TRANSLATIONAL GENETICS

Rare-disease genetics in the era of next-generation sequencing: discovery to translation

Kym M. Boycott, Megan R. Vanstone, Dennis E. Bulman and Alex E. MacKenzie

Abstract | Work over the past 25 years has resulted in the identification of genes responsible for ~50% of the estimated 7,000 rare monogenic diseases, and it is predicted that most of the remaining disease-causing genes will be identified by the year 2020, and probably sooner. This marked acceleration is the result of dramatic improvements in DNA-sequencing technologies and the associated analyses. We examine the rapid maturation of rare-disease genetic analysis and successful strategies for gene identification. We highlight the impact of discovering rare-disease-causing genes, from clinical diagnostics to insights gained into biological mechanisms and common diseases. Last, we explore the increasing therapeutic opportunities and challenges that the resulting expansion of the 'atlas' of human genetic pathology will bring.

Orphan drugs
Pharmaceutical agents developed for the treatment of a rare disease (often referred to as an orphan disease). The assignment of 'orphan' status is a matter of public policy and is possible in only some countries.

Next-generation sequencing
(NGS). Highly parallel DNA-sequencing technologies that produce many hundreds of thousands or millions of short reads (25–500 bp) for a low cost and in a short time.

Department of Pediatrics, University of Ottawa and Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada.
Correspondence to: K.M.B.
e-mail: kboycott@cheo.on.ca
doi:10.1038/nrg3555
Published online
3 September 2013

Rare diseases caused by altered functions of single genes can be chronically debilitating and life-limiting. Notwithstanding their severity, some rare diseases are compatible with a good quality of life if they are diagnosed early and optimally managed. Although the individual diseases are rare (defined as affecting fewer than 200,000 people in the United States or fewer than 1 in 2,000 people in Europe), they are collectively common, affecting millions of individuals worldwide[1,2]. Unfortunately, effective therapies for these diseases are themselves comparatively rare. Thus, in addition to the effects on patients and their families, these diseases have a tremendous cost for health care systems and societies[3–5].

The number of rare genetic diseases is difficult to gauge precisely. An interrogation of Online Mendelian Inheritance in Man (OMIM)[6], a catalogue of human genes and associated genetic diseases, and Orphanet, a comprehensive reference portal for rare diseases (including an inventory of such diseases and orphan drugs[7]), results in a best estimate of between 6,000 and 7,000 rare genetic diseases. However, the number of phenotypes that remain to be defined may be considerably higher[8]. Predictions regarding the total number of rare-disease-causing genes, based on estimates of the rate of human genome mutation and the number of essential genes, have resulted in estimates of 7,000 to 15,000 (REF. 9). The molecular aetiology of >3,500 rare diseases has been determined. This has been primarily accomplished through linkage mapping and candidate gene analysis, a labour- and resource-intensive process that often takes as long as a decade and costs hundreds of thousands of US dollars. Many rare genetic diseases have been refractory to traditional gene discovery approaches for several reasons: locus heterogeneity, the availability of only a small number of patients or families to study and substantially reduced reproductive fitness as a result of such diseases.

The advent of next-generation sequencing (NGS) surmounts these issues and has changed the landscape of rare-genetic-disease research, with causative genes being identified at an accelerating rate (FIG. 1). In this Review, we discuss the current status of NGS-based gene discovery for rare diseases and look ahead to the future prospects for this field of research. We examine successful strategies for gene identification and the incorporation of NGS into clinical practice for patients with rare diseases, who — in addition to patients with cancer — represent the first health care beneficiaries of this revolutionary technology. We highlight how the discovery of rare-disease-causing genes has provided insights into the biological mechanisms underlying rare diseases, and into complex and common diseases. We also explore the increasing therapeutic opportunities and challenges, and the need for low-cost generalizable approaches to translational research for rare diseases. We conclude by highlighting the crucial need for

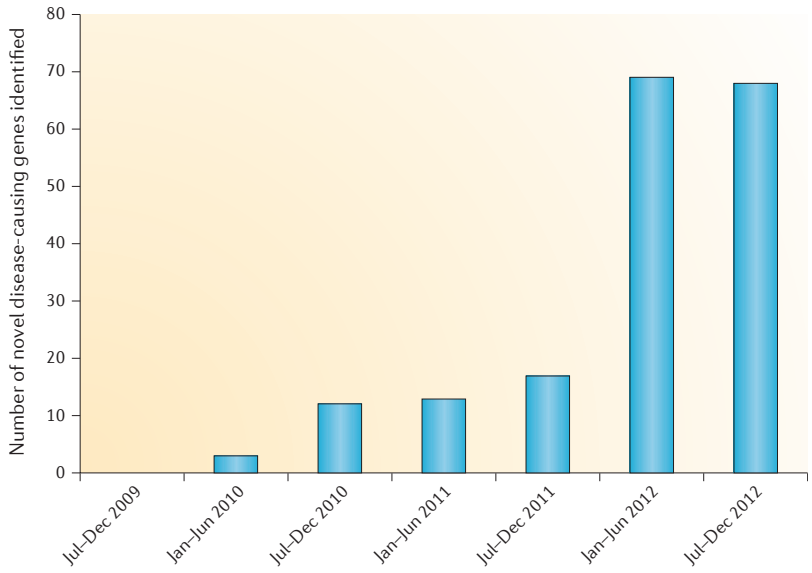

Figure 1 | **Pace of discovery of novel rare-disease-causing genes using whole-exome sequencing.** Since the first whole-exome sequencing (WES) proof-of-concept experiment[11], the discovery of disease-causing genes using WES has increased rapidly, with a marked jump from 2011 to 2012 and a stabilization of this rate in the latter half of 2012. More than 180 novel genes have been discovered in this manner. As next-generation sequencing technology becomes less costly and more widely used, we anticipate another jump in the rate of discovery, provided that the infrastructure for the large-scale sharing of deep phenotypic and genetic data sets emerges. Data were collected from PubMed until the end of 2012; duplicate discoveries were removed, and novel phenotypes associated with a known disease-causing gene were not included. Detailed information is available in Supplementary information S1 (table).

cooperation and coordination in solving the remaining several thousand single-gene disorders.

NGS-based rare-disease-causing-gene discovery

Both whole-genome sequencing (WGS) and whole-exome sequencing (WES) are powerful, unbiased approaches for detecting genetic variation within an individual. However, because of the breadth and inherent complexity (as well as the greater cost) of WGS, WES is currently the more popular platform for the discovery of rare-disease-causing genes. In WES, the ~1% protein-coding portion of the human genome (the exome) is enriched by one of several capture approaches, sequenced by NGS and then analysed (see REF. 10 for an in-depth review on WES technology and data analysis). The initial proof-of-concept for using WES in rare-disease research came with the identification of genes responsible for the dominant Freeman–Sheldon syndrome[11], recessive Miller syndrome[12] and dominant Schinzel–Giedion syndrome[13]. To comprehensively assess the pace of gene identification since these initial findings, we have summarized the literature describing the WES-based identification of rare-disease-causing genes (see Supplementary information S1 (table)). Although the processes, strategies and optimal conditions for the identification of rare-disease-causing genes by NGS are still being refined, this field of research has matured, with more than 180 novel genes having been discovered using this technology.

Capture approaches
Technologies based on hybridization using RNA or DNA baits to target and enrich for a genomic region of interest for subsequent next-generation sequencing.

Freeman–Sheldon syndrome
A rare genetic disease characterized by contractures of the hands and feet, oropharyngeal abnormalities and distinctive facial features, including a very small mouth, puckered lips and an H-shaped dimple on the chin.

Miller syndrome
A rare genetic disease characterized by extensive facial and limb defects, including malar hypoplasia, down-slanting palpebral fissures, micrognathia, cleft lip and palate, cup-shaped ears, lower-lid ectropion, postaxial limb deficiencies and syndactyly.

Strategies for the identification of rare-disease-causing genes. Standard pipelines are now in place to process the sequencing data generated by WES or WGS, including mapping, variant calling and annotation. The sequence data can be compared with various public databases (including the single-nucleotide polymorphism (SNP) database (dbSNP)[14], the 1000 Genomes Project[15], the Exome Variant Server and International HapMap Project[16]), as well as internal control databases. These comparisons reduce the ~20,000 variants that are typically identified by WES (when compared with a reference genome) to <500 rare variants (defined as occurring at a frequency of ≤1% in controls) per exome. Initially, both inherited variants and *de novo* variants are catalogued; the subsequent validation of a variant as definitively disease causing is frequently the rate-limiting step (see REFS 17,18 for in-depth reviews on establishing causality). For a well-defined rare disease, the detection of mutations in the same gene in unrelated individuals or families results in a comparatively straightforward genetic validation; when only one family is available for genetic analysis, pathogenicity might be supported by functional studies. In the following sections, we discuss specific strategies for the identification of disease-causing genes based on the inherited, *de novo* or somatic mosaic nature of the disease mutation (or mutations).

Identifying inherited mutations. When there is familial recurrence of a defined rare phenotype or parental consanguinity, the likelihood that a rare disease is monogenic is high. The mode of inheritance influences the selection and number of individuals to sequence, as well as the analytical approach used (FIG. 2a).

Autosomal recessive disorders have been over-represented in the early stages of NGS-based gene discovery, with >115 novel genes being identified so far (see Supplementary information S1 (table)). In families demonstrating autosomal recessive inheritance, in the absence of consanguinity or occurrence in an isolated population, compound heterozygous mutations are predicted. The identification of such mutations is comparatively straightforward, as they are few in number. In some instances, an autosomal recessive disease-causing gene can be identified using just one affected sibpair, as was the case with the compound heterozygous *DDHD2* mutations that were found to be responsible for a complex form of hereditary spastic paraparesis. This was followed by the rapid identification of additional *DDHD2* variants in patients with a similar phenotype through international networks[19].

The identification of genes for recessively inherited diseases in consanguineous pedigrees, in which homozygous mutations are anticipated, has also substantially contributed to this discovery rate. The number of homozygous variants in the exome of an individual born to related parents reflects the degree of consanguinity and affects the number of variants that need to be considered. For example, the exome of an individual born to first-cousin parents contains 15–20 homozygous rare variants (defined as occurring at a frequency of ≤1% in controls), whereas the exome of offspring of

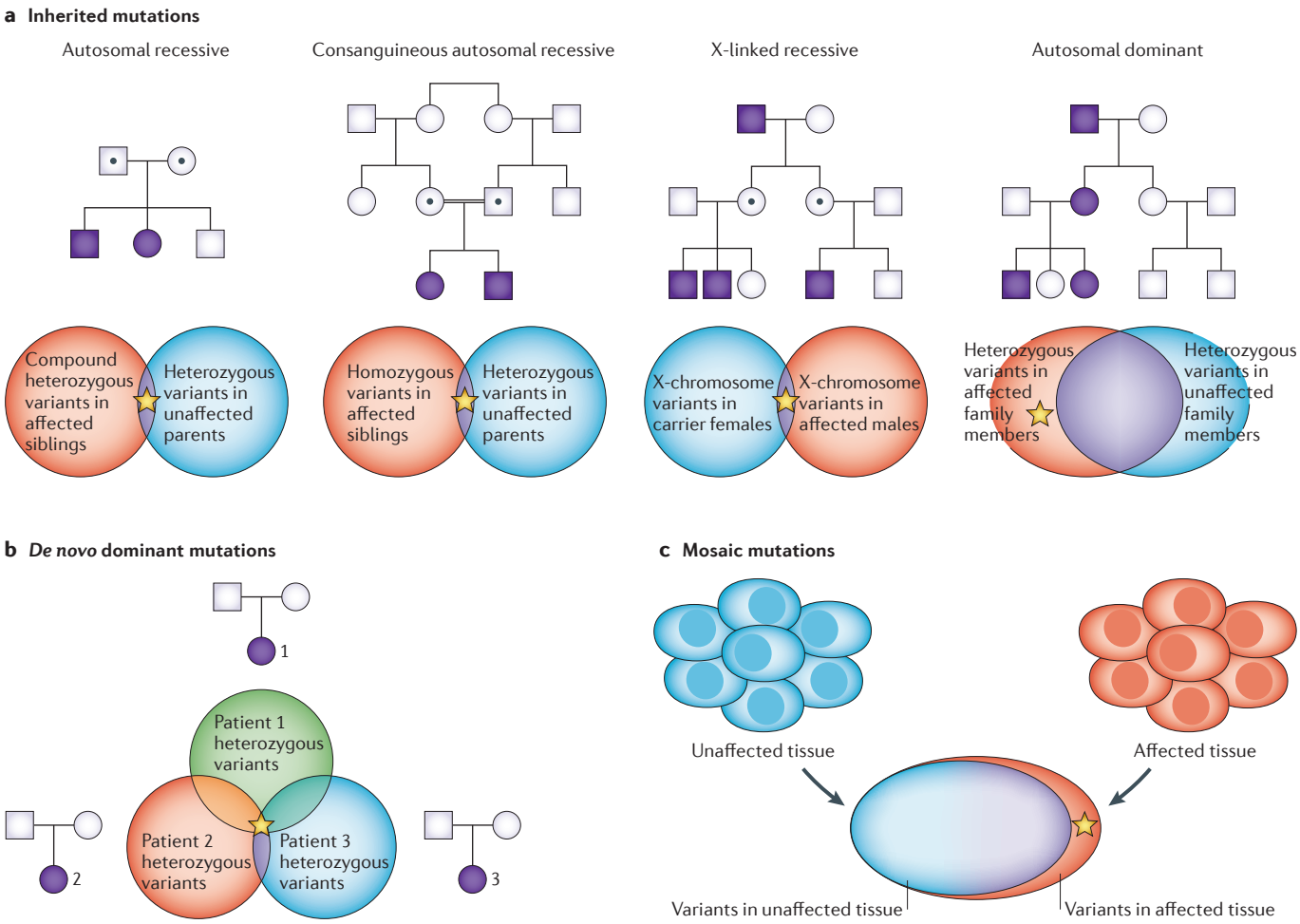

Figure 2 | **Gene identification approaches for different categories of rare diseases.** Representative family structures are indicated by the pedigrees for each type of mutation. Males are indicated by squares, and females by circles. Purple symbols indicate individuals affected by the rare disease. A dot in the centre of a symbol indicates that the individual is a carrier of the rare-disease-causing mutation. Stars highlight the search space that is predicted to contain the disease-causing gene. Selection of the appropriate gene discovery approach is contingent on whether the mutations are anticipated to be inherited, *de novo* or mosaic. When there is either a familial recurrence of a rare phenotype or the presence of consanguinity, the likelihood of a monogenic disease is high. The mode of inheritance influences the selection and number of individuals sequenced and the analytical approach used. **a** | For autosomal recessive disorders, sibpair analysis is often needed to reduce the number of gene variants to one or a few candidates; this is true even in consanguineous families. For X-linked recessive diseases, the favoured strategy is to analyse the two most remotely related male family members. For autosomal dominant disorders, the mapping of the gene to a discrete chromosomal region (for example, <2 Mb) may allow gene identification from the analysis of one individual; larger genomic regions or diseases which are not mapped require the analysis of a greater number of individuals. **b** | Analysis of whole-exome sequencing data from unaffected parents–affected child trios generally produces a handful of *de novo* variants for further analysis; comparison of these variants between as few as two families will generally reduce these to a single candidate gene. **c** | The comparison of sequence data from a patient's affected and unaffected tissue is frequently sufficient to identify *de novo* mosaic disease-causing mutations.

Schinzel–Giedion syndrome
A rare genetic disease characterized by severe mental retardation, distinctive facial features and multiple congenital malformations (including skeletal, genitourinary, renal and cardiac malformations).

Isolated population
A group of individuals who are descended from a small number of settlers (founders) and remain genetically (reproductively) isolated.

Compound heterozygous mutations
Two different mutations present in the same gene but arranged in *trans*, such that each copy of the gene in a diploid organism carries one of the mutations.

third-cousin parents will have only 5–10 homozygous rare variants. The number of candidate homozygous rare variants present genome-wide can be further reduced if a genomic region shared between affected siblings is identified. For example, analysis of three affected siblings from a consanguineous Pakistani family with autosomal recessive postaxial polydactyly type A identified a shared region containing a single homozygous *ZNF141* variant[20]. However, for rare disorders like this one, reported in only one family, the challenge becomes either genetic or functional validation of the gene as disease causing, owing to the difficulty in identifying additional families for genetic analysis and the lack of understanding of gene function, respectively.

X-linked pedigrees have also been tractable to NGS analyses. In families showing unequivocal X-linked inheritance, one can safely disregard autosomal variants. However, the pedigree structure frequently does

not allow differentiation between X-linked and autosomal recessive inheritance, as was the case for Diamond–Blackfan anaemia[21]. A WES analysis of affected male siblings identified no variants that fit an autosomal recessive mode of inheritance. However, a single *GATA1* variant on the X chromosome demonstrated appropriate segregation, and the gene was subsequently validated using an additional cohort of male patients with Diamond–Blackfan anaemia[21].

NGS-based gene identification for familial autosomal dominant disorders has been more challenging; most successes have been associated with a defined disease interval identified by linkage analyses of a large pedigree (or pedigrees). For example, recent analysis of a Norwegian family with a novel dominant familial diarrhoea syndrome identified a 2.9 Mb linked region containing just one rare variant, a heterozygous missense mutation in *GUCY2C*[22]; subsequent functional analyses in cells from patients suggested a gain-of-function effect[22]. However, in general, genes causing dominantly inherited disorders have proved to be more difficult to identify. This can be either because of a small family size and thus a large number of candidate heterozygous variants, or because of the absence of obvious disease-causing variants in the mapped region (or regions) in larger families, raising the possibility of non-coding mutations being responsible.

Identifying* de novo *dominant mutations. *De novo* mutations causing autosomal dominant disorders have proved to be much easier to identify, given that each individual carries very few variants that are not also found in their parents, resulting in a data set that is much less complex. The analysis of just two unrelated parent–child trios with a sporadic presentation of the same presumed autosomal dominant disorder can be sufficient for the identification of *de novo* mutations in the disease-causing gene (FIG. 2b). This was the case in the recent identification of *EZH2* as the causal gene for Weaver syndrome[23]. When the analysis of parent–child trios is not possible, the intersection of heterozygous variants across unrelated probands with the same *de novo* autosomal dominant disorder can also be used, although more patients are generally needed for this approach. For example, an analysis of five unrelated patients identified a mutation in *SRCAP* as the causative mutation for Floating–Harbour syndrome[24]. Thus, both approaches are effective, and the analytical design can be tailored to the patient and/or family resources available at the outset.

Identifying mosaic mutations. One of the most exciting applications of NGS has been the sequencing of DNA from a patient's affected and unaffected tissues to identify disease-causing mosaic mutations (FIG. 2c). Usually, a dominant mutation would be present in approximately 50% of the sequence reads; in the case of mosaicism, the frequency of the mutation will be lower and tissue dependent. For example, a WES study carried out on patients with Proteus syndrome identified a somatic *AKT1* oncogene-activating mutation in 1–50% of reads from affected versus normal tissues from patients, consistent with mosaicism[25]. NGS data can also be analysed to detect low levels of mosaicism in a single tissue from a patient. For example, WES was used to look for *de novo* mutations that are causal for megalencephaly–capillary malformation (MCAP) syndrome in leukocytes from a trio consisting of unaffected parents and an affected child. Adapting the analysis to identify mosaicism, a *de novo* mutation was identified as being present in 11% of the reads in *PIK3CA* in the child. Visual inspection of WES data from six patients with MCAP revealed mosaic mutations in *PIK3CA*. Although the accurate differentiation between sequencing error and a bona fide mosaic mutation at a target site is a challenge, mosaicism levels as low as 2% could be distinguished from sequencing errors[26]. Mutations in this gene were subsequently identified in 29 patients in total, with mutant allele frequencies ranging from 10% to 50% in multiple tissues[26].

Success rate of NGS-based gene discovery. The success rate of NGS-based identification of rare-disease-causing genes is difficult to determine precisely because negative studies tend not to be published. It is clear that the genes causing recessive and *de novo* dominant disorders are easier to identify than others and have thus been over-represented so far. Of 100 clinically well-defined disorders that entered the FORGE Canada (Finding of Rare Disease Genes Canada) discovery platform, 66 were explained at the molecular level, and 50% of the 66 were novel gene discoveries. Importantly, the FORGE disorders were those that are most likely to be genetic and thus tractable by WES. Of the 34 disorders that were studied but are not currently understood at the molecular level, roughly two-thirds had too many credible candidate variants, and the remaining one-third had none. Interestingly, those disorders in the latter category tend to be autosomal dominant, and the regions containing the critical gene had been mapped for many of them. This suggests, as discussed above, that the causal mutations reside outside the exome.

Maintaining the pace of novel gene discovery using NGS. WES has led to the identification of more than 180 distinct novel disease-causing genes, more than 130 of which were reported in 2012 alone (FIG. 1). It is likely that NGS technology will continue to become less costly and WES analyses of individuals with single-gene rare diseases will increase in number. The same is likely to be true for WGS when the cost and interpretability of this technique approach those of WES. However, the current increase in the pace of discovery for disease-causing genes may be moderated as fewer tractable diseases are analysed. These diseases include those which are so rare that the identification of additional families worldwide is required, and those for which the causal mutation lies in a non-coding region. Non-coding causal mutations might prove especially refractory to analysis, as their identification will require WGS, which, compared with WES, will identify many more sequence polymorphisms that have no connection to the disease. Because

Sibpair
Two siblings with both parents in common.

Postaxial polydactyly type A
A congenital anomaly characterized by fifth-digit duplications in hands and/or feet. In the type A disorder, the extra digit is well formed and articulates with the fifth or an extra metacarpal.

Diamond–Blackfan anaemia
A rare genetic disease characterized by congenital erythroid aplasia and congenital anomalies, particularly of the upper limb and craniofacial regions.

Gain-of-function effect
Pertaining to a mutation: the acquisition of a new and abnormal function by a gene product when the mutation is present in the heterozygous state.

Weaver syndrome
A rare genetic disease characterized by pre- and postnatal overgrowth, accelerated osseous maturation, characteristic craniofacial appearance and developmental delay.

Floating–Harbour syndrome
A rare genetic disease characterized by proportionate short stature, delayed bone age, delayed speech development and typical facial features.

Mosaic mutations
Mutations that are present in only a proportion of cells in the body.

Proteus syndrome
A rare genetic disease characterized by patchy or mosaic overgrowth and hyperplasia of various tissues and organs.

there is currently no rigorous means by which to filter the variants (as is the case with coding mutations) and no clear means of delineating biological causation, pathogenic validation will be a rate-limiting step. The speed of completion of the rare-genetic-disease 'atlas' is thus contingent on the proportion of rare diseases that are caused by non-coding mutations; if this class constitutes a large fraction of the remaining disorders, it might be a decade or more before the atlas is completed.

Thus, it is clear that an unprecedented level of international collaboration will be needed to accelerate or even maintain the pace of gene discovery. This will require increasingly sophisticated infrastructure and tools to share and integrate carefully curated and annotated phenotypic and genetic data sets. In this regard, the Human Phenotype Ontology now has more than 10,000 terms, each of which describes a phenotypic abnormality seen in human disease[27]. Ontologies such as this one provide a basis for computational analysis of the human phenotype and facilitate the capture, storage and exchange of such data. However, the algorithms required to integrate phenotypic and genetic data sets on a large scale are still in their infancy; with the emergence of such algorithms and increased collaboration, the rate of novel gene discovery might grow appreciably in the near future. A promising initiative towards infrastructure development and large-scale collaboration is seen in the International Rare Disease Research Consortium (IRDiRC), an umbrella organization for >30 participating global funding organizations and their aligned research projects, launched in 2011. One of the main objectives of IRDiRC is the provision of diagnostic tools for all rare genetic diseases by the year 2020. The current rate of gene discovery will have to at least triple if the molecular definition of the remaining 3,500 rare disorders is to occur within this time.

Insights from gene discovery

Depth of rare diseases. NGS-based discovery of rare-disease-causing genes has so far yielded several interesting findings regarding ostensibly novel phenotypes. These findings may necessitate a re-estimation of the total number of human disease-causing genes — the true depth of rare diseases. Rare diseases that were thought to be novel (that is, unlike any previously described conditions) can, in reality, be two rare diseases segregating in the same family. One example is the case of a Newfoundland family in which two affected siblings were thought to have a novel, variably penetrant syndrome characterized by ocular and skin hypopigmentation, congenital neutropenia, immune dysregulation and Crohn's disease. Using WES, both siblings were shown to have severe congenital neutropenia type 4 with the addition of oculocutaneous albinism type 4 in one sibling[28]. It is likely that other 'unique' phenotypes represent a conflation of two known phenotypes in this way.

For known diseases, atypical or unusually complex presentations that were initially suspected to be novel phenotypes are also being observed. These presentations are often the result of a dramatic increase or decrease in severity compared with most cases of the disease. This was recently illustrated in the case of two brothers with an undiagnosed neurodegenerative syndrome. In this case, analysis of WES data showed that the brothers were presenting with a previously unreported, milder subtype (type 4) of the known condition D-bifunctional protein deficiency[29]. In addition, known disease-causing genes will be increasingly implicated in phenotypes that are new and are caused by a distinct pathological mechanism. For example, Hajdu–Cheney syndrome was found, by WES studies, to be caused by carboxy-terminal truncating mutations in *NOTCH2* (REFS 30–32), a gene in which missense mutations have also been shown to cause a variant form of Alagille syndrome[33].

We anticipate that the proportion of ostensibly novel phenotypes that will be reassigned as atypical presentations of known disorders will increase as the compendium of human mutations is filled in. In addition, we expect that there will be a growing list of genes in which distinct mutations cause separate rare diseases. In our review of the more than 300 WES reports in the literature, we have identified ~15% that describe novel and distinct phenotypes for mutations in genes that were previously associated with a different rare disease (see Supplementary information S1 (table)). As we come closer to understanding the genetic aetiology of all rare diseases, it is likely that we will increasingly rediscover known genes and that the approach to completion of the disease compendium will be asymptotic. The accessibility of comprehensive NGS-based diagnostics for patients with rare diseases will be absolutely central to understanding the complete atlas of human genetic pathology. This will enable the phenotypes that are understood at a genetic level to be quickly identified, and for those that require it, further research can be undertaken to discover novel disease-causing genes.

NGS as a diagnostic tool. The search for the underlying cause of a patient's or family's novel rare disease frequently ends with the identification of a mutation (or mutations) in a known disease-causing gene. This is a readily interpretable result that requires no further analysis, and NGS is effectively a diagnostic rather than a discovery tool in these cases. In our literature analysis, of the more than 300 publications concerning rare diseases studied by WES, ~25% report mutations that are in known disease-causing genes which result in a phenotype that, in retrospect, matches the clinical presentation of the patient being investigated (see Supplementary information S1 (table)); one can safely assume that the number of unpublished studies identifying known disease-causing genes is much higher. One-third of all the projects studied by the FORGE Canada Consortium resulted in the identification of mutations in a known disease gene for which the patient's phenotype matched the clinical presentation of the identified disease-causing gene. As the list of known rare-disease-causing genes grows, so does the probability that mutations in these genes will be identified, resulting in a de facto conversion of NGS from a research tool to a diagnostic platform (BOX 1). The application of NGS in

Hajdu–Cheney syndrome
A rare skeletal disorder characterized by short stature, coarse and dysmorphic facial features, bowing of the long bones, vertebral anomalies, acroosteolysis and generalized osteoporosis.

Alagille syndrome
A rare disease characterized by cholestasis (caused by bile duct paucity), congenital cardiac defects, posterior embryotoxon in the eye, typical facial features and butterfly vertebrae.

Box 1 | Translation of next-generation sequencing into clinical diagnostics for patients with rare diseases

Many patients with rare diseases who are undergoing traditional diagnostic assessment never receive a molecular diagnosis. This results in a lack of knowledge regarding mutational content, a rudimentary insight into natural history and adverse impacts on patient care. Because of clinical and genetic heterogeneity, the rarity of a given condition or an atypical presentation, many diagnostic assessments are lengthy, expensive and ultimately futile, often costing more than US$10,000 (REF. 55). For example, the diagnosis of conditions such as retinitis pigmentosa or Charcot–Marie–Tooth disease, which may result from a mutation in any of the more than 50 candidate genes, presents an expensive challenge using conventional technologies. The implementation of next-generation sequencing (NGS)-based diagnostics for rare diseases using approaches such as targeted resequencing, whole-exome sequencing (WES) and whole-genome sequencing (WGS) is well underway, although hurdles remain for the broad implementation of the unbiased approaches of WES and WGS (see REF. 56 for an in-depth review on the future of molecular genetic testing).

Ultimately, WES or WGS will become part of a standard assessment for most patients suspected of having a rare genetic disorder, dramatically shortening the diagnostic process. The pace of this transition will depend on the presence of the requisite infrastructure, regulatory standards, training and best practice guidelines for reporting, and will vary widely between and even within countries. Regulatory standards and evaluation protocols focusing on nucleic acid preparation, the sequencing process and bioinformatic analyses, including data storage, will all need to be formulated. Numerous computational analytical approaches are currently in various stages of development and use; the standardization of these programs is required. The analytical process and its output must be set to ensure that sequence variants, copy number variation and mosaicism can be reliably identified.

Extensive and well-annotated reference data sets will be needed, not only for disease mutations but also for innocuous genomic variations. There exist a myriad of disease- and gene-specific mutation databases, but there is currently no accurately annotated, clinical-grade mutation database. In a recent study, 27% of mutations cited in the literature were found to be common polymorphisms or misannotated[57], underscoring the need for better mutation databases as part of the comprehensive reporting of NGS-based clinical testing. Frameworks for the reporting of diagnostic results[58,59] will help to facilitate the dialogue between clinician and patient to determine the types of incidental findings that will be returned. More broadly, standardized educational programmes, including those that cover NGS technology and data analysis, must be integrated into the training programmes for molecular diagnosticians and medical geneticists. This integration is needed to ensure that the appropriate and highly qualified personnel are ready to facilitate the translation of NGS technology into the rare-diseases clinic.

Genetic heterogeneity
Pertaining to a phenotype: caused by the alteration of one of many different genes.

Chromosomal microarray
An approach based on probe–target hybridization to detect amplifications or deletions of chromosomal regions in a patient's tissue.

Post-zygotic mutations
Mutations that an organism acquires during its lifespan. Also known as somatic mutations.

Megalencephaly–capillary malformation syndrome and megalencephaly–polymicrogyria–polydactyly–hydrocephalus syndromes
A class of rare genetic diseases characterized by congenital or early postnatal megalencephaly (large brain), prenatal overgrowth, brain and body asymmetry, cutaneous vascular malformations, digital anomalies and connective tissue dysplasia.

Hemimegalencephaly
A rare developmental malformation characterized by the enlargement of one-half of the brain. Also known as unilateral megalencephaly.

Coffin–Siris syndrome
A rare genetic disease characterized by mental retardation, coarse facial features, hypertrichosis and hypoplastic or absent nails on the fifth fingers or toes.

Joubert syndrome
A rare genetic disease characterized by hypotonia, developmental delay and hypoplasia of the cerebellar vermis with the characteristic neuroradiologic 'molar tooth sign'.

Loss-of-function alleles
Alleles that partly or fully eliminate normal protein activity.

Dominant negative
Pertaining to a mutation: having a negative impact on the biological function of the remaining wild-type gene product when the mutation is present in the heterozygous state.

the rare-diseases clinic will blur the boundaries between clinical genetic testing and research, as did the detection of chromosomal anomalies by chromosomal microarray in the previous decade.

Biological insight. NGS-based gene identification is making this an exciting time in biomedical research, as these studies frequently place proteins in known disease pathways or elucidate new pathways. For example, analysis of WES data has recently linked several overgrowth disorders to the dysregulation of the well-studied AKT–PI3K–mTOR pathway. A mosaic activating mutation in *AKT1* has been identified as the underlying cause of Proteus syndrome[25]; germline mutations (in *AKT3* and *PIK3R2*) or post-zygotic mutations (in *PIK3CA*) have been identified as causative for megalencephaly–capillary malformation and megalencephaly–polymicrogyria–polydactyly–hydrocephalus syndromes[26]; and mosaic *PIK3CA*, *AKT3* and *mTOR* mutations have been shown to be causal for hemimegalencephaly[34]. These findings have linked this central pathway to vascular, limb and brain development. Other disease-causing genes identified by NGS have implicated central cellular pathways in human development and disease; one example is the identification of mutations in six chromatin-modelling SWI/SNF genes (*SMARCB1*, *SMARCA4*, *SMARCA2*, *SMARCE1*, *ARID1A* and *ARID1B*) in the multiple congenital-anomaly disorder Coffin–Siris syndrome[35]. In addition, a newly identified gene can be placed in an established category of disorders with a common pathogenesis, although the exact pathogenic mechanism might be obscure. An example is the recent identification of *C5ORF42*, a gene with no known functional domains, as being causal for the classic Joubert syndrome ciliopathy in French Canadians[36]. This finding clearly suggests that the encoded protein is involved in the development and/or function of cilia, but in an unknown manner. Finally, just as the identification of rare-disease-causing genes in the past has provided insight into the aetiology of the risk and pathology of complex and common diseases, we anticipate that this collateral benefit will increase with NGS-based discoveries (BOX 2).

In general, we are hampered by our incomplete understanding of the biological function of most genes and proteins; the linking of a poorly characterized gene to a human disease does not necessarily make the protein function clearer. Although *in vitro* analysis with cell lines from patients can considerably contribute to our understanding of protein function, often a more comprehensive investigation at the tissue, organ or whole-organism level is required. Thus, there is a need for coordinated model-organism research platforms to put disease-causing genes into a biological context. One such effort is the International Knockout Mouse Consortium, which aims to mutate all protein-coding genes in the mouse and is providing resources to many laboratories that are studying the effects of loss-of-function alleles in human disease. As more gain-of-function and dominant-negative rare-disease-causing mutations are identified, there will be an additional need for knock-in models to recapitulate

Box 2 | **Insights into complex diseases and common disorders**

In the past, the identification of rare disease-causing genes has provided insight into the aetiology of the risk and pathology of complex diseases (for example, Alzheimer's disease, reviewed in REF. 60); we anticipate that similar collateral benefit will increase with next-generation sequencing (NGS)-based discovery of rare-disease-causing genes. For example, whole-exome sequencing (WES) has recently identified mutations in the actin-binding profilin 1 (*PFN1*) gene in familial as well as sporadic cases of amyotrophic lateral sclerosis (ALS), implicating cytoskeletal dysregulation in the pathogenesis of ALS[61]. This is also an example of a subtype of a complex disease being caused by a mutation in a highly penetrant gene for a substantial subset of patients.

Intellectual disability and autism are both common disorders, each with a prevalence estimated at almost 1% in the general population[62,63]. A link between genetics and these neurodevelopmental disorders has long been recognized. Recently, NGS-based analysis has revealed the role of *de novo* mutations in these disorders (reviewed in REF. 64). WES studies of patients with non-syndromic intellectual disability have shown that a diagnosis can be achieved in a substantial proportion of patients (16%) and that novel genes implicated in the disease can be identified[65,66]. As a result, an important role for copy number variation and *de novo* point mutations is emerging in the aetiology of non-syndromic intellectual disability, and it seems that a substantial subset of cases of this particular neurodevelopmental disorder is likely to turn out to be a large number of distinct rare diseases. Similarly, studies of patients with autism reveal extreme genetic heterogeneity, but in this instance, the emerging model of disease is oligogenic, with the sum of rare *de novo* and inherited mutations contributing to the overall genetic risk in a particular patient[67–70]. As the research field moves forwards, it is likely that the biological pathways which have been revealed to underlie syndromic forms of neurodevelopmental disorders will continue to be implicated more broadly in idiopathic intellectual disability and autism.

these diseases. Knock-in models will also be important for the study of these human recessive disorders that involve partial, instead of complete, loss-of-function mutations. An important and growing challenge is to overcome the disconnect between the static number of research laboratories and the increasing number of disease-causing genes that have been identified. To this end, the generation of a standardized tool box for the majority of human disease-causing genes will enhance our ability to study the disease pathways associated with these newly discovered genes and to move a subset of these disorders closer to effective therapies[37].

Paths forwards for rare-disease therapy

The exact number of therapies that are currently available for rare diseases is difficult to gauge. In Europe, there are 144 medicinal products with European marketing authorization for the treatment of more than 120 rare diseases between them. In the United States, there are 340 medicinal products with both marketing authorization and a designation as orphan drugs for the treatment of 355 different rare-disease indications. There are 66 medicinal products that are on both lists, bringing the total number of treatments for rare diseases to approximately 420 (S. Aymé and V. Hivert, personal communication). This is the number of medicinal products only; non-pharmaceutical therapies will substantially add to this list.

The increased pace of identification for genes that cause rare diseases means that, in effect, there is an almost commensurate increase in the number of molecularly defined, readily diagnosable, but nonetheless untreatable, diseases. Occasionally, with the identification of a causative gene and the resultant molecular insight into a given disorder, an obvious and facile therapeutic approach will present itself. For example, this happened following the WES-based identification of *SLC18A2* as the causative gene for an infantile-onset movement disorder characterized by severe parkinsonism, non-ambulation, mood disturbance, autonomic instability and developmental delay[38]. The gene encodes VMAT2 (also known as SVM2), a translocator of dopamine and serotonin into synaptic vesicles, suggesting that direct dopamine agonists could be used as therapeutic agents; consistent with this hypothesis, treatment with dopamine agonists resulted in a marked improvement in symptoms and the resumption of development. Similarly, the identification of mutations in the riboflavin transporter genes *SLC52A3* (REFS 39,40) and *SLC52A2* (REF. 41) as the cause of Brown–Vialetto–Van Laere syndrome (progressive sensorineural deafness in combination with childhood amyotrophic lateral sclerosis) suggested that riboflavin could be used as a treatment; initial work has shown encouraging results for riboflavin therapy in terms of both the biomarker response and clinical end-points[42].

Such outcomes, although exciting, are rare. Indeed, for a large number of diseases, given their congenital nature and the fact that they often involve structural defects in early development or are overwhelmingly severe, the prospect of a definitive therapy is unlikely. For the remainder of disorders for which there may be an existing therapeutic opportunity, the diversity and number of rare diseases combined with the small numbers of patients for each disorder effectively precludes, for all but a fraction of conditions, traditional costly drug discovery approaches. To provide further context, based on the current rate of approval of orphan drug products in Europe, we predict that there will be approximately 75 new approvals in the next 20 years (FIG. 3), little more than a drop in the ocean when considering the large number of potentially treatable rare genetic diseases.

Therapeutic approaches for rare diseases. The majority of genetic disorders can be viewed as a dysregulation of dosage that results in supraphysiological or infraphysiological levels, usually of protein activity; translational approaches can thus attempt to normalize these pathogenic levels of protein activity (TABLE 1). The potential therapy can be direct in nature, aimed at rectifying the immediate impact of the genetic lesion. For disorders that result from loss-of-function mutations, which are usually recessive, one can potentially replace the relevant DNA or protein, or increase transcript or protein levels. For gain-of-function disorders, which are usually dominant, approaches can be used to decrease mRNA or protein levels, or to inhibit protein activity. In addition to addressing the causal DNA or protein lesion, a rare-disease-causing mutation often has specific downstream consequences that may account for the majority of the pathogenic burden and can therefore be credible therapeutic targets.

Supraphysiological
At levels greater than those normally found in the body.

Infraphysiological
At levels below those normally found in the body.

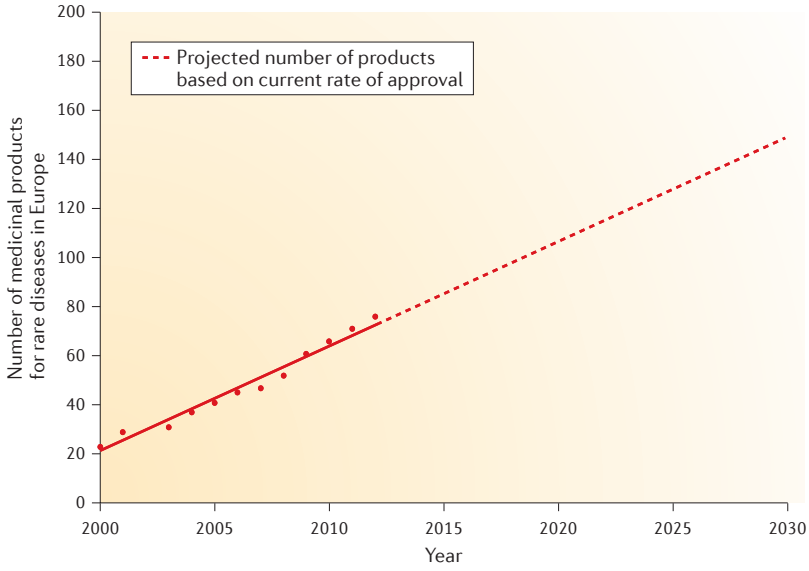

Figure 3 | **Rate of approval of orphan drug products.** Extrapolation from the current rate of medicinal-product approval for rare-disease treatments in Europe (obtained from the January 2013 Orphanet Report Series: Lists of medicinal products for rare diseases in Europe; see Orphanet for the latest report) suggests that, at current levels, only an additional 75 products will become available for patients with rare diseases in the next ~20 years. This is less than half the number of new genes discovered annually and underlines the widening gap between the identification of rare-disease-causing genes and the development of new therapies.

There exists a clear and growing gap between the identification of rare-disease-causing genes and the formulation of effective therapies; this is exacerbated by the fact that the majority of rare diseases are not the focus of a research laboratory or any translational research programme and for most there is no interested charitable agency. Research into rare-disease therapeutics will require more creative strategies, such as translational research programmes across a broad spectrum of rare diseases using rapid, low cost, low risk and generalizable approaches (BOX 3). In addition to these approaches, there are various other avenues that might prove successful. These include the use of proteasomal inhibitors, which have recently been shown as promising for the treatment of Pompe disease[43] and may also be effective for treating a larger number of recessive disorders with partially disabling mutations. The establishment of a library of diverse, clinically tested gene-therapy vectors (such as self-complementary adeno-associated viruses) with well-characterized and distinct human tissue tropism would be a resource of considerable value for the rare-disease community. Similarly, the exploration of new, rapid and high-yield protein production methods to lower the cost of enzyme replacement therapy would be valuable. More generally, we need to determine how often the reversal of cellular phenotypes for monogenic disorders can predict *in vivo* responses with a useful degree of accuracy. Such accuracy would facilitate the rapid identification of treatments using medium-throughput assays with clinically approved agents.

Pompe disease
A rare lysosomal storage disease characterized by cardiomyopathy and muscular weakness. Also known as glycogen storage disease type II.

A new model for rare-disease translational research. Funding allocation for rare-disease research is highly variable and correlates poorly with the severity and prevalence of a given disease. The development of academic and pharmaceutical-industry research programmes for some of the less rare and better known diseases has fostered an asymmetrical growth in these areas[44]. It might be argued that this is a necessary asymmetry; the initial focus of resources on some conditions will ultimately allow a solution to be found for the tractable remainder. In this regard, it is worth noting that two of the more 'common' rare diseases, cystic fibrosis and Duchenne muscular dystrophy (DMD), have been the subject of hundreds of millions of US dollars' worth of preclinical and clinical assessment. However, only recently has an effective therapy been developed for one specific and comparatively rare subclass of cystic fibrosis[45], and the potential therapies for DMD treatment are still mainly at preclinical stages. Notwithstanding this slow progress, many of the approaches which showed promise for treating other rare disorders have been pioneered in preclinical work on both cystic fibrosis and DMD.

The role of the pharmaceutical industry in the development of therapies for rare diseases is, to some extent, circumscribed by the fact that only a minority of such disorders have sufficient prevalence to maintain a traditional drug discovery programme. Although NGS-based molecular diagnoses will reveal higher prevalences for many conditions (perhaps doubling or even tripling the number of 'marketable' diseases in the future), the number of disorders studied by the pharmaceutical industry will always include only a very small proportion of all genetic diseases. Given this reality and the general difficulty of new-drug discovery, there is a growing interest in pre-competitive, open-access translational research in which work is carried out transparently and results are freely available to all[44]. In addition, new funding models are needed for rare-disease research; such models must emphasize coordinated support from public funders and charitable organizations, as well as from industry and venture capital if possible.

Regulatory issues. Even when there is a credible therapeutic avenue for a rare disease, onerous regulatory commitments designed for more common disorders can often hamper the progress of developing such therapies. The small number of patients with a particular rare disease precludes the development of a strong evidence base for drug efficacy, or even safety, which makes drug approval by standard approaches problematic. A modification of current regulatory processes that balances patient safety with accurate clinical assessment is clearly called for; regulatory agencies are increasingly recognizing the need to rework existing licensing processes to meet the special challenges of the development of rare-disease therapies. This reform will need to include the judicious revisiting of allowable evidence when assessing drug efficacy and safety, including the use of surrogate outcomes.

Table 1 | **Direct therapeutic approaches to treat rare diseases**

Approach	Intervention	Disease examples	Refs
Loss-of-function, usually recessive disorders			
DNA replacement	Gene therapy	Severe combined immunodeficiency	75
	Bone marrow transplantation	Mucopolysaccharidoses	76
Splicing correction	Antisense oligonucleotides	Duchenne muscular dystrophy (preclinical)	77
	Small molecules	Familial dysautonomia (preclinical)	78
mRNA increase	Small molecules	Spinal muscular atrophy (preclinical)	71,72
Protein replacement	Enzyme replacement therapy	Lysosomal storage diseases	79
Increase in protein activity, stability or level	Translational readthrough	Duchenne muscular dystrophy (preclinical)	80
	Chaperonin therapy	Cystic fibrosis	45
		Transthyretin amyloidosis	81
	Proteasome inhibition	Pompe disease (preclinical)	43
Gain-of-function, usually dominant disorders			
Transcriptional downregulation	Antisense oligonucleotides	Myotonic dystrophy (preclinical)	82
	RNA interference	Huntington disease (preclinical)	83
Protein inhibition	Small molecules	Noonan syndrome (preclinical)	84

Future perspectives

The necessity for a Human 'Phenome' Project. The next 5–10 years will witness an unprecedented growth in our understanding of human biology in terms of disease as well as normal development and health. However, the large amount of genetic data generated from NGS-based approaches brings with it the parallel need for large-scale phenotypic annotation, as would be encapsulated in a Human Phenome Project[46]. How the research community prepares for such a project was the focus of a recent forum of the Human Variome Project[47], which is a global initiative to collect and curate all human genetic variation affecting health, with the ultimate goal of reducing disease. Currently, our limited ability to capture, analyse and share deep phenotypic data and genomic-variation data on a large scale is a major impediment to progress in our understanding of rare genetic diseases and of human genome biology in general. Several initiatives are valuable examples for a Human Phenome Project. For instance, the International Standards for Cytogenomic Arrays Consortium has collected phenotypic data from patients investigated in more than 28,500 array-based studies[48]. Similarly, projects focused on the mouse[49], zebrafish[50], and rat[51,52] are collecting phenotypic data under standardized conditions in a centralized database. However, there is currently no coordinated effort towards understanding the human phenome by correlating standardized disease descriptions with genomic and other types of data.

Recent steps forwards for the collection, storage and analysis of standardized deep phenotypic data for patients participating in projects to identify rare-disease-causing genes include PhenoDB[53] and PhenoTips[54], both of which are freely available web-based tools. Such a phenome project will enable the rapid understanding of not only the highly penetrant mutations that cause rare genetic diseases but also the associated clinical variability. Identification and understanding of disease modifiers will enable the prediction of clinical severity so that in the future, we can inform patients and their families about this important aspect of disease management.

The evolving role of the medical geneticist. The NGS era of gene discovery and, ultimately, molecular diagnoses heralds a fundamental change for medical genetics and medical geneticists. In addition to continuing to define phenotypes, the medical geneticist of the NGS era will have an unprecedented opportunity to identify human disease-causing genes, enable the translation of NGS into diagnostic tools and understand the role of disease modifiers to advance the care of patients with rare diseases. These geneticists will also become vital members of collaborative teams along with providers from other specialties as they use genomic results to guide patient care. As such, medical geneticists will need to complement their clinical skills with expertise in the clinical interpretation of NGS data. With high-throughput testing becoming available for rare diseases, the diagnostic process will be re-engineered in the rare-disease clinic. Currently, time and money are spent on gathering data that can be used to group patients together or split them into precise categories, which facilitates the selection of one or more genetic tests to establish a molecular diagnosis. In the future, medical geneticists will instead phenotype patients to facilitate the interpretation of the large data set generated by WES or WGS. Perhaps most importantly, as increasing molecular insight inevitably suggests novel therapeutic avenues, some medical geneticists will also become interventionists, configuring, trialling and instituting treatments for rare disorders.

Deep phenotypic data
Data from the precise and comprehensive annotation of phenotypic abnormalities (clinical features) using a standard set of agreed descriptors (ontology).

Box 3 | **Rapid low-cost, low-risk approaches to therapy for rare genetic diseases**

Occasionally, rapid, low-cost, low-risk and potentially high-yield therapeutic approaches will present themselves when the molecular pathogenesis of a rare disease is revealed.

Dietary-substrate omission or supplementation

This approach can be considered for disorders in which there is a partially defective or haploinsufficient protein that catalyses, transports or in some other manner binds and processes a substrate. Either the accumulation of product is the key pathogenic factor, in which case dietary omission of the substrate may have a clinically beneficial effect, or a specific innocuous substrate can be given as a dietary supplement to result in increased production of the product.

Repurposing of clinically approved agents

Rare diseases are frequently found to be protein dosage problems. Drugs can have an impact on mRNA and protein levels in unanticipated ways; thus, the identification of established drugs that modify the levels of the relevant mRNA and protein at a clinically meaningful level is a promising avenue[37]. This approach is enabled by the mining of system-wide databases of drug-elicited transcriptional responses as well as the use of software programs predicting gene regulation pathways to rapidly identify drugs and drug classes that might affect a disease-modifying gene. A second approach uses the available increasingly detailed transcription factor binding-site maps that incorporate the majority of annotated genes. For a number of transcription factors, there are well-characterized, clinically approved agents which are expected to upregulate those genes containing the transcription factor-binding site; the use of this information has led to potential therapeutics for spinal muscular atrophy[71,72].

Occasionally, a rare-disease pathogenic pathway will overlap with one that is being targeted in common disorders. This phenomenon will occur with increasing frequency as the molecular 'atlas' of human genetic pathology is completed and the modern pharmacopoeia grows, and is best exemplified by Marfan syndrome and the beneficial TGFβ pathway modulation (reviewed in REF. 73). Another example is the study of the effect of kinase inhibitors originally formulated for cancer treatment on cardiac function in Noonan syndrome, an inherited kinasopathy caused by activation of the RAS signalling pathway[74].

Conclusion

Patients with rare genetic diseases are among the first beneficiaries of the NGS revolution; their experience will inform personalized medicine in other areas over the next decade. When we have elucidated the genetic basis of all rare genetic diseases, the emphasis will continue to shift towards understanding disease mechanisms and phenotypic heterogeneity, as well as developing high-throughput or generalizable therapies for many diseases. From our perspective, the focal point of this future should be the translation of these gene discoveries and all that comes with this new knowledge to the well-being of the patients and families living with rare diseases.

1. Carter, C. O. Monogenic disorders. *J. Med. Genet.* **14**, 316–320 (1977).
2. Baird, P. A., Anderson, T. W., Newcombe, H. B. & Lowry, R. B. Genetic disorders in children and young adults: a population study. *Am. J. Hum. Genet.* **42**, 677–693 (1988).
3. Yoon, P. W. *et al.* Contribution of birth defects and genetic diseases to pediatric hospitalizations. A population-based study. *Arch. Pediatr. Adolesc. Med.* **151**, 1096–1103 (1997).
4. McCandless, S. E., Brunger, J. W. & Cassidy, S. B. The burden of genetic disease on inpatient care in a children's hospital. *Am. J. Hum. Genet.* **74**, 121–127 (2004).
5. Dye, D. E. *et al.* The impact of single gene and chromosomal disorders on hospital admissions of children and adolescents: a population-based study. *Public Health Genomics* **14**, 153–161 (2010).
6. McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am. J. Hum. Genet.* **80**, 588–604 (2007).
7. Ayme, S., Urbero, B., Oziel, D., Lecouturier, E. & Biscarat, A. C. Information on rare diseases: the Orphanet project. *Rev. Med. Interne* **19**, 376S–377S (in French) (1998).
8. Samuels, M. E. Saturation of the human phenome. *Curr. Genom.* **11**, 482–499 (2010).
9. Cooper, D. N. *et al.* Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Hum. Mutat.* **31**, 631–655 (2010).
10. Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Rev. Genet.* **12**, 745–755 (2011).
11. Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
Using Freeman–Sheldon syndrome as a proof-of-concept, the authors demonstrate for the first time that WES of a small number of unrelated affected individuals can identify the disease-causing gene.
12. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nature Genet.* **42**, 30–35 (2010).
The identification of *DHODH* as the causative gene for Miller syndrome is the first successful application of WES for the discovery of a gene causing a rare Mendelian disorder.
13. Hoischen, A. *et al. De novo* mutations of *SETBP1* cause Schinzel–Giedion syndrome. *Nature Genet.* **42**, 483–485 (2010).
The first demonstration of WES as a means to identify *de novo* mutations in a rare clinical syndrome.
14. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
15. Abecasis, G. R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
16. Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
17. Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Rev. Genet.* **12**, 628–640 (2011).
18. Sunyaev, S. R. Inferring causality and functional significance of human coding DNA variants. *Hum. Molec. Genet.* **21**, R10–R17 (2012).
19. Schuurs-Hoeijmakers, J. H. *et al.* Mutations in *DDHD2*, encoding an intracellular phospholipase A(1), cause a recessive form of complex hereditary spastic paraplegia. *Am. J. Hum. Genet.* **91**, 1073–1081 (2012).
20. Kalsoom, U. E. *et al.* Whole exome sequencing identified a novel zinc-finger gene *ZNF141* associated with autosomal recessive postaxial polydactyly type A. *J. Med. Genet.* **50**, 47–53 (2012).
21. Sankaran, V. G. *et al.* Exome sequencing identifies *GATA1* mutations resulting in Diamond–Blackfan anemia. *J. Clin. Invest.* **122**, 2439–2443 (2012).
22. Fiskerstrand, T. *et al.* Familial diarrhea syndrome caused by an activating *GUCY2C* mutation. *N. Engl. J. Med.* **366**, 1586–1595 (2012).
23. Gibson, W. T. *et al.* Mutations in *EZH2* cause Weaver syndrome. *Am. J. Hum. Genet.* **90**, 110–118 (2011).
24. Hood, R. L. *et al.* Mutations in *SRCAP*, encoding SNF2-related CREBBP activator protein, cause Floating–Harbor syndrome. *Am. J. Hum. Genet.* **90**, 308–313 (2012).
25. Lindhurst, M. J. *et al.* A mosaic activating mutation in *AKT1* associated with the Proteus syndrome. *N. Engl. J. Med.* **365**, 611–619 (2011).
The first application of WES to discover somatic *de novo* mutations as the cause of a genetic disorder.
26. Riviere, J. B. *et al. De novo* germline and postzygotic mutations in *AKT3, PIK3R2* and *PIK3CA* cause a spectrum of related megalencephaly syndromes. *Nature Genet.* **44**, 934–940 (2012).
27. Robinson, P. N. *et al.* The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am. J. Hum. Genet.* **83**, 610–615 (2008).
28. Fernandez, B. A. *et al.* Adult siblings with homozygous *G6PC3* mutations expand our understanding of the severe congenital neutropenia type 4 (SCN4) phenotype. *BMC Med. Genet.* **13**, 111 (2012).
29. McMillan, H. J. *et al.* Specific combination of compound heterozygous mutations in 17β-hydroxysteroid dehydrogenase type 4 (*HSD17B4*) defines a new subtype of D-bifunctional protein deficiency. *Orphanet J. Rare Dis.* **7**, 90 (2012).

30. Isidor, B. *et al.* Truncating mutations in the last exon of *NOTCH2* cause a rare skeletal disorder with osteoporosis. *Nature Genet.* **43**, 306–308 (2011).
31. Majewski, J. *et al.* Mutations in *NOTCH2* in families with Hajdu–Cheney syndrome. *Hum. Mutat.* **32**, 1114–1117 (2011).
32. Simpson, M. A. *et al.* Mutations in *NOTCH2* cause Hajdu–Cheney syndrome, a disorder of severe and progressive bone loss. *Nature Genet.* **43**, 303–305 (2011).
33. McDaniell, R. *et al. NOTCH2* mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *Am. J. Hum. Genet.* **79**, 169–173 (2006).
34. Lee, J. H. *et al. De novo* somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nature Genet.* **44**, 941–945 (2012).
35. Tsurusaki, Y. *et al.* Mutations affecting components of the SWI/SNF complex cause Coffin–Siris syndrome. *Nature Genet.* **44**, 376–378 (2012).
Mutations in six of 15 genes encoding subunits of a single complex are found to be causal for Coffin–Siris syndrome.
36. Srour, M. *et al.* Mutations in *C5ORF42* cause Joubert syndrome in the French Canadian population. *Am. J. Hum. Genet.* **90**, 693–700 (2012).
37. Beaulieu, C. L. *et al.* A generalizable pre-clinical research approach for orphan disease therapy. *Orphanet J. Rare Dis.* **7**, 39 (2012).
A valuable review which discusses generalizable approaches to the therapeutic configuration for rare diseases.
38. Rilstone, J. J., Alkhater, R. A. & Minassian, B. A. Brain dopamine-serotonin vesicular transport disease and its treatment. *N. Engl. J. Med.* **368**, 543–550 (2012).
A novel disease-causing gene is identified for a rare neurological disease, and the new insight suggests a successful treatment.
39. Green, P. *et al.* Brown–Vialetto–Van Laere syndrome, a ponto-bulbar palsy with deafness, is caused by mutations in *c20orf54*. *Am. J. Hum. Genet.* **86**, 485–489 (2010).
40. Johnson, J. O., Gibbs, J. R., Van Maldergem, L., Houlden, H. & Singleton, A. B. Exome sequencing in Brown–Vialetto–van Laere syndrome. *Am. J. Hum. Genet.* **87**, 567–569 (2010).
41. Johnson, J. O. *et al.* Exome sequencing reveals riboflavin transporter mutations as a cause of motor neuron disease. *Brain* **135**, 2875–2882 (2012).
42. Bosch, A. M. *et al.* The Brown–Vialetto–Van Laere and Fazio Londe syndrome revisited: natural history, genetics, treatment and future perspectives. *Orphanet J. Rare Dis.* **7**, 83 (2012).
43. Shimada, Y. *et al.* Proteasome inhibitors improve the function of mutant lysosomal α-glucosidase in fibroblasts from Pompe disease patient carrying c.546G > T mutation. *Biochem. Biophys. Res. Commun.* **415**, 274–278 (2011).
A therapeutic approach using protein stabilization; mutant α-glucosidase in patients with Pompe disease can be stabilized by treatment with a proteasome inhibitor.
44. Institute of Medicine (US) Committee on Accelerating Rare Diseases Research and Orphan Product Development. *Rare Diseases and Orphan Products: Accelerating Research and Development* (ed. Field, M.J. & Boat, T.F.) (National Academies Press, 2010).
45. Ramsey, B. W. *et al.* A CFTR potentiator in patients with cystic fibrosis and the *G551D* mutation. *N. Engl. J. Med.* **365**, 1663–1672 (2011).
The first effective therapy for a specific and comparatively rare subclass of cystic fibrosis.
46. Freimer, N. & Sabatti, C. The human phenome project. *Nature Genet.* **34**, 15–21 (2003).
47. Oetting, W. S. *et al.* Getting ready for the Human Phenome Project: the 2012 forum of the Human Variome Project. *Hum. Mutat.* **34**, 661–666 (2013).
48. Riggs, E. R. *et al.* Towards an evidence-based process for the clinical interpretation of copy number variation. *Clin. Genet.* **81**, 403–412 (2012).
49. Maddatu, T. P., Grubb, S. C., Bult, C. J. & Bogue, M. A. Mouse Phenome Database (MPD). *Nucleic Acids Res.* **40**, D887–894 (2012).
50. Cheng, K. C., Xin, X., Clark, D. P. & La Riviere, P. Whole-animal imaging, gene function, and the Zebrafish Phenome Project. *Curr. Opin. Genet. Dev.* **21**, 620–629 (2011).
51. Mashimo, T., Voigt, B., Kuramoto, T. & Serikawa, T. Rat Phenome Project: the untapped potential of existing rat strains. *J. Appl. Physiol.* **98**, 371–379 (2005).
52. Serikawa, T. *et al.* National BioResource Project-rat and related activities. *Exp. Anim.* **58**, 333–341 (2009).
53. Hamosh, A. *et al.* PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. *Hum. Mutat.* **34**, 566–571 (2013).
54. Girdea, M. *et al.* PhenoTips: patient phenotyping software for clinical and research use. *Hum. Mut.* **34**, 1057–1065 (2013).
55. Kingsmore, S. F. & Saunders, C. J. Deep sequencing of patient genomes for disease diagnosis: when will it become routine? *Sci. Transl. Med.* **3**, 87ps23 (2011).
56. Katsanis, S. H. & Katsanis, N. Molecular genetic testing and the future of clinical genomics. *Nature Rev. Genet.* **14**, 415–426 (2013).
57. Bell, C. J. *et al.* Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med.* **3**, 65ra4 (2011).
58. Berg, J. S., Khoury, M. J. & Evans, J. P. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. *Genet. Med.* **13**, 499–504 (2011).
59. Green, R. C. *et al.* Exploring concordance and discordance for return of incidental findings from clinical sequencing. *Genet. Med.* **14**, 405–410 (2012).
A valuable study providing recommendations about which specific conditions and types of genetic variants should be returned as incidental findings in clinical sequencing.
60. Bettens, K., Sleegers, K. & Van Broeckhoven, C. Genetic insights in Alzheimer's disease. *Lancet Neurol.* **12**, 92–104 (2013).
61. Wu, C. H. *et al.* Mutations in the *profilin 1* gene cause familial amyotrophic lateral sclerosis. *Nature* **488**, 499–503 (2012).
62. Leonard, H. & Wen, X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment. Retard. Dev. Disabil. Res. Rev.* **8**, 117–134 (2002).
63. Elsabbagh, M. *et al.* Global prevalence of autism and other pervasive developmental disorders. *Autism Res.* **5**, 160–179 (2012).
64. Veltman, J. A. & Brunner, H. G. *De novo* mutations in human genetic disease. *Nature Rev. Genet.* **13**, 565–575 (2012).
65. Vissers, L. E. *et al.* A *de novo* paradigm for mental retardation. *Nature Genet.* **42**, 1109–1112 (2010).
The first study to identify *de novo* mutations as causative for intellectual disability using WES of unaffected parents–affected child trios.
66. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
67. Sanders, S. J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
68. Neale, B. M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
69. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
70. Iossifov, I. *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
71. Farooq, F., Balabanian, S., Liu, X., Holcik, M. & MacKenzie, A. p38 mitogen-activated protein kinase stabilizes SMN mRNA through RNA binding protein HuR. *Hum. Mol. Genet.* **18**, 4035–4045 (2009).
72. Farooq, F. *et al.* Prolactin increases *SMN* expression and survival in a mouse model of severe spinal muscular atrophy via the STAT5 pathway. *J. Clin. Invest.* **121**, 3042–3050 (2011).
73. Akhurst, R. J. & Hata, A. Targeting the TGFβ signalling pathway in disease. *Nature Rev. Drug Discov.* **11**, 790–811 (2012).
74. MacKenzie, A. & Boycott, K. MEK inhibition in *BRAF*-mutated melanoma. *N. Engl. J. Med.* **367**, 1364–1365 (2012).
75. Montiel-Equihua, C. A., Thrasher, A. J. & Gaspar, H. B. Gene therapy for severe combined immunodeficiency due to adenosine deaminase deficiency. *Curr. Gene Ther.* **12**, 57–65 (2012).
76. Prasad, V. K. & Kurtzberg, J. Transplant outcomes in mucopolysaccharidoses. *Semin. Hematol.* **47**, 59–69 (2009).
77. Nakamura, A. & Takeda, S. Exon-skipping therapy for Duchenne muscular dystrophy. *Neuropathology* **29**, 494–501 (2009).
78. Axelrod, F. B. *et al.* Kinetin improves *IKBKAP* mRNA splicing in patients with familial dysautonomia. *Pediatr. Res.* **70**, 480–483 (2009).
79. Ohashi, T. Enzyme replacement therapy for lysosomal storage diseases. *Pediatr. Endocrinol. Rev.* **10**, 26–34 (2012).
80. Malik, V. *et al.* Gentamicin-induced readthrough of stop codons in Duchenne muscular dystrophy. *Ann. Neurol.* **67**, 771–780 (2010).
81. de Lartigue, J. Tafamidis for transthyretin amyloidosis. *Drugs Today (Barc.)* **48**, 331–337 (2012).
82. Wheeler, T. M. *et al.* Targeting nuclear RNA for *in vivo* correction of myotonic dystrophy. *Nature* **488**, 111–115 (2012).
83. Mantha, N., Das, S. K. & Das, N. G. RNAi-based therapies for Huntington's disease: delivery challenges and opportunities. *Ther. Deliv.* **3**, 1061–1076 (2012).
84. Wu, X. *et al.* MEK-ERK pathway modulation ameliorates disease phenotypes in a mouse model of Noonan syndrome associated with the *Raf1*^{L613V} mutation. *J. Clin. Invest.* **121**, 1009–1025 (2011).

Acknowledgements
The authors thank the Finding of Rare Disease Genes (FORGE) Canada Consortium for providing the opportunity to explore the underpinnings of rare diseases with a spirit of national collaboration. They also thank J. Schwartzentruber, J. Majewski and C. Beaulieu for their insight and counsel into the best practices for the discovery of rare-disease-causing genes.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
1000 Genomes Project: http://www.1000genomes.org
SNP database (dbSNP): http://www.ncbi.nlm.nih.gov/SNP
EURODIS: http://www.eurordis.org
Exome Variant Server: http://evs.gs.washington.edu/EVS
FORGE Canada Consortium: http://www.care4rare.ca
Genetic Alliance: http://www.geneticalliance.org
International HapMap Project: http://hapmap.ncbi.nlm.nih.gov
Human Phenotype Ontology: http://www.human-phenotype-ontology.org
Human Variome Project: http://www.humanvariomeproject.org
International Rare Disease Research Consortium (IRDiRC): http://www.irdirc.org
International Knockout Mouse Consortium: http://www.knockoutmouse.org
Online Mendelian Inheritance in Man (OMIM): http://omim.org
Orphanet: http://www.orpha.net
PhenoDB: http://phenodb.net
PhenoTips: http://phenotips.org

SUPPLEMENTARY INFORMATION
See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF