Genome **Biology**

Unlocking Mendelian disease using exome sequencing

Christian Gilissen*, Alexander Hoischen, Han G Brunner and Joris A Veltman

Abstract

Exome sequencing is revolutionizing Mendelian disease gene identification. This results in improved clinical diagnosis, more accurate genotype-phenotype correlations and new insights into the role of rare genomic variation in disease.

The identification of the causative mutation for a Mendelian disease enables molecular diagnosis and carrier testing in the patient and his or her family. This is of great importance for patient management and family counseling, and serves as a starting point for therapeutic interventions [1]. Furthermore, the identification of Mendelian disease genes contributes to our understanding of gene functions and biological pathways underlying health and disease in general [2], and lessons learned from rare diseases are often also relevant to common disease [3]. Research aimed at the identification of genes that cause Mendelian disease has received a boost over the past couple of years by the introduction of new technologies that enable the sequencing of DNA at a much higher throughput and at much lower costs than previously possible [4].

Although traditional gene mapping approaches (such as karyotyping [5], linkage analysis [6] homozygosity mapping [7] and copy number variation (CNV) analysis [8]) have led to great insights into Mendelian disease over the past few decades (Figure 1), they are unable to detect all forms of genomic variation (Table 1). The approach applied is dependent on whether the disease is, for example, caused by single nucleotide mutations or by CNVs, which is difficult to predict in advance. In addition, mapping approaches would often not reduce the number of candidate genes sufficiently for

straightforward follow-up by Sanger sequencing [9]. For example, genome-wide single nucleotide polymorphism analysis in a large Dutch pedigree with autosomal-dominant familial exudative vitreoretinopathy (FEVR, MIM 613310), a retinal disorder, identified a linkage peak of about 40 Mb on chromosome 7, containing more than 300 genes [10]. Even after adding linkage data from a second FEVR family the region was still too large for straightforward disease-gene identification, and Sanger sequencing of a few candidate genes did not identify causative mutations. Next generation sequencing (NGS) has the potential to identify all kinds of genetic variation at base-pair resolution throughout the human genome in a single experiment. This can be performed much faster and more cost efficiently than with traditional techniques (the sequencing of a genome by traditional techniques needed many years and cost millions of dollars, whereas NGS technology can sequence a genome for less than $7,000 and within a week [11]). This enables the detailed genomic analysis of large numbers of patients [12]. In the case of the two families with FEVR, we [10] used next generation sequencing to investigate the entire coding sequence of the 40-Mb region in a single affected individual from the first family and identified mutations in tetraspanin 12 (*TSPAN12*) to be the cause of FEVR in both families and in three additional families. For most Mendelian disorders, however, there is no disease locus known and an unbiased approach is required.

There are two unbiased sequencing approaches for detecting genetic variation within an individual: whole genome sequencing and whole exome sequencing. Whole genome sequencing is the ultimate approach for detecting all genomic variation in a patient's genome in a single experiment. However, current NGS instruments are limited in terms of throughput and cost efficiency. Therefore, this approach is limited to gene discovery projects in large genome sequencing centers and service companies. More cost-efficient sequencing strategies have been developed to study the approximately 1% of our genome that is protein-coding (the exome), by using various capturing approaches to enrich before NGS. Exome sequencing has rapidly become one of the main tools for studying the genetic causes of Mendelian disease

*Correspondence: c.gilissen@antrg.umcn.nl
Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands
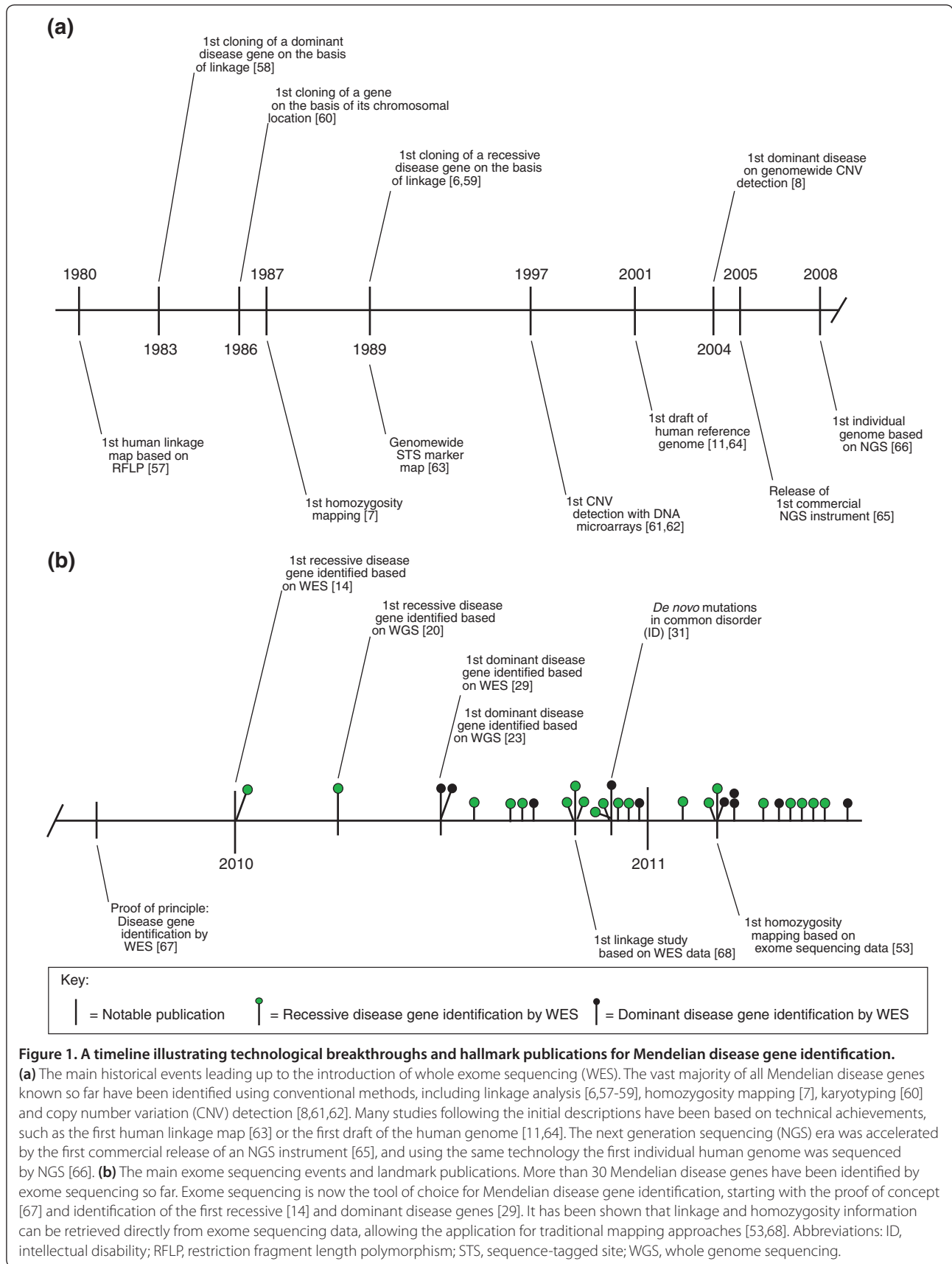
BioMed Central

Figure 1. A timeline illustrating technological breakthroughs and hallmark publications for Mendelian disease gene identification.
(a) The main historical events leading up to the introduction of whole exome sequencing (WES). The vast majority of all Mendelian disease genes known so far have been identified using conventional methods, including linkage analysis [6,57-59], homozygosity mapping [7], karyotyping [60] and copy number variation (CNV) detection [8,61,62]. Many studies following the initial descriptions have been based on technical achievements, such as the first human linkage map [63] or the first draft of the human genome [11,64]. The next generation sequencing (NGS) era was accelerated by the first commercial release of an NGS instrument [65], and using the same technology the first individual human genome was sequenced by NGS [66]. **(b)** The main exome sequencing events and landmark publications. More than 30 Mendelian disease genes have been identified by exome sequencing so far. Exome sequencing is now the tool of choice for Mendelian disease gene identification, starting with the proof of concept [67] and identification of the first recessive [14] and dominant disease genes [29]. It has been shown that linkage and homozygosity information can be retrieved directly from exome sequencing data, allowing the application for traditional mapping approaches [53,68]. Abbreviations: ID, intellectual disability; RFLP, restriction fragment length polymorphism; STS, sequence-tagged site; WGS, whole genome sequencing.

Table 1. Mendelian disease gene identification approaches

Approach	Applies to	Advantages	Disadvantages
Candidate gene	Any disease	Easy to perform for one or two genes; requires no mapping, can directly identify the causative variant/mutation	Relies heavily on current biological knowledge; success rate very low
Genetic mapping by karyotyping	Any disease	Easy to perform; no familial cases required; can detect (large) balanced events	Low resolution, only detects large chromosomal aberrations; mutation detection requires second step
Genetic mapping by linkage analysis	Inherited disease	Easy to perform	Requires large families, often identifies large loci; mutation detection requires second step
Genetic mapping by homozygosity mapping	Recessive monogenic diseases	Small families can be used	Most useful for consanguineous families; often identifies large loci; mutation detection requires second step
Genetic mapping by CNV analysis	Monogenic/ monolocus disease	High resolution CNV screening; no familial cases required; can potentially identify small loci	Only investigates CNVs; cannot detect balanced events, no base-pair resolution; mutation detection requires second step
Whole exome sequencing (WES)	Any disease	Base-pair resolution exome-wide; detects most types of genomic variation; can directly identify the causative variant/mutation	Unable to detect non-coding variants; limited resolution for CNVs and other structural variation; coverage variability due to enrichment process; relatively expensive
Whole genome sequencing (WGS)	Any disease	Base-pair resolution genome-wide; detects all types of genomic variation; can directly identify the causative variant/mutation	Data analysis complex; even more expensive than exome sequencing

[13] because academic groups with access to only one or two NGS systems can use this approach to study the exomes of hundreds of patients with Mendelian diseases per year and the bioinformatic challenges are modest when compared with whole genome sequencing. Although sequencing a complete genome may take only 1 week on a single machine, one can sequence more than 20 exomes in the same time. Since November 2009, exome sequencing has led to the identification of over 30 new genes in Mendelian diseases (Table 2) [14,15]. Although publication bias makes it difficult to assess the actual success rate of these approaches, from our own studies we estimate that whole exome sequencing identifies the major disease gene in at least 50% of the projects focused on rare but clinically well-defined Mendelian diseases (our unpublished data). The rapid expansion of Mendelian disease genes by exome sequencing is providing new insights because technological limitations have probably biased our current knowledge. Here, we discuss how our view of Mendelian disease is changing as a result of whole exome sequencing experiments and a limited number of whole genome sequencing approaches.

Lessons learned from exome sequencing

In the past 2 years we have seen many proof-of-concept studies using exome or genome studies to identify new disease genes for recessive and dominant disorders. These publications paint a mixed picture of phenotypes, genes and mutations underlying Mendelian disease (Figure 1, Table 2). There is a bias towards recessive disorders as their genetic causes are easier to identify than those that cause dominant disorders. This is because

genes carrying rare homozygous or compound heterozygous variants are not frequent in the unaffected population, and these can easily be prioritized for follow-up. In addition, past sample collection has mainly focused on familial cases with recessive inheritance.

Improving clinical diagnosis

From a review of the Mendelian diseases studied (Table 2), it is clear that not every whole exome sequencing experiment will result in the identification of a new disease gene. There are several examples in which the underlying genetic cause was not evident from the phenotype, yet whole exome sequencing revealed mutations in a known disease gene. For example, Choi *et al.* [16] identified mutations in the gene solute carrier family 26, member 3 (*SLC26A3*), encoding an epithelial Cl^-/HCO_3^- exchanger, in a case with the initial differential diagnosis of Bartter syndrome, a renal salt-wasting disease. After these mutations were identified the clinical diagnosis was re-evaluated and changed to congenital chloride diarrhea (*CLD*, MID 214700), a disease that was already known to result from mutations in this gene [17]. Similarly, Worthey *et al.* [18] reported a case in which a diagnosis of intractable inflammatory bowel disease (MIM 266600) was initially missed but was revealed after the identification of a missense mutation in *XIAP*, the X-linked inhibitor of apoptosis gene, by exome sequencing. Both these examples [16,18] show that unbiased whole exome sequencing can have an enormous impact on patient management by assisting clinicians in making the proper diagnosis, a phenomenon known as reverse phenotyping [19].

Mutations in genes that are known to cause disease will also be identified frequently when whole exome or

Table 2. Mendelian disease gene identifications by exome or genome sequencing

Disorder	Inheritance	Gene identified	Scope	References
Congenital chloride diarrhea	Recessive	*SLC26A3*	Exome	Choi *et al.* [16]
Miller syndrome	Recessive	*DHODH*	Exome	Ng *et al.* [14]
Charcot-Marie-Tooth neuropathy	Recessive	*SH3TC2*	Genome	Lupski *et al.* [20]
Metachondromatosis	Dominant	*PTPN11*	Genome	Sobreira *et al.* [23]
Schinzel-Giedion syndrome	Dominant	*SETBP1*	Exome	Hoischen *et al.* [29]
Nonsyndromic hearing loss	Recessive	*GPSM2*	Exome	Walsh *et al.* [69]
Perrault syndrome	Recessive	*HSD17B4*	Exome	Pierce *et al.* [25]
Hyperphosphatasia mental retardation syndrome	Recessive	*PIGV*	Exome	Krawitz *et al.* [68]
Sensenbrenner syndrome	Recessive	*WDR35*	Exome	Gilissen *et al.* [26]
Cerebral cortical malformations	Recessive	*WDR62*	Exome	Bilguvar *et al.* [70]
Kaposi sarcoma	Recessive	*STIM1*	Exome	Byun *et al.* [71]
Spinocerebellar ataxia	Dominant	*TGM6*	Exome	Wang *et al.* [72]
Combined hypolipidemia	Recessive	*ANGPTL3*	Exome	Musunuru *et al.* [40]
Complex I deficiency	Recessive	*ACAD9*	Exome	Haack *et al.* [52]
Autoimmune lymphoproliferative syndrome	Recessive	*FADD*	Exome	Bolze *et al.* [73]
Amyotrophic lateral sclerosis	Dominant	*VCP*	Exome	Johnson *et al.* [74]
Nonsyndromic mental retardation	Dominant	Various	Exome	Vissers *et al.* [31]
Kabuki syndrome	Dominant	*MLL2*	Exome	Ng *et al.* [30]
Inflammatory bowel disease	Dominant	*XIAP*	Exome	Worthey *et al.* [18]
Nonsyndromic mental retardation	Recessive	*TECR*	Exome	Caliskan *et al.* [75]
Retinitis pigmentosa	Recessive	*DHDDS*	Exome	Züchner *et al.* [56]
Osteogenesis imperfecta	Recessive	*SERPINF1*	Exome	Becker *et al.* [53]
Dilated cardiomyopathy	Dominant	*BAG3*	Exome	Norton *et al.* [24]
Hajdu-Cheney syndrome	Dominant	*NOTCH2*	Exome	Simpson *et al.* [76]
Hajdu-Cheney syndrome	Dominant	*NOTCH2*	Exome	Isidor *et al.* [77]
Skeletal dysplasia	Recessive	*POP1*	Exome	Glazov *et al.* [78]
Amelogenesis	Recessive	*FAM20A*	Exome	O'Sullivan *et al.* [80]
Chondrodysplasia and abnormal joint development	Recessive	*IMPAD1*	Exome	Vissers *et al.* [80]
Progeroid syndrome	Recessive	*BANF1*	Exome	Puente *et al.* [81]
Infantile mitochondrial cardiomyopathy	Recessive	*AARS2*	Exome	Götz *et al.* [82]
Sensory neuropathy with dementia and hearing loss	Dominant	*DNMT1*	Exome	Klein *et al.* [49]
Autism	Dominant	Various	Exome	O'Roak *et al.* [32]

genome sequencing is applied to genetically heterogeneous diseases that can be caused by monogenic mutations in many different genes. The technological limitations of Sanger sequencing often did not allow routine analysis of all known disease genes in patients with genetically heterogeneous disorders before whole exome approaches. The most prominent example of this was the identification of mutations in the SH3 domain and tetratricopeptide repeat domain 2 gene (*SH3TC2*, a gene known to cause neuropathy) as the cause of Charcot-Marie Tooth neuropathy (MIM 601596) in a family by whole genome sequencing [20]. An unbiased base-pair resolution approach can also reveal mutations in multiple genes that jointly explain a combination of two Mendelian phenotypes. For example, a study identified mutations in dihydroorotate dehydrogenase (*DHODH*) and dynein, axonemal, heavy chain 5 (*DNAH5*) in two siblings as the explanation of the combined phenotype of Miller syndrome (postaxial acrofacial dysostosis; MIM 263750) and primary ciliary dyskinesia, respectively [14]. Traditional mapping approaches would probably have missed the mutations in *DNAH5* as these were unique to this sibling and not present in other patients with Miller syndrome, which severely complicates mapping. Unbiased whole exome sequencing, on the other hand, identifies all variants and

allows for detailed analysis of individual cases and families. This is advantageous because the clinical spectrum of a disease can often be wider than previously appreciated, and multiple mutations might jointly explain a more complex phenotype.

Phenotypic heterogeneity

Allelic heterogeneity, in which a disease can be caused by mutations in different genes, has long been recognized to occur in Mendelian diseases. For example, in the case of Fanconi anemia (MIM 227650), a myelodysplastic disorder with chromosome breakage affecting all bone marrow elements and associated with cardiac, renal and limb malformations, mutations in many genes give rise to the exact same phenotype [21]. Conversely, it is also clear that different mutations in the same gene can result in completely different phenotypes, as is the case for tumor protein p63, in which different mutations can lead to several monogenic malformation syndromes [22]. This extreme form of phenotypic variability suggests completely different biological pathways are involved.

Unbiased whole exome and/or genome sequencing has identified further examples of unrelated phenotypes caused by different mutations in the same gene. Sobreira *et al.* [23] sequenced the complete genome of a single patient with metachondromatosis (MIM 156250), a skeletal dysplasia. They identified pathogenic loss-of-function mutations in protein tyrosine phosphatase, nonreceptor-type 11 (*PTPN11*) for which gain-of-function mutations are known to cause Noonan syndrome (MIM 163950), characterized by short stature and facial features, and pulmonary stenosis. Another example comes from Norton *et al.* [24], who identified both truncating and missense mutations in Bcl2-associated athanogene 3 (*BAG3*) in patients with dilated cardiomyopathy (MIM 115200), characterized by cardiac dilatation and reduced systolic function. A missense variant in this gene was previously known to cause myofibrillar myopathy (MIM 612954), a strikingly different phenotype characterized by skeletal muscle weakness associated with cardiac conduction blocks, arrhythmias, and restrictive heart failure. The authors [24] also observed phenotypic variability in a zebrafish model, where translation initiation blocking of the whole gene gave a single phenotype of heart failure. By contrast, a second morpholino oligonucleotide that splices out exon 2 resulted in axis curvature that might be analogous to skeletal myopathy. This observation led the authors [24] to suggest fundamental differences in mechanisms of disease. Similarly, Pierce *et al.* [25] observed mutations in *HSD174B*, which encodes 17β-hydroxysteroid dehydrogenase type 4, in two adult patients with Perrault syndrome, characterized by ovarian dysgenesis and deafness. Different classes of *HSD17B4* mutations have previously been associated with three different types of D-bifunctional protein (DBP) deficiency (MIM 261515), which is generally fatal within the first 2 years of life. The authors [25] suggested that the specific mutations they identified lead to a much milder phenotype that allows patients to survive to puberty and causes ovarian dysgenesis in females in addition to the known neurological defects associated with DBP. Using traditional approaches, these genes might not have been considered likely candidates for the disease.

In addition, for some novel disease genes the type of mutation may be specific to the disease in which these were observed. In two patients with Sensenbrenner syndrome, an autosomal recessive disorder characterized by ectodermal features, craniosynostosis and hypodontia, whole exome sequencing revealed a missense and a nonsense mutation in WD-repeat-containing protein 35 (*WDR35*) [26]. It may well be that two missense mutations result in a milder phenotype, whereas other mutations could result in a more severe phenotype, as confirmed by a subsequent study of patients with a lethal short rib polydactyly syndrome [27].

The role of *de novo* mutations in rare and common disorders

Many dominant Mendelian disorders occur sporadically because the severity and early onset of the disorder preclude transmission to subsequent generations. Consequently, there are no families available for genetic studies. Genetic variants associated with these diseases are under strong negative selection and will be rapidly eliminated from the genetic pool [28]. For these reasons, many genes that cause 'sporadic' disease remain to be identified. Important new insights have come from studying these sporadic forms of Mendelian disease by whole exome sequencing.

Exome sequencing first identified *de novo* mutations causing rare syndromic forms of dominant sporadic Mendelian disease, such as Schinzel-Giedion syndrome (MIM 269150) and Kabuki syndrome (MIM 147920) [29,30]. Both disorders are characterized by intellectual disability and typical facial features, and were anticipated to be largely caused by mutations in a single gene, which facilitated interpretation of exome sequencing data. Multiple unrelated patients with the same syndrome were sequenced and variant prioritization was focused on genes showing severe mutations in multiple if not all patients. For Schinzel-Giedion syndrome this resulted in the identification of heterozygous mutations in *SETBP1* (encoding the SET binding protein 1, a histone-lysine *N*-methyltransferase) in 12 out of 13 patients tested [29]. All mutations in *SETBP1* occurred *de novo* in the patient and were not detected in DNA from the unaffected parents. The disease mutations clustered in a genomic stretch of

just 11 nucleotides, affecting three of four consecutive amino acids. Interestingly, individuals with partial chromosome 18 deletions affecting *SETBP1* do not show clinical overlap with Schinzel-Giedion syndrome. Collectively, this indicates that these mutations are likely to confer a gain of function. Exome sequencing is particularly useful for identifying these types of mutations for which no other genome-wide approach is applicable. We suspect that gain-of-function mutations are largely underrepresented in the current databases, and that many more will be detected by exome sequencing.

After the detection of *de novo* mutations in rare Mendelian disorders, subsequent studies focused on their roles in common neurodevelopmental disorders, such as intellectual disability [31] and autism (MIM 209850) [32]. The high population frequency of these disorders has been hypothesized to reflect *de novo* mutations that compensate for allele loss due to severely reduced fecundity. If this were the case, the frequency of these disorders would reflect the size of the mutational target; disease caused by *de novo* mutations in a single gene will be very rare, whereas disorders caused by *de novo* mutations in many genes can occur at a high prevalence (Figure 2). In other words, the complexity of

many common diseases might be primarily due to genetic heterogeneity, with novel defects in different genes causing the same disease [33]. This would explain both the genetic and the clinical heterogeneity observed for mental illnesses and would explain why these disorders have a low recurrence risk [34]. To investigate this hypothesis, Vissers *et al.* [31] sequenced the exome of ten patients with intellectual disability and their healthy parents. After filtering out all inherited genomic variation, nine *de novo* mutations were identified in ten patients, between none and two per patient. Two *de novo* mutations were observed in genes previously linked to intellectual disability, and an analysis of the gene function and the mutation type indicated that an additional four *de novo* mutations in novel genes are likely to be pathogenic.

Similar results were recently reported by O'Roak *et al.* [32], who studied 20 patients with sporadic autism using the same approach [32]. Related work has also been reported for patients with schizophrenia (MIM 181500), although not yet using unbiased whole genome or exome approaches. In this case the authors [32] performed Sanger sequencing for 401 synapse-expressed genes in 143 patients with schizophrenia and identified eight *de novo* mutations, two of which were in SH3 and multiple
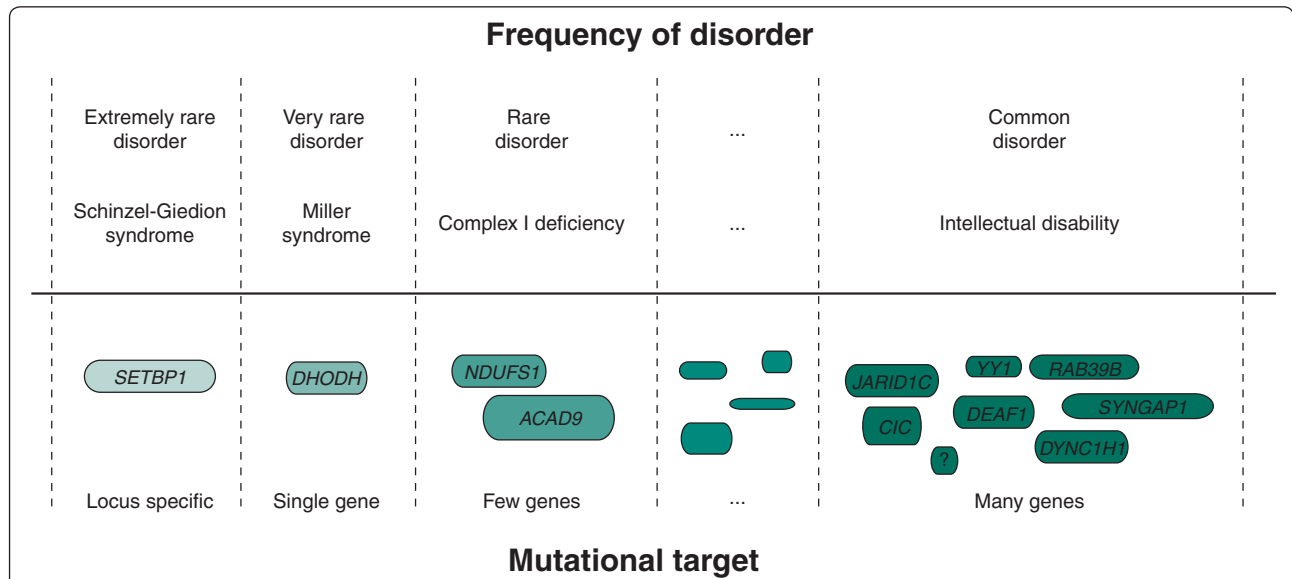
Figure 2. A representation of the relationship between the size of the mutational target and the frequency of disease for disorders caused by *de novo* mutations. Dashed lines separate different sizes of mutational target. Rounded rectangles represent examples of genes. Disease frequency categories range from extremely rare disorders (that is, only a few cases described) to disorders that occur more commonly within the population (such as intellectual disability, which has a frequency in the general population of more than 1%). Underneath each of these categories an example disorder is given. The lower part shows some of the implicated disease gene(s), ranging from a specific domain in a single gene, to single gene disorders, to multiple gene disorders, to disorders with extreme genetic heterogeneity. From left to right: SET binding protein 1 (SETBP1); dihydroorotate dehydrogenase (DHODH); NADH dehydrogenase (ubiquinone) Fe-S protein 1 (NDUFS1); acyl-CoA dehydrogenase family, member 9 (ACAD9); jumonji, AT rich interactive domain 1C (JARID1C); capicua homolog (CIC); deformed epidermal autoregulatory factor 1 (DEAF1); YY1 transcription factor (YY1); dynein, cytoplasmic 1, heavy chain 1 (DYNC1H1); member RAS oncogene family (RAB39B); synaptic Ras GTPase activating protein 1 (SYNGAP1).

ankyrin repeat domains 3 (*SHANK3*), a gene known to cause schizophrenia [35,36]. If confirmed in larger follow-up studies, this would indicate that a significant proportion of these common disorders is caused by rare *de novo* mutations. This would provide strong support for the hypothesis that rare variants substantially contribute to common diseases and would also represent a change of focus in Mendelian genetic disease research, where the emphasis has been on studying familial forms of disease.

Severe early onset disorders, such as intellectual disability, autism and childhood schizophrenia, are clearly the first group of disorders on which to test this *de novo* mutation hypothesis, as inherited mutations are less likely to have a major role in these disorders because of the reduced fitness of those affected by the disease [37]. The contribution of *de novo* mutations to adult-onset diseases with less impact on fertility is expected to be lower. However, we note that this is mainly dependent on the size of the mutational target and remains to be determined. A problem in this respect, however, is the availability of surviving parents of offspring who have a late-onset disease in order to prove *de novo* occurrence [38]. Related to this, Conrad *et al.* [39] recently published the first genome sequencing study in two healthy trios of offspring with both their parents and validated 49 *de novo* germline mutations in one offspring and 35 in the other. This study [39] also highlighted the fact that cell lines are not the preferred material for this analysis, as 952 and 643 non-germline *de novo* mutations were observed in both offspring that are likely to have been caused by cell line creation and culturing. Given the apparent role of rare *de novo* mutations in disease, it is evident that we need to learn much more about the general occurrence of these mutations in our population and perform detailed comparisons of *de novo* mutations in patients versus controls.

Finding the Mendelian contribution to common traits

Although some studies support the arguments that a significant proportion of common disorders (such as intellectual disability, autism and schizophrenia) represent an accumulation of rare disorders, true multigenic/complex diseases can also benefit from exome sequencing. Until NGS studies of the size of genome-wide association studies are broadly applicable and affordable (that is, until one can run and analyze 1,000 exomes routinely), insights can be obtained by studying small cohorts from the extreme ends of the phenotypic spectrum of common traits, because the Mendelian forms of these traits are expected to be overrepresented in this group. The first example of this approach was provided by whole exome sequencing in patients with extremely low low-density lipoprotein (LDL) cholesterol levels, which identified mutations in *ANGPTL3* [40]. The identified gene is secreted and expressed primarily in the liver and encodes the angiopoietin-like 3 protein. *ANGPTL3* has a role in lipoprotein lipase and endothelial lipase inhibition, thereby increasing plasma triglyceride and high-density lipoprotein (HDL) cholesterol levels. The very low LDL levels in these individuals could therefore be explained by two mutated alleles for this gene.

From genes to pathways

Although only a few exome studies have been conducted so far (Table 2), these have already provided new insights into human gene networks. For example, it was known through classical disease gene identification that histone modifiers have an important role in human developmental diseases. Haploinsufficiency of histone methyltransferases, such as *NSD1* (encoding nuclear receptor binding SET domain protein 1), *NSD2* (encoding nuclear SET domain-containing protein 2 and also called Wolf-Hirschhorn syndrome candidate 1, *WHSC1*) and *EHMT1* (encoding euchromatic histone-lysine *N*-methyltransferase 1), cause several congenital diseases [41]. Exome sequencing studies have confirmed the importance of this class of genes. Among the few exome studies that identified dominant genes for Mendelian disorders, three disease genes function in histone modification: (i) mutations in the histone methyltransferase gene *MLL2* (H3K4me) have been shown to be the cause of Kabuki syndrome [30]; (ii) point mutations in *SETBP1* – encoding the SET binding protein 1, a histone-lysine *N*-methyltransferase – have been identified as the cause of Schinzel-Giedion syndrome [29]; (iii) most recently, *ASXL1*, encoding the protein Additional sex-combs-like 1, was implicated in Bohring-Opitz syndrome (MIM 605039), characterized by severe malformations and intellectual disability [42]. ASXL1 is an interactor of lysine-specific demethylase 1 (LSD1) and therefore *ASXL1* can be considered another histone modification gene [43].

Histone modifiers might have a dual role in disease. Although germline mutations in these genes can lead to developmental disorders, somatic mutations have been reported in leukemias and other malignancies. Translocations of *NSD1* to *NUP98* (encoding nucleoporin 98 kDa) and of *NSD2/WHSC1* to *IgH* occur in some hematologic malignancies [41]. In addition, *SETBP1* translocations with *NUP98* have been described in pediatric acute T-cell lymphoblastic leukemia [44], and somatic mutations in *ASXL1* occur in several forms of leukemia [45,46]. Further credence for this dual role for genes involved in histone modification has been given by recent studies. It was already anticipated that the DNMT3A, DNMT3B and DNMT3L proteins were primarily responsible for the establishment of genomic DNA methylation patterns and should therefore have an

important role in human developmental, reproductive and mental health [47]. Exome sequencing identified *DNMT3A* (encoding DNA methyltransferase 3A) mutations in monocytic leukemia, confirming the suspected link with cancer development [48]. Exome sequencing also led to the identification of mutations in *DNMT1* that cause both central and peripheral neurodegeneration in a form of hereditary sensory and autonomic neuropathy with dementia and hearing loss [49]. These exome sequencing findings add to a growing list of genes for which somatic mutations have been identified in malignancies and germline mutations in developmental disorders.

The importance of evolutionary conservation

The interpretation of missense variants in Mendelian diseases is challenging. A common method to address pathogenicity is to assume that purifying selection 'constrains' evolutionary divergence at phenotypically important nucleotides and amino acids [50]. Following an earlier suggestion [51], exome studies have now confirmed that pathogenic missense variants indeed tend to affect highly conserved nucleotides (using GERP, PhyloP or PhastCons scores), or amino acids (by multiple sequence alignment). All scores in some way measure the difference between the number of nucleotide substitutions that have occurred at a site during evolution and the number of substitutions that are expected from neutral evolution. As an example, the nucleotides affected by *de novo* mutations in *SETBP1* that cause Schinzel-Giedion syndrome are among the most highly conserved bases in the genome, according to PhyloP vertebrate conservation scores [29]. This observation is unlikely to be explained by an ascertainment bias, as evolutionary conservation is generally used only to strengthen the existing evidence of pathogenicity, rather than to prioritize variants for follow-up.

Towards therapy for Mendelian disease

Exome studies have revealed mutations causing Mendelian disease in more than 30 genes, many of which had no previously known function. This knowledge can help to identify essential biological pathways disrupted in these often severe disorders. For example, Haack *et al.* [52] obtained a promising clinical response to a multivitamin scheme including daily riboflavin treatment in a patient with a complex 1 deficiency (MIM 252010) after whole exome sequencing revealed mutations in *ACAD9*, a member of the mitochondrial acyl-CoA dehydrogenase protein family. Another example involves *SERPINF1*, in which mutations were detected by whole exome sequencing in patients with osteogenesis imperfecta [53]. There is much to learn from the first gene therapy trials with pigment epithelium-derived

factor (PEDF, encoded by *SERPINF1)*. PEDF is thought to counteract the effect of vascular endothelial growth factor (VEGF) – a signal protein produced by cells that stimulates vasculogenesis and angiogenesis – and trials are dedicated to the treatment of the wet form of age-related macular degeneration [54]. In a mouse model of ischemia-induced retinal angiogenesis, PEDF eliminated aberrant neovascularization [55], which suggests that PEDF might have potential in treating osteogenesis imperfecta. Finally, Züchner *et al.* [56] used whole exome sequencing to identify mutations in dehydrodolichyl diphosphate synthase (*DDHDS*) in patients with retinitis pigmentosa, linking this disease to *N*-linked glycosylation pathways, suggesting new possibilities for therapeutic interventions. Although the number of these promising examples is small, one might foresee that several clinical conditions will be better understood after the identification of the underlying disease gene, and in some cases this will enable usable therapy.

Conclusions

Exome sequencing and, in a few cases, genome sequencing, has significantly progressed the field of Mendelian disease in the past 2 years. The unbiased nature of these approaches is providing significant insights into the genetic causes of Mendelian disease in general, and of sporadic disease in particular, by revealing rare *de novo* mutations as a common cause of disease. This approach is crucial for drawing accurate genotype-phenotype correlations and will undoubtedly improve diagnosis for the millions of individuals with Mendelian disease, improve family counseling and reveal new therapeutic targets. The next challenge in disease research will be to systematically study the role of variation in the non-coding part of our genome in health and disease. The study of Mendelian diseases will be crucial in this endeavor, as they offer the advantage of high penetrant mutations, the ability to perform family studies and look for segregation of variation with disease, and the possibility of finding recurrent mutations in unrelated patients with similar phenotypes.

References
1. Antonarakis SE, Beckmann JS: Mendelian disorders deserve more attention. *Nat Rev Genet* 2006, **7**:277-282.
2. Oti M, Brunner HG: The modular nature of genetic diseases. *Clin Genet* 2007, **71**:1-11.
3. Peltonen L, Perola M, Naukkarinen J, Palotie A: Lessons from studying monogenic disease for common disease. *Hum Mol Genet* 2006, **15** Spec No 1:R67-R74.

4. Shendure J, Ji H: **Next-generation DNA sequencing.** *Nat Biotechnol* 2008, **26:**1135-1145.

5. Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Hasegawa T, Chinen Y, Tomita Ha HA, Kinoshita A, Mizuguchi T, Yoshiura Ki K, Ohta T, Kishino T, Fukushima Y, Niikawa N, Matsumoto N: **Haploinsufficiency of NSD1 causes Sotos syndrome.** *Nat Genet* 2002, **30:**365-366.

6. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC: **Identification of the cystic fibrosis gene: genetic analysis.** *Science* 1989, **245:**1073-1080.

7. Lander ES, Botstein D: **Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children.** *Science* 1987, **236:**1567-1570.

8. Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van dV, Huys EH, de Jong PJ, Hamel BC, Schoenmakers EF, Brunner HG, Veltman JA, van Kessel AG: **Mutations in a new member of the chromodomain gene family cause CHARGE syndrome.** *Nat Genet* 2004, **36:**955-957.

9. Botstein D, Risch N: **Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease.** *Nat Genet* 2003, **33 Suppl:**228-237.

10. Nikopoulos K, Gilissen C, Hoischen A, van Nouhuys CE, Boonstra FN, Blokland EA, Arts P, Wieskamp N, Strom TM, Ayuso C, Tilanus MA, Bouwhuis S, Mukhopadhyay A, Scheffer H, Hoefsloot LH, Veltman JA, Cremers FP, Collin RW: **Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy.** *Am J Hum Genet* 2010, **86:**240-247.

11. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, *et al.*: **The sequence of the human genome.** *Science* 2001, **291:**1304-1351.

12. Mardis ER: **A decade's perspective on DNA sequencing technology.** *Nature* 2011, **470:**198-203.

13. Teer JK, Mullikin JC: **Exome sequencing: the sweet spot before whole genomes.** *Hum Mol Genet* 2010, **19:**R145-R151.

14. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ: **Exome sequencing identifies the cause of a Mendelian disorder.** *Nat Genet* 2010, **42:**30-35.

15. Ng SB, Nickerson DA, Bamshad MJ, Shendure J: **Massively parallel sequencing and rare disease.** *Hum Mol Genet* 2010, **19:**119.

16. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP: **Genetic diagnosis by whole exome capture and massively parallel DNA sequencing.** *Proc Natl Acad Sci U S A* 2009, **106:**19096-19101.

17. Hoglund P, Haila S, Socha J, Tomaszewski L, Saarialho-Kere U, Karjalainen-Lindsberg ML, Airola K, Holmberg C, de la CA, Kere J: **Mutations of the Down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea.** *Nat Genet* 1996, **14:**316-319.

18. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP: **Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease.** *Genet Med* 2011, **13:**255-262.

19. Schulze TG, McMahon FJ: **Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping.** *Hum Hered* 2004, **58:**131-138.

20. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio DD, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA: **Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy.** *N Engl J Med* 2010, **362:**1181-1191.

21. Neveling K, Endt D, Hoehn H, Schindler D: **Genotype-phenotype correlations in Fanconi anemia.** *Mutat Res* 2009, **668:**73-91.

22. Brunner HG, Hamel BC, Bokhoven HH: **P63 gene mutations and human developmental syndromes.** *Am J Med Genet* 2002, **112:**284-290.

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

24. Norton N, Li D, Rieder MJ, Siegfried JD, Rampersaud E, Züchner S, Mangos S, Gonzalez-Quintana J, Wang L, McGee S, Reiser J, Martin E, Nickerson DA, Hershberger RE: **Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy.** *Am J Hum Genet* 2011, **88:**273-282.

25. Pierce SB, Walsh T, Chisholm KM, Lee MK, Thornton AM, Fiumara A, Opitz JM, Levy-Lahad E, Klevit RE, King MC: **Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault syndrome.** *Am J Hum Genet* 2010, **87:**282-288.

26. Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M, van Reeuwijk J, Kant SG, Roepman R, Knoers NV, Veltman JA, Brunner HG: **Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome.** *Am J Hum Genet* 2010, **87:**418-423.

27. Mill P, Lockhart PJ, Fitzpatrick E, Mountford HS, Hall EA, Reijns MA, Keighren M, Bahlo M, Bromhead CJ, Budd P, Aftimos S, Delatycki MB, Savarirayan R, Jackson IJ, Amor DJ: **Human and mouse mutations in WDR35 cause short-rib polydactyly syndromes due to abnormal ciliogenesis.** *Am J Hum Genet* 2011, **88:**508-515.

28. Eyre-Walker A, Keightley PD: **The distribution of fitness effects of new mutations.** *Nat Rev Genet* 2007, **8:**610-618.

29. Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BB, Veltman JA: ***De novo* mutations of SETBP1 cause Schinzel-Giedion syndrome.** *Nat Genet* 2010, **42:**483-485.

30. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J: **Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.** *Nat Genet* 2010, **42:**790-793.

31. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA: **A *de novo* paradigm for mental retardation.** *Nat Genet* 2010, **42:**1109-1112.

32. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE: **Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations.** *Nat Genet* 2011, **43:**585-589.

33. McClellan J, King MC: **Genetic heterogeneity in human disease.** *Cell* 2010, **141:**210-217.

34. Crow YJ, Tolmie JL: **Recurrence risks in mental retardation.** *J Med Genet* 1998, **35:**177-182.

35. Gauthier J, Champagne N, Lafrenière RG, Xiong L, Spiegelman D, Brustein E, Lapointe M, Peng H, Côté M, Noreau A, Hamdan FF, Addington AM, Rapoport JL, Delisi LE, Krebs MO, Joober R, Fathalli F, Mouaffak F, Haghighi AP, Néri C, Dubé MP, Samuels ME, Marineau C, Stone EA, Awadalla P, Barker PA, Carbonetto S, Drapeau P, Rouleau GA; S2D Team: ***De novo* mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia.** *Proc Natl Acad Sci U S A* 2010, **107:**7863-7868.

36. Awadalla P, Gauthier J, Myers RA, Casals F, Hamdan FF, Griffing AR, Côté M, Henrion E, Spiegelman D, Tarabeux J, Piton A, Yang Y, Boyko A, Bustamante C, Xiong L, Rapoport JL, Addington AM, DeLisi JL, Krebs MO, Joober R, Millet B, Fombonne E, Mottron L, Zilversmit M, Keebler J, Daoud H, Marineau C, Roy-Gagnon MH, Dubé MP, *et al.*: **Direct measure of the *de novo* mutation rate in autism and schizophrenia cohorts.** *Am J Hum Genet* 2010, **87:**316-324.

37. Julie G, Hamdan FF, Rouleau GA: **A strategy to identify *de novo* mutations in common disorders such as autism and schizophrenia.** *J Vis Exp* 2011, doi: 10.3791/2534.

38. Pamphlett R, Morahan JM, Yu B: **Using case-parent trios to look for rare *de novo* genetic variants in adult-onset neurodegenerative diseases.** *J Neurosci Methods* 2011, **197:**297-301.

39. Conrad DF, Keebler JE, Depristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, Zilversmit M, Cartwright R, Rouleau GA, Daly M, Stone EA, Hurles ME, Awadalla P: **Variation in genome-wide mutation rates within and between human families.** *Nat Genet* 2011, **43:**712-714.

40. Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ, Fennell T, Banks E, Ambrogio L, Cibulskis K, Kernytsky A, Gonzalez E, Rudzicz N, Engert JC, DePristo MA, Daly

MJ, Cohen JC, Hobbs HH, Altshuler D, Schonfeld G, Gabriel SB, Yue P, Kathiresan S: **Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia.** *N Engl J Med* 2010, **363:**2220-2227.

41. Nimura K, Ura K, Kaneda Y: **Histone methyltransferases: regulation of transcription and contribution to human disease.** *J Mol Med* 2010, **88:**1213-1220.

42. Hoischen A, van Bon BW, Rodríguez-Santiago B, Gilissen C, Vissers LE, de Vries P, Janssen I, van Lier B, Hastings R, Smithson SF, Newbury-Ecob R, Kjaergaard S, Goodship J, McGowan R, Bartholdi D, Rauch A, Peippo M, Cobben JM, Wieczorek D, Gillessen-Kaesbach G, Veltman JA, Brunner HG, de Vries BB: *De novo* **nonsense mutations in ASXL1 cause Bohring-Opitz syndrome.** *Nat Genet* 2011, doi: 10.1038/ng.868.

43. Lee SW, Cho YS, Na JM, Park UH, Kang M, Kim EJ, Um SJ: **ASXL1 represses retinoic acid receptor-mediated transcription through associating with HP1 and LSD1.** *J Biol Chem* 2010, **285:**18-29.

44. Panagopoulos I, Kerndrup G, Carlsen N, Strombeck B, Isaksson M, Johansson B: **Fusion of NUP98 and the SET binding protein 1 (SETBP1) gene in a paediatric acute T cell lymphoblastic leukaemia with t(11;18)(p15;q12).** *Br J Haematol* 2007, **136:**294-296.

45. Chou WC, Huang HH, Hou HA, Chen CY, Tang JL, Yao M, Tsay W, Ko BS, Wu SJ, Huang SY, Hsu SC, Chen YC, Huang YN, Chang YC, Lee FY, Liu MC, Liu CW, Tseng MH, Huang CF, Tien HF: **Distinct clinical and biological features of *de novo* acute myeloid leukemia with additional sex comb-like 1 (ASXL1) mutations.** *Blood* 2010, **116:**4086-4094.

46. Gelsi-Boyer V, Trouplin V, Roquain J, Adélaíde J, Carbuccia N, Esterni B, Finetti P, Murati A, Arnoulet C, Zerazhi H, Fezoui H, Tadrist Z, Nezri M, Chaffanet M, Mozziconacci MJ, Vey N, Birnbaum D: **ASXL1 mutation is associated with poor prognosis and acute transformation in chronic myelomonocytic leukaemia.** *Br J Haematol* 2010, **151:**365-375.

47. Chedin F: **The DNMT3 family of mammalian *de novo* DNA methyltransferases.** *Prog Mol Biol Transl Sci* 2011, **101:**255-285.

48. Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ: **Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia.** *Nat Genet* 2011, **43:**309-315.

49. Klein CJ, Botuyan MV, Wu Y, Ward CJ, Nicholson GA, Hammans S, Hojo K, Yamanishi H, Karpf AR, Wallace DC, Simon M, Lander C, Boardman LA, Cunningham JM, Smith GE, Litchy WJ, Boes B, Atkinson EJ, Middha S, B Dyck PJ, Parisi JE, Mer G, Smith DI, Dyck PJ: **Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss.** *Nat Genet* 2011, **43:**595-600.

50. Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, Shendure J, Nickerson DA: **Single-nucleotide evolutionary constraint scores highlight disease-causing mutations.** *Nat Methods* 2010, **7:**250-251.

51. Sidow A: **Sequence first. Ask questions later.** *Cell* 2002, **111:**13-16.

52. Haack TB, Danhauser K, Haberberger B, Hoser J, Strecker V, Boehm D, Uziel G, Lamantea E, Invernizzi F, Poulton J, Rolinski B, Iuso A, Biskup S, Schmidt T, Mewes HW, Wittig I, Meitinger T, Zeviani M, Prokisch H: **Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency.** *Nat Genet* 2010, **42:**1131-1134.

53. Becker J, Semler O, Gilissen C, Li Y, Bolz HJ, Giunta C, Bergmann C, Rohrbach M, Koerber F, Zimmermann K, de Vries P, Wirth B, Schoenau E, Wollnik B, Veltman JA, Hoischen A, Netzer C: **Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta.** *Am J Hum Genet* 2011, **88:**362-371.

54. Campochiaro PA, Nguyen QD, Shah SM, Klein ML, Holz E, Frank RN, Saperstein DA, Gupta A, Stout JT, Macko J, DiBartolomeo R, Wei LL: **Adenoviral vector-delivered pigment epithelium-derived factor for neovascular age-related macular degeneration: results of a phase I clinical trial.** *Hum Gene Ther* 2006, **17:**167-176.

55. Stellmach V, Crawford SE, Zhou W, Bouck N: **Prevention of ischemia-induced retinopathy by the natural ocular antiangiogenic agent pigment epithelium-derived factor.** *Proc Natl Acad Sci U S A* 2001, **98:**2593-2597.

56. Züchner S, Dallman J, Wen R, Beecham G, Naj A, Farooq A, Kohli MA, Whitehead PL, Hulme W, Konidari I, Edwards YJ, Cai G, Peter I, Seo D, Buxbaum JD, Haines JL, Blanton S, Young J, Alfonso E, Vance JM, Lam BL, Peričak-Vance MA: **Whole-Exome sequencing links a variant in DHDDS to retinitis pigmentosa.** *Am J Hum Genet* 2011, **88:**201-206.

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

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

59. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL: **Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA.** *Science* 1989, **245:**1066-1073.

60. Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, Cole FS, Curnutte JT, Orkin SH: **Cloning the gene for an inherited human disorder--chronic granulomatous disease--on the basis of its chromosomal location.** *Nature* 1986, **322:**32-38.

61. Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P: **Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances.** *Genes Chromosomes Cancer* 1997, **20:**399-407.

62. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG: **High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.** *Nat Genet* 1998, **20:**207-211.

63. Olson M, Hood L, Cantor C, Botstein D: **A common language for physical mapping of the human genome.** *Science* 1989, **245:**1434-1435.

64. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.

65. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, *et al.*: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437:**376-380.

66. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM: **The complete genome of an individual by massively parallel DNA sequencing.** *Nature* 2008, **452:**872-876.

67. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J: **Targeted capture and massively parallel sequencing of 12 human exomes.** *Nature* 2009, **461:**272-276.

68. Krawitz PM, Schweiger MR, Rödelsperger C, Marcelis C, Kölsch U, Meisel C, Stephani F, Kinoshita T, Murakami Y, Bauer S, Isau M, Fischer A, Dahl A, Kerick M, Hecht J, Köhler S, Jäger M, Grünhagen J, de Condor BJ, Doelken S, Brunner HG, Meinecke P, Passarge E, Thompson MD, Cole DE, Horn D, Roscioli T, Mundlos S, Robinson PN: **Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome.** *Nat Genet* 2010, **42:**827-829.

69. Walsh T, Shahin H, Elkan-Miller T, Lee MK, Thornton AM, Roeb W, Abu RA, Loulus S, Avraham KB, King MC, Kanaan M: **Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82.** *Am J Hum Genet* 2010, **87:**90-94.

70. Bilgüvar K, Oztürk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalnizoğlu D, Tüysüz B, Cağlayan AO, Gökben S, Kaymakçalan H, Barak T, Bakircioğlu M, Yasuno K, Ho W, Sanders S, Zhu Y, Yilmaz S, Dinçer A, Johnson MH, Bronen RA, Koçer N, Per H, Mane S, Pamir MN, Yalçinkaya C, Kumandaş S, Topçu M, Ozmen M, Sestan N, *et al.*: **Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations.** *Nature* 2010, **467:**207-210.

71. Byun M, Abhyankar A, Lelarge V, Plancoulaine S, Palanduz A, Telhan L, Boisson B, Picard C, Dewell S, Zhao C, Jouanguy E, Feske S, Abel L, Casanova JL: **Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma.** *J Exp Med* 2010, **207:**2307-2312.

72. Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, Jiang H, Zhang P, Shen L, Guo JF, Li N, Li YR, Lei LF, Zhou J, Du J, Zhou YF, Pan Q, Wang J, Wang J, Li RQ, Tang BS: **TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing.** *Brain* 2010, **133:**3510-3518.

73. Bolze A, Byun M, McDonald D, Morgan NV, Abhyankar A, Premkumar L, Puel A, Bacon CM, Rieux-Laucat F, Pang K, Britland A, Abel L, Cant A, Maher ER, Riedl SJ, Hambleton S, Casanova JL: **Whole-exome-sequencing-based discovery of human FADD deficiency.** *Am J Hum Genet* 2010, **87:**873-881.

74. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, Gibbs JR, Brunetti M, Gronka S, Wuu J, Ding J, McCluskey L, Martinez-Lage M, Falcone D, Hernandez DG, Arepalli S, Chong S, Schymick JC, Rothstein J, Landi F, Wang YD, Calvo A, Mora G, Sabatelli M, Monsurrò MR, Battistini S, Salvi F, Spataro R, Sola P, Borghero G; ITALSGEN Consortium, *et al*: **Exome sequencing reveals VCP mutations as a cause of familial ALS.** *Neuron* 2010, **68**:857-864.

75. Çalışkan M, Chong JX, Uricchio L, Anderson R, Chen P, Sougnez C, Garimella K, Gabriel SB, DePristo MA, Shakir K, Matern D, Das S, Waggoner D, Nicolae DL, Ober C: **Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13.** *Hum Mol Genet* 2011, **20**:1285-1289.

76. Simpson MA, Irving MD, Asilmaz E, Gray MJ, Dafou D, Elmslie FV, Mansour S, Holder SE, Brain CE, Burton BK, Kim KH, Pauli RM, Aftimos S, Stewart H, Kim CA, Holder-Espinasse M, Robertson SP, Drake WM, Trembath RC: **Mutations in NOTCH2 cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss.** *Nat Genet* 2011, **43**:303-305.

77. Isidor B, Lindenbaum P, Pichon O, Bézieau S, Dina C, Jacquemont S, Martin-Coignard D, Thauvin-Robinet C, Le MM, Mandel JL, David A, Faivre L, Cormier-Daire V, Redon R, Le Caignec C: **Truncating mutations in the last exon of NOTCH2 cause a rare skeletal disorder with osteoporosis.** *Nat Genet* 2011, **43**:306-308.

78. Glazov EA, Zankl A, Donskoi M, Kenna TJ, Thomas GP, Clark GR, Duncan EL, Brown MA: **Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia.** *PLoS Genet* 2011, **7**:e1002027.

79. O'Sullivan J, Bitu CC, Daly SB, Urquhart JE, Barron MJ, Bhaskar SS, Martelli-Júnior H, dos Santos Neto PE, Mansilla MA, Murray JC, Coletta RD, Black GC, Dixon MJ: **Whole-exome sequencing identifies FAM20A mutations as a cause of amelogenesis imperfecta and gingival hyperplasia syndrome.** *Am J Hum Genet* 2011, **88**:616-620.

80. Vissers LE, Lausch E, Unger S, Campos-Xavier AB, Gilissen C, Rossi A, Del Rosario M, Venselaar H, Knoll U, Nampoothiri S, Nair M, Spranger J, Brunner HG, Bonafé L, Veltman JA, Zabel B, Superti-Furga A: **Chondrodysplasia and abnormal joint development associated with mutations in IMPAD1, Encoding the Golgi-resident nucleotide phosphatase, gPAPP.** *Am J Hum Genet* 2011, **88**:608-615.

81. Puente XS, Quesada V, Osorio FG, Cabanillas R, Cadiñanos J, Fraile JM, Ordóñez GR, Puente DA, Gutiérrez-Fernández A, Fanjul-Fernández M, Lévy N, Freije JM, López-Otín C: **Exome sequencing and functional analysis identifies BANF1 mutation as the cause of a hereditary progeroid syndrome.** *Am J Hum Genet* 2011, **88**:650-656.

82. Götz A, Tyynismaa H, Euro L, Ellonen P, Hyötyläinen T, Ojala T, Hämäläinen RH, Tommiska J, Raivio T, Oresic M, Karikoski R, Tammela O, Simola KO, Paetau A, Tyni T, Suomalainen A: **Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy.** *Am J Hum Genet* 2011, **88**:635-642.