and their allele frequencies in human populations[27–29]. By using these enhanced high-resolution genotyping methods, it should be possible, even for rare diseases, to identify short genomic intervals that are shared by affected individuals; common disease-associated haplotypes could be identified in unrelated patients who carry a common ancestral founder mutation, thereby pinpointing the location of the disease allele.

Enhanced genomic analysis and bioinformatics tools. Genome sequencing and comparative genome analyses have uncovered additional functional genomic elements that could be targets of pathogenic mutations; these include the functionally conserved sequences that are not protein-coding[57], miRNAs and other non-coding RNAs[58,59], and numerous insertion–deletion/copy-number variations of different human genomes[53–55]. Therefore, positional cloning is now much easier because of the precise knowledge of the human sequence[1] and the availability of rapid methods for mutation analysis, which in combination can uncover both point mutations and larger genomic lesions. Determination of gene-expression differences between affected and normal cells and tissues can also pinpoint candidate genes. What used to be a laborious effort could now be greatly facilitated by merging various methodologies and bioinformatics tools.

Model organisms. The use of sequence conservation and experimentation in model organisms provides ways to validate the dysfunctional nature of mutant proteins. It is worth noting that the principal bottleneck is not the mapping of a disease locus or the identification of candidate pathogenic alleles, but the demonstration of the pathogenicity of the true disease-related allele and its subsequent pathophysiological characterization.

Advanced molecular genetic tools. The increased ease with which conditional or tissue-specific (knocked-in or knocked-out) animal or cellular models can be generated is a great adjunct to understanding and validating the functions of suspected disease-linked genes. Indeed, linking a mutation within a defined gene or functional genomic element to a particular phenotype remains essential to assign a function to this genomic element.

Interaction databases. Databases of protein–protein interactions increase the number of candidate molecules that could underlie phenotypically related disorders[60–62]. For example, Fanconi anaemia subtypes are caused by mutant alleles of genes encoding proteins that form a complex involved in DNA repair, function as a downstream effector of this complex or modify this effector[63]. Using these mutated alleles as 'Ariadne's thread' allows one to track additional genes involved in related pathologies, thereby progressively reconstructing the normal and pathophysiological pathways in which these genes participate.

5-methylcytosine DNA and histone acetylation, affecting their gene-expression profile; these epigenetic differences might partially explain how different phenotypes can originate from the same genotype[32].

A continuum from simple to complex disorders. The above examples show that the concept of a monogenic disorder can no longer be understood *sensu stricto*. These disorders provide instructive examples of oligogenic inheritance, such as digenic inheritance[33–35], that in turn provide clues to understanding polygenic inheritance. Allelic variation in genes or other functional DNA sequences that modify the phenotypic severity of a monogenic disorder or control variation in gene expression provide links to additional genomic causes that are related to phenotypic variability. Mutations that affect regulatory sequences, such as promoters, or mRNA stability and turnover rate[36] can modulate variation in expression levels and so result in phenotypic alterations. Furthermore, regulated co-expression or biochemical interactions of gene products also result in more gene–phenotype links. Eventually, starting from an allelic variant that causes a monogenic disorder, one could end up with several other genes linked to the same phenotype, a situation similar to the expected molecular genetic causes of polygenic complex phenotypes. A good example is that of thalassaemias, sickle-cell disease or the methaemoglobinaemias. Allelic variation that underlies phenotypic variability in these disorders lies not only within the β-globin gene on chromosome 11 (haemoglobin β-chain, *HBB*) but also in *HBG*, *HBA* (haemoglobin α-chain on chromosome 16), X-linked *ATRX* (α-thalassaemia/mental retardation) and other genes[25]. Furthermore, more attention to the phenotypic variability of a Mendelian trait could unravel novel functions or interactions of the mutant gene. A striking recent example is that of *HBS* heterozygosity, which not only protects from malaria but also enhances acquired immunity to the parasite[37].

Caveats. We do not wish to give the misleading impression that research on Mendelian disorders is simple and that the molecular genetic basis of monolocus diseases can be understood shortly after the identification of the mutant allele. As mentioned above, the identification of genetic modifiers is complex. Even for the most common and well-studied Mendelian disorders such as sickle-cell disease, thalassaemias and cystic fibrosis, in which some modifiers have been identified, our current knowledge of all the genomic variation that results in phenotypic variability is rather poor[24,25,38]. Studies in mice suggest experimental approaches to identify certain genetic modifiers[39].

Proving that a pathogenic allele is causal to the disease in question is another difficult challenge. Much has been learnt about the phenotypic consequences of mutations[40], but there are many examples of missense mutations, variants in DNA elements of unknown function and silent changes in coding regions for which pathogenicity is questionable.

Conclusions

The classic argument for advocating more research on monogenic disorders is more relevant than ever: a mutant gene that gives rise to a monogenic disorder usually provides an opportunity to understand the allelic variability for susceptibility to a similar common polygenic phenotype. For example, the discovery of *APC* (adenomatous polyposis coli) mutations in hereditary colon cancer led to the discovery of additional genes, mutant alleles of which also cause or predispose to colon cancer. Additional gene interactions might reveal the full spectrum of mutant alleles related to this common phenotype. Furthermore, many rare allelic variants in three candidate genes — *ABCA1* (ATP-binding cassette A1), *APOA1* (apolipoprotein A1), and *LCAT* (lecithin:cholesterol acyltransferase) — that cause pathogenically low levels of HDL-cholesterol in plasma were also found in individuals with the common, complex version of the low-HDL-cholesterol trait[41,42]. Similarly, the identification of loci for susceptibility to schizophrenia, mapped by studying monogenic diseases[43–44], might contribute significantly to unraveling this severe condition.

It is widely accepted that most of the lessons learnt from Mendelian disorders will help us to identify susceptibility alleles for complex phenotypes. Observations of mutation mechanisms and phenomena such as anticipation[45], gene dosage effects[46], uniparental disomy[47], imprinting[48], variation in copy-number repeats and effects of allelic series not only in dichotomous but also in quantitative traits (for example, different haemoglobin levels from different β+ mutations in the *HBB* gene[25]) are additional examples of the complex issues that emerged from seemingly simple Mendelian traits. Good examples of disease-associated variation in copy-number repeats include the contraction of the D4Z4 repeats in facioscapulohumeral muscular dystrophy[49], and the gene-dosage

Glossary

Anticipation
A phenomenon that describes the property of some disorders to increase in severity in successive generations.

Copy-number polymorphism (CNP) or variant (CNV)
A structural genomic variant, resulting in confined copy-number changes in a defined chromosomal region. If its population allele frequency is less than 1%, then one refers to it as a variant; the term polymorphism refers instead to variants that occur at an allelic frequency of 1% or higher.

Discordant monozygotic twins
Twins that share identical genomes but are not affected with the same disorder; one twin being affected but the other not.

Epistatic interaction
The influence of the interaction of multiple loci on phenotypic variation.

Expressivity
The extent to which a genetic defect is expressed.

High-resolution tiling path CGH arrays
Arrays for comparative genomic hybridization (CGH) offering a resolution in the order of kilobases. The arrays are currently based on BACs or long oligonucleotides.

Mendelian disease
A disease for which alternative genotypes fall into distinct, discrete phenotypic classes, following Gregor Mendel's laws of inheritance.

Monogenic disorder
A disease that is mainly caused by variants in a single gene.

Multifactorial
Multifactorial traits are determined by a combination of genetic as well as non-genetic factors, each contributing to the overall phenotype.

Parental imprinting
Differential allele expression that depends on parental origin.

Penetrance
The frequency with which a given genotype manifests itself in a given phenotype.

Triplet-repeat copy number
Triplet-repeat copy-number mutations are dynamic mutations that are due to the expansion or contraction in the number of triplet nucleotide repeats.

Uniparental disomy
(UPD). A state wherein both homologues (alleles) at a locus derive from the same parent; for some chromosomal segments, UPD generates characteristic syndromes.

effects of the normal, unmutated versions of *APP* in one form of Alzheimer disease[50] and of α-synuclein in Parkinson disease[51,52]. Last but not least, the discovery[53–55] of the prevalence and abundance of copy-number polymorphisms (CNPs) or variants (CNVs) as well as insertions/deletions now presents us with the interesting challenge to differentiate between neutral submicroscopic rearrangements and those that contribute to phenotypic variability. The recent demonstration that copy-number variation in the *FCGR3* gene predisposes to glomerulonephritis is an instructive first example of the role of such variation in the aetiology of phenotypes[56].

There is no *a priori* reason to believe that any of these diverse and complex issues regarding mutation, gene regulation and epigenetic processes that are seen in monogenic diseases do not also contribute to common multifactorial traits. It is probable that additional clues for understanding complex phenotypes will come first from a fresh and more sophisticated look at these instructive Mendelian traits. So, all lessons from the latter are more than relevant for the new genetic challenge. Moreover, the exploration of rare orphan diseases might necessitate the development of new identification strategies which, once validated, will also benefit the dissection of complex traits.

We argue that the time is now opportune for more research into the genetic dissection of monogenic disorders, as they are instructive about molecular mechanisms, and therefore can lead the way to understanding complex traits. This effort will provide clues for gene function and aid the annotation of the human genome. This is not to say that efforts for the molecular analysis of polygenic, complex phenotypes is not appropriate or timely; we simply argue that conventional monogenic phenotypes, because they themselves can be oligogenic and complex, are likely to continue to provide unexpected and ground-breaking knowledge towards the common goal of developing early diagnosis and treatment for all genetic disorders.

Stylianos E. Antonarakis is at the Department of Genetic Medicine and Development, University of Geneva Medical School, and University Hospital of Geneva, 1 rue Michel-Servet, 1211 Geneva, Switzerland.

Jacques S. Beckmann is at the Department of Medical Genetics, University of Lausanne and Centre Hospitalier Universitaire Vaudois, 2 Avenue Pierre Decker, 1011 Lausanne, Switzerland.

e-mails: Stylianos.Antonarakis@medecine.unige.ch; Jacques.Beckmann@chuv.ch

doi:10.1038/nrg1826
Published online 14 March 2006

1. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
2. The ENCODE (ENCyclopedia Of DNA Elements) project. *Science* **306**, 636–640 (2004).
3. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514–D517 (2005).
4. Antonarakis, S. E. & McKusick, V. A. OMIM passes the 1,000-disease-gene mark. *Nature Genet.* **25**, 11 (2000).
5. Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* **21**, 577–581 (2003).
6. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
7. Sonnichsen, B. *et al.* Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434**, 462–469 (2005).
8. Bittles, A. Consanguinity and its relevance to clinical genetics. *Clin. Genet.* **60**, 89–98 (2001).
9. Hackstein, J. H., Hochstenbach, R. & Pearson, P. L. Towards an understanding of the genetics of human male infertility: lessons from flies. *Trends Genet.* **16**, 565–572 (2000).
10. Strachan, T., Abitbol, M., Davidson, D. & Beckmann, J. S. A new dimension for the human genome project: towards comprehensive expression maps. *Nature Genet.* **16**, 126–132 (1997).
11. de Vries, B. B. *et al.* Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* **77**, 606–616 (2005).
12. Vissers, L. E. *et al.* Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nature Genet.* **36**, 955–957 (2004).
13. Auwerx, J. *et al.* The European dimension for the mouse genome mutagenesis program. *Nature Genet.* **36**, 925–927 (2004).
14. Austin, C. P. *et al.* The knockout mouse project. *Nature Genet.* **36**, 921–924 (2004).
15. Ben Yaou, R. *et al.* Genetics of laminopathies. *Novartis Found. Symp.* **264**, 81–90; discussion 90–97, 227–230 (2005).
16. Liu, J. *et al.* Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nature Genet.* **20**, 31–36 (1998).
17. Bashir, R. *et al.* A gene related to *Caenorhabditis elegans* spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. *Nature Genet.* **20**, 37–42 (1998).
18. Gambetti, P., Parchi, P. & Chen, S. G. Hereditary Creutzfeldt–Jakob disease and fatal familial insomnia. *Clin. Lab. Med.* **23**, 43–64 (2003).
19. Phinney, A. L. *et al.* Mouse models of Alzheimer's disease: the long and filamentous road. *Neurol. Res.* **25**, 590–600 (2003).
20. Kuehn, M. R., Bradley, A., Robertson, E. J. & Evans, M. J. A potential animal model for Lesch–Nyhan syndrome through introduction of HPRT mutations into mice. *Nature* **326**, 295–298 (1987).
21. Pearson, C. E., Edamura, K. N. & Cleary, J. D. Repeat instability: mechanisms of dynamic mutations. *Nature Rev. Genet.* **6**, 729–742 (2005).
22. Rabionet, R., Gasparini, P. & Estivill, X. Molecular genetics of hearing impairment due to mutations in gap junction genes encoding β connexins. *Hum. Mutat.* **16**, 190–202 (2000).
23. Buratti, E., Brindisi, A., Pagani, F. & Baralle, F. E. Nuclear factor TDP-43 binds to the polymorphic TG repeats in CFTR intron 8 and causes skipping of exon 9: a functional link with disease penetrance. *Am. J. Hum. Genet.* **74**, 1322–1325 (2004).
24. Cutting, G. R. Modifier genetics: cystic fibrosis. *Annu. Rev. Genomics Hum. Genet.* **6** 237–260 (2005).
25. Weatherall, D. J. Phenotype–genotype relationships in monogenic disease: lessons from the thalassaemias. *Nature Rev. Genet.* **2**, 245–255 (2001).
26. Karmous-Benailly, H. *et al.* Antenatal presentation of Bardet–Biedl syndrome may mimic Meckel syndrome. *Am. J. Hum. Genet.* **76**, 493–504 (2005).
27. Hinds, D. A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072–1079 (2005).
28. Consortium., T. I. H. The international HapMap project. *Nature* **426**, 789–796 (2003).
29. Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
30. Abbadi, N. *et al.* Additional case of female monozygotic twins discordant for the clinical manifestations of Duchenne muscular dystrophy due to opposite X-chromosome inactivation. *Am. J. Med. Genet.* **52**, 198–206 (1994).
31. Bras, J. M. *et al.* G2019S dardarin substitution is a common cause of Parkinson's disease in a Portuguese cohort. *Mov. Disord.* **20**, 1653–1655 (2005).
32. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* **102**, 10604–10609 (2005).
33. Kajiwara, K., Berson, E. L. & Dryja, T. P. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**, 1604–1608 (1994).
34. Katsanis, N. *et al.* Triallelic inheritance in Bardet–Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256–2259 (2001).
35. Beckmann, J. S. The Reunion paradox and the digenic model. *Am. J. Hum. Genet.* **59**, 1400–1402 (1996).
36. Wang, D., Johnson, A. D., Papp, A. C., Kroetz, D. L. & Sadee, W. Multidrug resistance polypeptide 1 (MDR1, ABCB1) variant 3435C > T affects mRNA stability. *Pharmacogenet. Genomics* **15**, 693–704 (2005).
37. Williams, T. N. *et al.* An immune basis for malaria protection by the sickle cell trait. *PLoS Med.* **2**, e128 (2005).
38. Nagel, R. L. Epistasis and the genetics of human diseases. *C. R. Biol.* **328**, 606–615 (2005).
39. Nadeau, J. H. Modifier genes in mice and humans. *Nature Rev. Genet.* **2**, 165–174 (2001).
40. Antonarakis, S. E., Krawczak, M., Cooper, D. M. in *The Metabolic and Molecular Basis of Inherited Disease* (eds Scriver, C. R. *et al.*) 343–378 (McGraw–Hill, New York, 2001).
41. Cohen, J. C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869–872 (2004).
42. Frikke-Schmidt, R., Nordestgaard, B. G., Jensen, G. B. & Tybjaerg-Hansen, A. Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J. Clin. Invest.* **114**, 1343–1353 (2004).
43. Lindsay, E. A. *et al.* Schizophrenia and chromosomal deletions within 22q11.2. *Am. J. Hum. Genet.* **56**, 1502–1503 (1995).
44. Millar J. K., *et al.* DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* **310**, 1187–1191 (2005).
45. Harper, P. S., Harley, H. G., Reardon, W. & Shaw, D. J. Anticipation in myotonic dystrophy: new light on an old problem. *Am. J. Hum. Genet.* **51**, 10–16 (1992).
46. Ibanez, P. *et al.* Causal relation between α-synuclein gene duplication and familial Parkinson's disease. *Lancet* **364**, 1169–1171 (2004).
47. Ledbetter, D. H. & Engel, E. Uniparental disomy in humans — development of an imprinting map and its implications for prenatal diagnosis. *Hum. Mol. Genet.* **4**, 1757–1764 (1995).
48. da Rocha, S. T. & Ferguson-Smith, A. C. Genomic imprinting. *Curr. Biol.* **14**, R646–R649 (2004).
49. Tupler, R. & Gabellini, D. Molecular basis of facioscapulohumeral muscular dystrophy. *Cell. Mol. Life Sci.* **61**, 557–566 (2004).
50. Rovelet-Lecrux, A. *et al. APP* locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nature Genet.* **38**, 24–26 (2006).
51. Singleton, A. B. *et al.* α-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841 (2003).
52. Eriksen, J. L., Przedborski, S. & Petrucelli, L. Gene dosage and pathogenesis of Parkinson's disease. *Trends Mol. Med.* **11**, 91–96 (2005).
53. Sharp, A. J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
54. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
55. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature Rev. Genet.* **7**, 85–97 (2006).
56. Aitman, T. J. *et al.* Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
57. 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).
58. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.* **5**, 522–531 (2004).

59. Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.* **2**, 919–929 (2001).
60. Brunner, H. G. & van Driel, M. A. From syndrome families to functional genomics. *Nature Rev. Genet.* **5**, 545–551 (2004).
61. Gavin, A. C. & Superti-Furga, G. Protein complexes and proteome organization from yeast to man. *Curr. Opin. Chem. Biol.* **7**, 21–27 (2003).
62. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, 399–403 (2002).
63. D'Andrea, A. D. & Grompe, M. The Fanconi anaemia/BRCA pathway. *Nature Rev. Cancer* **3**, 23–34 (2003).

Acknowledgements
We thank all the past and present members of our laboratories for ideas, debates and discussions. Mrs J. Amberger at OMIM, John Hopkins University, is gratefully acknowledged for her assistance in providing the data of figure 1a. S.E.A.'s laboratory is supported by the Swiss National Science Foundation (SNF), European Union, National Institutes of Health and the Lejeune and ChildCare Foundations. Work in J.S.B.'s laboratory is funded by grants from the SNF and the University of Lausanne.

Competing interests statement
The authors declare no competing financial interests.

DATABASES

The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
α-synuclein | *ABCA1* | *APC* | *APOA1* | *APP* | *ATRX* | *CHD7* | *CX26* | *DMD* | *HBA* | *HBB* | *HBG* | *HPRT* | lamin A/C | *LCAT* | *LRRK2*
OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Alzheimer disease | Bardet–Biedl syndrome | CHARGE syndrome | cystic fibrosis | Duchenne muscular dystrophy | facioscapulohumeral muscular dystrophy | Fanconi anaemia | hereditary prion disease | Lesch–Nyhan syndrome | limb girdle muscular dystrophy type 2B | Meckel syndrome | Miyoshi myopathy | Parkinson disease | sickle-cell anaemia | thalassaemias

FURTHER INFORMATION

Online Mendelian Inheritance in Man (OMIM): http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Human Gene Mutation Database (HGMD): http://www.hgmd.cf.ac.uk
The International Hapmap Project: http://www.hapmap.org
Affymetrix: http://www.affymetrix.com/index.affx
Illumina: http://www.illumina.com
Sequenom: http://www.sequenom.com
The ENCODE Project: http://www.genome.gov/10005107
Access to this links box is available online.