TRANSLATIONAL GENETICS

Paediatric genomics: diagnosing rare disease in children

Caroline F. Wright[1], David R. FitzPatrick[2] and Helen V. Firth[3,4]

Abstract | The majority of rare diseases affect children, most of whom have an underlying genetic cause for their condition. However, making a molecular diagnosis with current technologies and knowledge is often still a challenge. Paediatric genomics is an immature but rapidly evolving field that tackles this issue by incorporating next-generation sequencing technologies, especially whole-exome sequencing and whole-genome sequencing, into research and clinical workflows. This complex multidisciplinary approach, coupled with the increasing availability of population genetic variation data, has already resulted in an increased discovery rate of causative genes and in improved diagnosis of rare paediatric disease. Importantly, for affected families, a better understanding of the genetic basis of rare disease translates to more accurate prognosis, management, surveillance and genetic advice; stimulates research into new therapies; and enables provision of better support.

Developmental disorders
Diseases with their genesis in embryonic life or early fetal brain development.

[1]*University of Exeter Medical School, Institute of Biomedical and Clinical Science, Royal Devon and Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK.*
[2]*Medical Research Council Human Genetics Unit, The Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.*
[3]*Cambridge University Hospitals National Health Service Foundation Trust, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0QQ, UK.*
[4]*Wellcome Trust Sanger Institute, Wellcome Genome Campus, Cambridge, Hinxton CB10 1SA, UK.*

Correspondence to H.V.F.
hvf21@cam.ac.uk

doi:10.1038/nrg.2017.116
Published online 5 Feb 2018; corrected online 19th Feb 2018

Rare diseases are defined as those that affect fewer than 1 in 2,000 individuals in the general population[1]. These diseases remain rare in the general population because they usually adversely affect reproductive fitness[2,3]. There are estimated to be 7,000 rare diseases, ~80% of which are thought to have a genetic cause[4,5]. The majority (50–75%) of rare diseases affect children[1], and many are severe multisystem disorders with a range of phenotypes (FIG. 1). Collectively, they are responsible for 35% of deaths in the first year of life and are a significant cause of paediatric hospital admissions[6,7]; one-third of children born with a rare disease will not live to see their fifth birthday[1].

Although rare variants in nearly 1,500 genes have now been shown to cause developmental disorders, many more disease-causing genes still remain to be discovered[8,9]. Accurate diagnosis — here defined as discerning the precise molecular cause (genotype) that explains the clinical features of the rare disease (phenotype) — is the cornerstone of safe medical practice. For children with a rare genetic disorder, a robust genetic diagnosis unlocks access to a wealth of information in the literature that provides advice on management and therapy and also enables access to disorder-specific support groups, which reduces isolation for families affected by a rare disorder. A robust genetic diagnosis also enables accurate determination of risk to existing and future family members. However, finding a diagnosis for each individual remains a considerable challenge because of the genetic and phenotypic variability associated with these diseases and our incomplete knowledge. Addressing this challenge will fulfil the new vision of the International Rare Diseases Research Consortium (IRDiRC) for 2017–2027: to "enable all people living with a rare disease to receive an accurate diagnosis, care and available therapy within one year of coming to medical attention" (REF. 10).

Causative genetic variants can range in size from the substitution, deletion or duplication of a single base pair, to structural variants and to altered copy numbers of an entire chromosome (aneuploidy) or genome (as in diploid/triploid mosaicism). Not only are the diseases caused by these variants individually very rare or ultra-rare (<1 in 100,000) but there is also usually clinical variability in the penetrance of the disorder and in the expressivity of individual features within disease entities. A proportion of this variability in genetic diseases can be attributed to locus heterogeneity and allelic heterogeneity, which can be extreme; for example, the clinical feature 'intellectual disability' may be caused by monoallelic or biallelic variants in one or more of >700 different genes[11,12]. Other important sources of variability include genetic variants at one or more other loci (modifiers) and environmental factors, which are currently difficult to identify and quantify. As a consequence, the majority of rare genetic diseases do not have highly distinctive clinical presentations and are intractable to clinical diagnosis by classical approaches, which have typically relied on the recognition of a consistent pattern of clinical features, such as a combination of growth anomalies, site-specific malformations and a characteristic facial appearance. Moreover, not all rare paediatric disorders

Structural variants
Blocks of DNA > 1 kb that differ relative to the reference genome or general population, including inversions, balanced translocations and copy number variants (for example, deletions and duplications).

Mosaicism
The presence of two or more populations of cells with different genotypes in an individual who has developed from a single fertilized egg.

are genetic in origin, and worldwide, there remains a small but important contribution from teratogen exposure (for example, fetal alcohol syndrome, congenital toxoplasmosis or Zika virus)[13–15]. Thus, until recently, most children with rare disease, especially those with developmental disorders, have remained undiagnosed[16].

Next-generation sequencing (NGS) technologies have hugely improved the prospects of obtaining a genetic diagnosis because they are genetically and phenotypically agnostic in their approach. Simultaneous sequencing of every gene in the genome detects the vast majority of important variation present in an individual's genome and potentially enables diagnoses to be made across a range of paediatric phenotypes (FIG. 1), even for rare and ultra-rare conditions, including those with which the child's clinician may be unfamiliar[17]. Once the genetic basis of a syndrome is identified, the range of clinical presentations often increases (known as syndrome expansion) as milder or incomplete combinations of features are recognized to share the same molecular genetic basis[18]. Advances in genomic technologies have had a particular impact on rare paediatric disease owing to its early onset and the differing nature and frequency of causal variants compared with rare adult genetic disease, which means that the diagnostic yield is generally higher[19]. In severe developmental disorders, the reproductive fitness will be very low; without strong balancing selection — which is rare — or socially and/or geographically constrained mating — which is common — such variants will not be maintained in a population[20]. Thus, the prior expectation in severe paediatric genetic disease is that the causal variants will be ultra-rare and often either *de novo* (with a dominant inheritance pattern) or enriched in consanguineous or founder populations (with a recessive inheritance pattern)[21–23]. Although human genomes vary from each other at around 4–5 million locations, the vast majority of this variation is both common and benign[24–26]. Finding the primary genetic diagnosis in children is therefore simplified by the fact that such

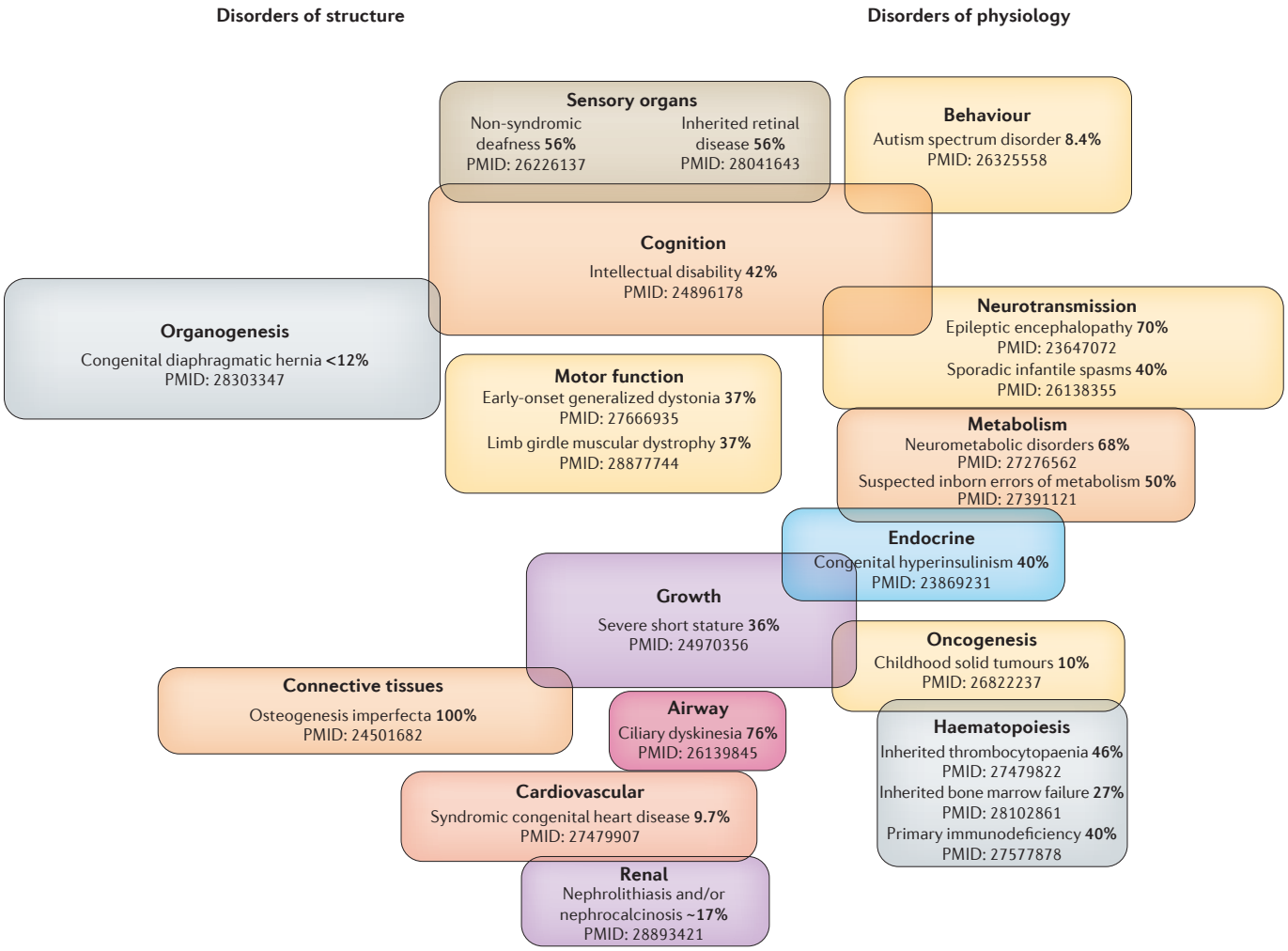


Figure 1 | **Available diagnostic rates based on whole-exome sequencing in classes of paediatric genomic diseases.** The diagnostic rate of whole-exome sequencing is given for each phenotype class, with the associated PubMed identification (PMID) number; only case series ≥10 were used, and only pathogenic or likely pathogenic variants in probable or known disease genes are included. There were remarkably few studies available in the published literature for many of the broad classes of disease with the exception of neurodevelopmental disorders. Box sizes are approximately proportionate to the prevalence of the phenotypes in paediatric practice.

rare damaging *de novo* variants (and biallelic variant combinations) are unlikely to be present in adult control populations.

This Review focuses primarily on the application of genomics to paediatric rare disease diagnosis; it does not discuss somatic variation in cancer or the application of genomics to the diagnosis or management of paediatric cancer. We begin by outlining the benefits of moving from traditional genetic diagnostic testing to NGS. Next, we highlight some of the issues around data analysis and consider the current primary clinical indications for whole-exome sequencing (WES) or whole-genome sequencing (WGS). Finally, we discuss the future of paediatric genomics in terms of improved workflows, wider applications and treatment opportunities.

From genetic tests to genomic tests

Traditional genetic testing. For most of its history, clinical genetics has relied upon two types of genetic tests: highly focused high-resolution molecular single-gene tests and low-resolution genome-wide cytogenetic tests[27] (FIG. 2a). In single-gene molecular tests, a particular gene is selected for Sanger sequencing or genotyping according to the clinical presentation of the patient. The probability of the test yielding a correct diagnosis is dependent on the diagnostician correctly identifying the underlying genetic cause of the condition and selecting the correct test. Thus, single-gene molecular tests are best suited for diagnosing highly distinct clinical conditions that are caused by just one or a few genes. For instance, most 'classic' inherited genetic conditions, such as cystic fibrosis (caused by variants in *CFTR*)[28] or Duchenne muscular dystrophy (caused by variants in *DMD*)[29], are diagnosed using this approach. By contrast, a low-resolution whole-genome approach, such as a G-banded microscopy-based karyotype (resolution ~5–7 Mb), can be used to diagnose common trisomies and segmental aneuploidy together with some apparently sporadic developmental disorders that are caused by rare recurrent or unique large imbalances of chromosomal material. Smaller structural variants, such as copy number variants (CNVs), can be detected by molecular karyotyping using genomic microarrays, which typically have a resolution of ~50–100 kb[30]. Genomic microarrays can detect disease-causing CNVs at any position in the genome, including recurrent variants associated with microdeletion and/or microduplication syndromes[30–32]. However, their sensitivity is low compared with NGS, and only ~10% of patients with a rare paediatric disease can be diagnosed using these approaches[33].

Modern genetic testing. NGS technology enables massively parallel sequencing of multiple genes[34–36], which has revolutionized clinical genetics[37–39]. Rather than selecting a single gene to test or a single type of variant to detect, clinicians can now request simultaneous sequencing of multiple genes or even entire exomes or genomes[40]. This high-resolution whole-genome approach combines the benefits of both individual molecular tests and genome-wide cytogenetic tests (FIG. 2a). Many paediatric conditions are caused by a genetic variant in one of numerous different genes, so the ability to sequence multiple genes simultaneously enables fast and comprehensive genetic analysis. For example, Bardet–Biedl syndrome can be caused by variants in over 20 genes with indistinguishable clinical presentations[41]; similarly, there are many causes of congenital sensorineural hearing impairment[42] that are clinically indistinguishable but genetically distinct. In such cases, simultaneous testing of all known genetic causes improves and expedites the diagnostic process.

NGS is particularly useful for diagnosing rare developmental disorders that might be caused by single nucleotide variants (SNVs) or small insertions and/or deletions (indels) in any one of thousands of genes in the genome where testing each candidate gene individually is no longer a feasible approach (FIG. 2b). This approach is exemplified by the application of NGS to the genetic diagnosis of infantile epileptic encephalopathy, a heterogeneous condition that can be caused by variants in any one of dozens of genes, where NGS-driven discovery has identified 31 novel genes between 2012 and 2015 (REFS 43,44). The sensitivity of NGS technologies to detect SNVs and indels is extremely high[45], and although Sanger sequencing is still considered to be the gold standard in many laboratories, NGS technologies may actually prove superior, particularly for detecting heterozygous changes, in which the variant of interest is present in just one copy of a chromosome pair, and mosaic variants that are only present in a subset of cells[46].

NGS data can also be used for detecting CNVs and other structural variants, although their analytical validity is currently lagging substantially behind that of microarrays[47–52]. Nonetheless, optimized gene panels sequenced at high depth using NGS have proved useful for detecting small exon deletions (<10 kb) missed by low-resolution microarrays, and both WES and WGS have been successfully used to detect large CNVs (>500 kb)[49]. The question of whether to undertake a genomic microarray before WES or WGS for children with developmental disorders is currently unresolved because the sensitivity and specificity of the respective tests depend on the precise characteristics of the test selected, the analysis pipeline deployed and the diagnostic yield for different sizes of CNVs.

Although NGS technologies allow every gene — and every *cis*-regulatory element (CRE) — in the genome to be sequenced, clinical reports in paediatrics usually focus on variants in genes known to be associated with rare childhood-onset disorders. However, ~70% of genes and almost all CREs in the genome still have no known function in human health and development[2,3], and so many variants that could be causal for disease will not typically be reported because they occur in genes of unknown relevance at the time of analysis. Moreover, short-read NGS does not reliably detect triplet repeat expansions, which are involved in paediatric diseases such as fragile X syndrome, congenital myotonic dystrophy and Friedreich ataxia[53]. It is also often poor at detecting small CNVs, such as exon deletions or

Penetrance
The proportion of individuals with a particular genotype who show features of the condition (however mildly). If some individuals with the genotype never show any features, the condition is said to have incomplete (or reduced) penetrance. If features develop with age, the condition is said to have age-dependent penetrance.

Expressivity
The phenotypic variability and severity that a given genotype shows in individuals penetrant for the condition.

Locus heterogeneity
When variants in a number of different genes independently cause the same phenotype.

Allelic heterogeneity
When different mutations at the same locus cause the same phenotype.

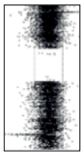
Monoallelic
Describes a mutation that affects only one copy of a gene. Autosomal dominant, *de novo* dominant or X-linked disorders are caused by a monoallelic pathogenic variant.

Biallelic variants
Mutations that affect both copies of a gene. Autosomal recessive disorders are caused by pathogenic biallelic variants.

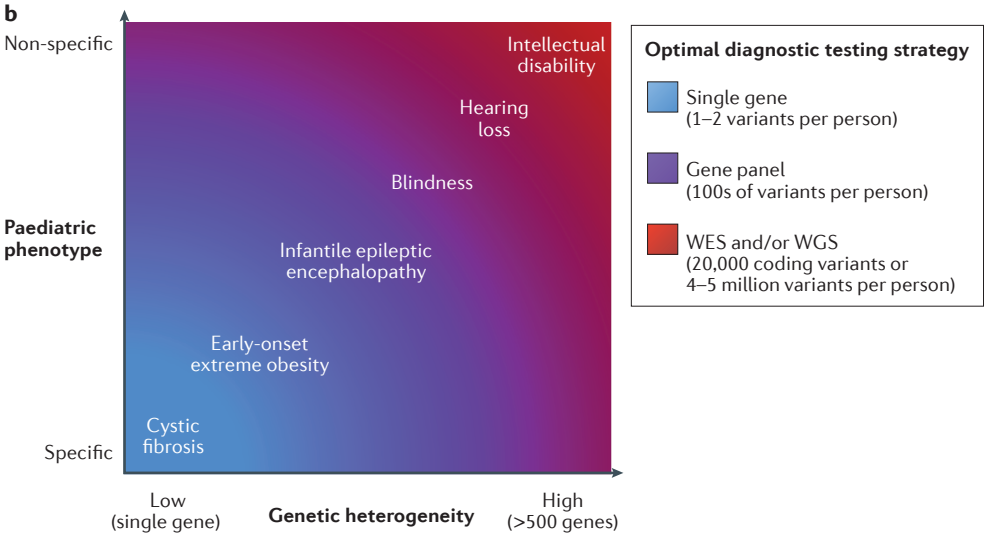
Whole-exome sequencing
(WES). Next-generation sequencing of the entire protein-coding portion of the genome. In humans, the total length of coding and splicing regions is estimated to be ~35 Mb and comprises ~20,000 genes (1–2% of the genome).

Whole-genome sequencing
(WGS). Next-generation sequencing of the entire genome, which, in humans, is typically ~3,000 Mb.

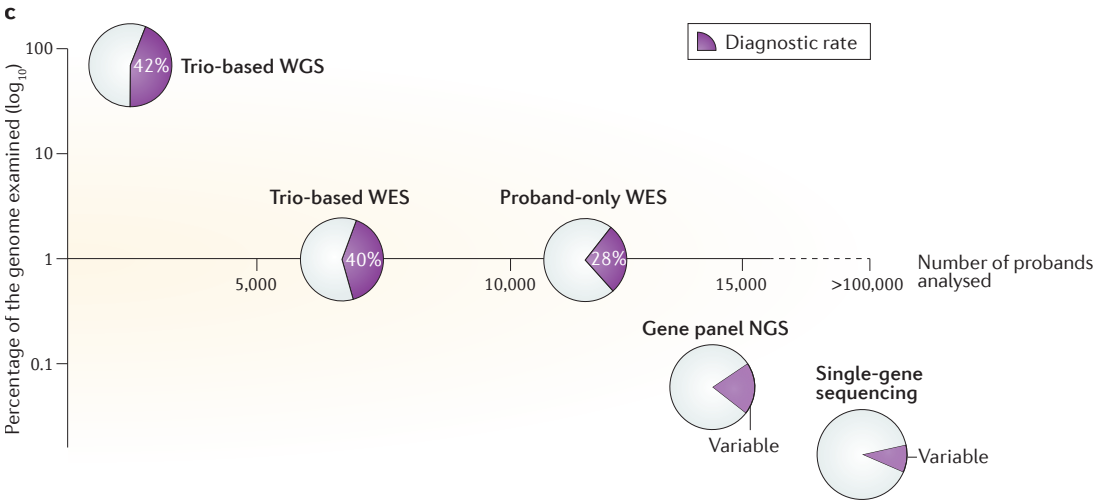
a

	Light microscope	G-banded karyotype	Microarray	Whole-exome sequence	Whole-genome sequence
Appearance				CGGATGATTACCCGTT G.....GCTC TAGCTAGCTATA....	CGGATGATTACCCGTT GATATAGCTCTCGCTC GCTCTAGCTAGCTATA GGCTATGGGTGGGGGC
Resolution	Entire chromosome	5–10 Mb	50–100 kb	1 bp	1 bp
Number of loci probed	N/A	~500	~0.05–2 million	~50 million	3 billion
Variants detected	Aneuploidy, polyploidy	Variants >5 Mb	Copy number variants	Coding regions	Majority of variants
Variants per person	0 or 1	0 or 1	10–100s	~20,000	4–5 million
Diagnostic yield	Low →				High
Incidental findings	Low →				High

b



c



Single-gene tests
Approaches that enable detailed analysis of a single gene. In addition to sequence analysis, they usually also include an assessment of dosage in order to detect exon-level deletions and duplications, which are often difficult to detect with current approaches to whole-exome sequencing and whole-genome sequencing.

Cytogenetic tests
Genome-wide tests that analyse the number and structure of chromosomes, including copy number variants, but do not provide information about the DNA sequence.

Karyotype
The chromosomal complement of a cell. Large-scale chromosomal imbalances can be detected using karyotyping approaches, such as imaging Giemsa-banded chromosomes with light microscopy.

Copy number variants
(CNVs). Structural variants that involve either a deletion or a duplication of a section of DNA relative to the reference genome.

Single nucleotide variants
(SNVs). Differences within a population, or between an individual and a reference genome, that affect a single base pair of DNA.

Gene panels
Subsets of genes (usually linked to a particular phenotype) that are incorporated into a laboratory-based gene capture kit or that form the basis of computer-based virtual gene panels, which are applied to a subset of variant data from a clinical exome or a whole-exome sequencing and/or whole-genome sequencing assay.

◀ Figure 2 | **Genomic testing strategies and clinical heterogeneity. a** | Genome-wide assays used in clinical genetics have developed from traditional methods for visualizing chromosomes using a light microscope to assaying copy number variation across the genome and to sequencing the entire genome. As the resolution of the test increases, the number of detectable variants also increases. Although this increase in the number of variants leads to an increase in diagnostic yield across a range of conditions, it also substantially increases the likelihood of detecting incidental findings and variants of uncertain clinical significance. **b** | The less specific the phenotype associated with a disease is, the more likely it is to be caused by variants in a large number of individual genes; that is, genetic heterogeneity increases as phenotypic specificity decreases. Testing a single gene or small number of genes may be preferable if the disease is phenotypically and/or genetically very homogeneous. However, for phenotypically and/or genetically heterogeneous conditions, many hundreds of genes may need to be tested, which will generate thousands of variants for every person. Whole-exome sequencing (WES) and whole-genome sequencing (WGS) are the most phenotypically agnostic assays and can be used to diagnose a wide range of disorders. **c** | There is a balance between the diagnostic potential of a sequencing strategy and its practicality and cost. In this Figure, the number of currently published probands analysed using different next-generation sequencing (NGS) techniques is plotted on the x-axis, and the percentage of the genome examined is plotted as a logarithmic scale on the y-axis. The diagnostic rate was obtained from the published literature. Trio-based WGS has the highest diagnostic yield but is also the most informatically demanding and expensive approach; hence, fewer studies have been reported to date. Given that most currently known pathogenic variants are in the coding portion of the genome, using WES only slightly lowers the diagnostic yield in severe intellectual disability (from ~42% for WGS to ~40% for WES) but hugely reduces the cost. On the other hand, although moving from a family trio-based approach to a proband-only approach has practical and financial advantages, it substantially reduces the diagnostic yield (to ~28%) because *de novo* status or phase cannot be directly assigned to observed genomic variants to determine if they are on the same or different chromosomes. NGS of gene panels or of single genes is the most common approach, but the rate of diagnosis varies considerably depending on phenotype and patient ascertainment, as this determines the prior likelihood of a detectable genetic cause for the clinical presentation, and on the proportion of patients with a given phenotype that can be explained by known disease genes (see FIG. 1). N/A, not applicable. Part **b** is adapted from the *Journal of Medical Genetics*, Boycott, K. *et al.*, **52**, 431–437 (2015) with permission from BMJ publishing group Ltd. (REF. 62). Part **c** is adapted from REF. 167, Macmillan Publishers Limited.

duplications, which are a major cause of disease in some conditions such as Duchenne muscular dystrophy[54] and spinal muscular atrophy (SMA) type 1 (REF. 55). Post-zygotic mosaic variants are also often missed in low coverage data in cases where they are the *de novo* cause of disease in the affected child or are inherited from an unaffected mosaic parent[56,57]. It is therefore crucial to consider what data may be missing from each report that is reviewed and to determine whether the assay and/or analysis covers the relevant gene or genes and detects the relevant variant type typically associated with disease.

Diagnostic tests based on next-generation sequencing. There are a number of ways that NGS can be used for clinical diagnostics, which vary in the number and type of target regions that are sequenced. Targeted approaches include sequencing whole single genes, for example, in cases where individual variants would previously have been genotyped; sequencing panels of disease-specific genes, with panels ranging in size from 2 to >2,000 genes; and sequencing all exons of the ~4,000 genes currently associated with monogenic disease, which are also known as the clinical exome or Mendeliome[58]. By contrast, sequencing all ~20,000 protein-coding genes by WES and entire genomes by WGS are essentially non-targeted tests[40]. All NGS-based diagnostic approaches generate large amounts of data, but the difference in scale between tests can be vast, ranging from a few hundred base pairs of DNA for a small gene, where only a handful of variants will be detected, to 3 billion base pairs for a whole genome containing 4–5 million variants per person (FIG. 2c). As more genes are tested in parallel and more data are generated, the sensitivity of the test increases, but its specificity decreases, and both the logistical and ethical challenges increase (BOX 1). In particular, the issue of how to deal with unsolicited secondary findings is extremely contentious because of the enormous number of variants in every genome and their breadth of applicability. WES or WGS offers the opportunity to screen genomes for a range of conditions, but legitimate concerns about overdiagnosis have resulted in alternative approaches being taken in different laboratories and across different jurisdictions[59–63].

Where a virtual gene panel approach is applied to analyse WES or WGS data, a balance must be struck between limiting the panel to a highly specific, conservative set of genes that are directly related to the patient's particular phenotype and including a broader set of genes that are associated with a number of related conditions (for instance, limiting testing to epilepsy-related genes versus including all genes associated with neurodevelopmental disorders). Using a limited gene panel is clinically attractive because it reduces off-target noise and incidental findings, but it can result in missed diagnoses owing to genetic heterogeneity. However, the number of benign genetic variants present in every individual and the narrative potential of every genome[64] means that the probability of over-interpretation leading to overdiagnosis is substantial, and it increases with the scale of the test. It is not yet clear how best to tune genomic analysis to maximize identification of true diagnostic variants and minimize identification of misleading variants. The sensitivity and specificity of particular assays are rarely documented or compared, as such outputs are dependent on many factors, including patient ascertainment, the genomic footprint of the phenotype in question and the availability of comparable data from unaffected individuals. Test data sets that would enable comparisons between assays are not routinely used in clinical practice.

Proband-only and trio-based diagnostic tests. One excellent strategy to manage the wealth of genetic variants identified by WES or WGS is to use a family trio analysis of samples from the child and both of their biological parents. This enables rare benign familial variants to be filtered out, *de novo* variants that are present only in the child to be easily identified, and the phase of variants in recessive or imprinted disorders to be established by inheritance. For families in which neither parent is affected by the same disorder as the child, sequencing of parent–child trios rather than the child alone (proband analysis) offers an approximately tenfold reduction in the number of candidate variants as well as a 50% increase in diagnostic yield

Box 1 | **Ethical, legal and social implications of paediatric genomics**

Paediatric genomics has many of the same ethical, legal and social issues that clinical genetics has been dealing with for decades, such as reproductive autonomy, informed consent for research, misattributed parentage and implications for family members[141–146]. However, the enormity and breadth of genomic data raise their own novel issues, which are more complicated for paediatric testing; the reduced capacity of the child to consent to testing and/or research means that parents and clinicians have an increased role in deciding what may be in the best interests of the child[147–150]. Most of the novel ethical issues in the era of genomics relate to the storage, interpretation and access of data.

Data storage. How or where clinical next-generation sequencing (NGS) data should be stored is unclear because of their size and the computer power required for their analysis. Furthermore, it is not clear who should have access to that data and when they should be allowed access to it. For instance, should medical records contain just the diagnostic variant, all potentially clinically relevant variants or all detected variants, or should they contain whole-exome or whole-genome sequences? If parental genomes have been sequenced as part of a family trio approach to paediatric diagnosis, how and where should family data be stored? Should access be limited to clinicians involved in the direct care of the family or opened to researchers in industry and/or academia? Although there may be clinical benefits to an individual having a genome sequence available for repeated analysis in years to come, the ongoing improvements in sequencing technologies mean that it may actually be cheaper, easier and more informative to simply re-sequence an individual's genome if it is clinically indicated than it would be to store it.

Confidentiality versus data access. After the initial results from NGS testing, a delicate balance must be struck between protecting individual patient confidentiality and the need to share data widely to improve interpretation (either for the benefit of that individual patient or other patients). This balance is even more complex in paediatrics, where parents are often asked to make decisions about their child's data that may have irreversible repercussions. Should these decisions be revisited when the child approaches and passes the age of majority? The question of who has access to the data relates not only to the individuals, such as family members, clinicians and researchers, among others, but also to the extent of data available to them. Attitudinal data suggest that parents are more concerned about future risk to their child from data-sharing than adult patients[151]. Access could be limited to specific diagnostic variants or a range of clinically relevant results or could include the entire sequence. We have previously proposed a proportionate approach that balances the depth of data against the breadth of sharing[152]. This approach includes sharing a limited number of potentially relevant variants completely openly to maximize the potential benefits for diagnosis while minimizing the plausible risks of re-identification. Although it is important to share individual whole-exome or whole-genome data too, the potential for re-identification is substantially higher, so databases with strictly managed access are often used[153]. Sharing anonymized, sparse phenotypic and genotypic data openly may increase the opportunity not just for diagnosis and discovery but also for patients with very rare conditions to participate in research or trials of new therapies.

Duty of care. For clinicians, the issue of data access is linked to the question of whether their duty of care is limited to finding a diagnosis for the child's immediate problems or whether it extends beyond the scope of the initial investigation[154]. This duty of care could extend to regular re-analysis of the data as new knowledge comes to light, potentially finding a diagnosis many years after the initial sequencing was performed or redacting misdiagnoses in light of new evidence. Although some studies have pioneered an approach to iterative reporting using a regularly updated virtual gene panel[155], which is likely to yield substantial diagnostic benefits in paediatrics because of the current rate of gene discovery, routine re-analysis of genomic data is hampered in the clinic by limited resources for updating and re-analysing data as well as for re-contacting patients[156].

The duty of care could also extend to looking for incidental predispositions to adult-onset conditions or to adverse drug reactions either in the child or their parents[157–160]. The practice of actively searching for additional 'looked-for' findings is currently recommended by the American College of Medical Genetics and Genomics, albeit being limited to pathogenic variants in 59 genes[60,161]. In general, investigating children for adult-onset conditions for which there is no early treatment is not recommended[162–164], and to date, British and European genetics societies have been cautious about extending the scope of genomic analyses beyond finding a diagnosis[165]. Unfortunately, it is often difficult to interpret likely pathogenic variants in asymptomatic families because most disease genes have been found in the context of individuals or families with a particular condition; as a result, the benefits and harms of opportunistic genome screening are currently unknown[166].

(FIG. 2c), thus substantially increasing the speed and likelihood of reaching an accurate diagnosis[8]. This increased analytical power needs to be weighed against the additional cost of parental sequencing, but the increased cost is often offset by the reduced requirement for segregation studies of candidate variants that may be necessary in proband-only approaches. Trio WES for severe developmental disorders that were undiagnosable using conventional genetic testing currently has a diagnostic rate of ~40% (REF. 65), taking the overall diagnostic rate across all genetic testing modalities to over 50% for these diverse and previously intractable disorders.

Integrating and interpreting data

Over the past decade, it has become increasingly clear that an individual genome cannot be interpreted in isolation. Knowledge of the background variation in a population is required to filter common variants. Ideally, this information would be derived from randomly sampled individuals who have not been selected for any specific clinical or social characteristic, and many thousands of such genomes are needed to improve variant calling, annotation and interpretation. Optimal variant classification also requires detailed and comprehensive clinical information about the individual being tested, rather than a simple diagnostic label. The scale of WES and WGS data necessitates automation of variant filtering to focus on the genomic intervals and/or variant characteristics relevant to the clinical question. Analytical workflows for the diagnosis of rare disease are not yet fully standardized[8,66] because the component workflows for processing raw data, identifying pathogenic variants and integrating clinical data to achieve a robust genetic diagnosis (FIG. 3) are complex and still being developed, and because each step has its own problems that require individualized evaluation.

Variant calling and annotation. Variant calling and annotation involve highly complex processes, and the

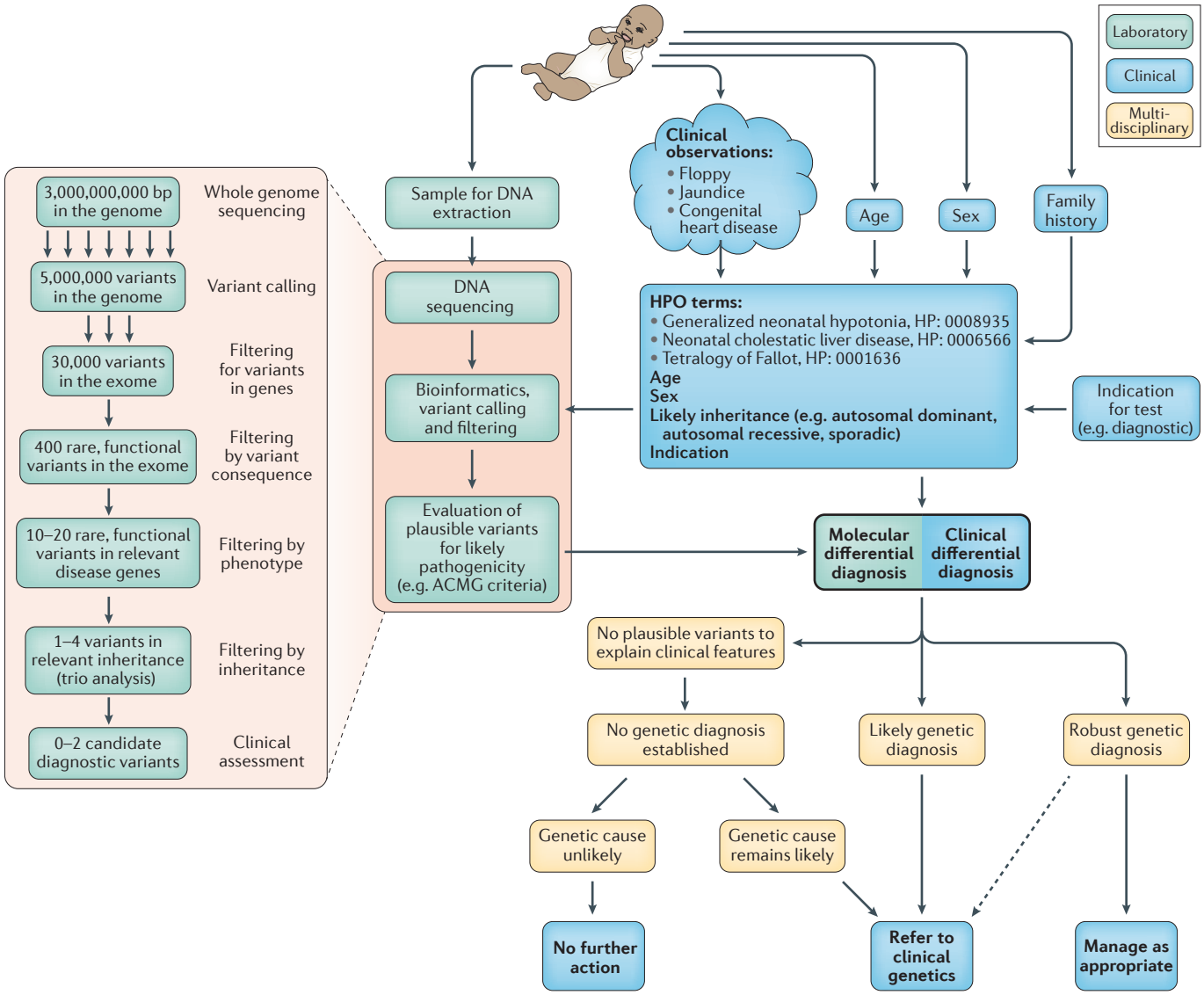


Figure 3 | **Integration of clinical and laboratory workflows to optimize rare disease diagnosis using next-generation sequencing.** The safe and effective implementation of a paediatric genomics service requires diligent design to optimize the expertise of clinicians, bioinformaticians and scientists in various parts of the workflow. Each step of the variant filtering process merits careful thought and evaluation, as the diagnostic variant or variants could be inadvertently discarded at any step, yet stringent filtering is necessary to reduce the number of variants to a sufficiently low number to enable expert review. Typically, variants are filtered out if they have a minor allele frequency of >1% or even >0.1%. Note that diagnostic variants may be missed if an incorrect transcript is chosen for a gene, the wrong inheritance mode is selected in a trio analysis (for example, a mildly affected parent is coded as unaffected) or the variant type responsible for the rare disease is not well captured by the sequencing and analysis pipeline (for instance, a triplet repeat expansion disorder such as fragile X). Clinical assessment of the patient takes place at the start and end of the diagnostic process. At the outset, assessment is used to determine the prior likelihood that the patient has a tractable genetic basis for their clinical presentation and to capture the phenotype and establish a clinical differential diagnosis. At the end of the process, the clinician should determine whether the candidate molecular diagnosis identified by the laboratory constitutes a robust genetic diagnosis (either in full or in part) and whether the confidence in this assertion is sufficient to guide future management of the patient and enable testing for relatives or pre-implantation or prenatal diagnosis. In difficult cases, rare candidate variants are often discussed in a multidisciplinary setting with expert clinicians and clinical scientists (see FIG. 4). ACMG, American College of Medical Genetics and Genomics; HPO, Human Phenotype Ontology.

reproducibility for repeat analysis of the same WGS sample using the same pipeline is generally less than 100% (REF. 47). Variant calling is used to identify allelic differences in the base composition of an individual by comparing each position represented in the data to the equivalent position in a reference human genome. The sensitivity of calling is dependent on the quality of both the test data and reference data and on the algorithms employed. Low-quality or low-coverage data can result in important variants being absent from the output. Variant calling, particularly of small indels, is difficult in repetitive sequences[67], which can lead to diagnoses being missed despite the variant being present in the raw data. Variant call data are generally

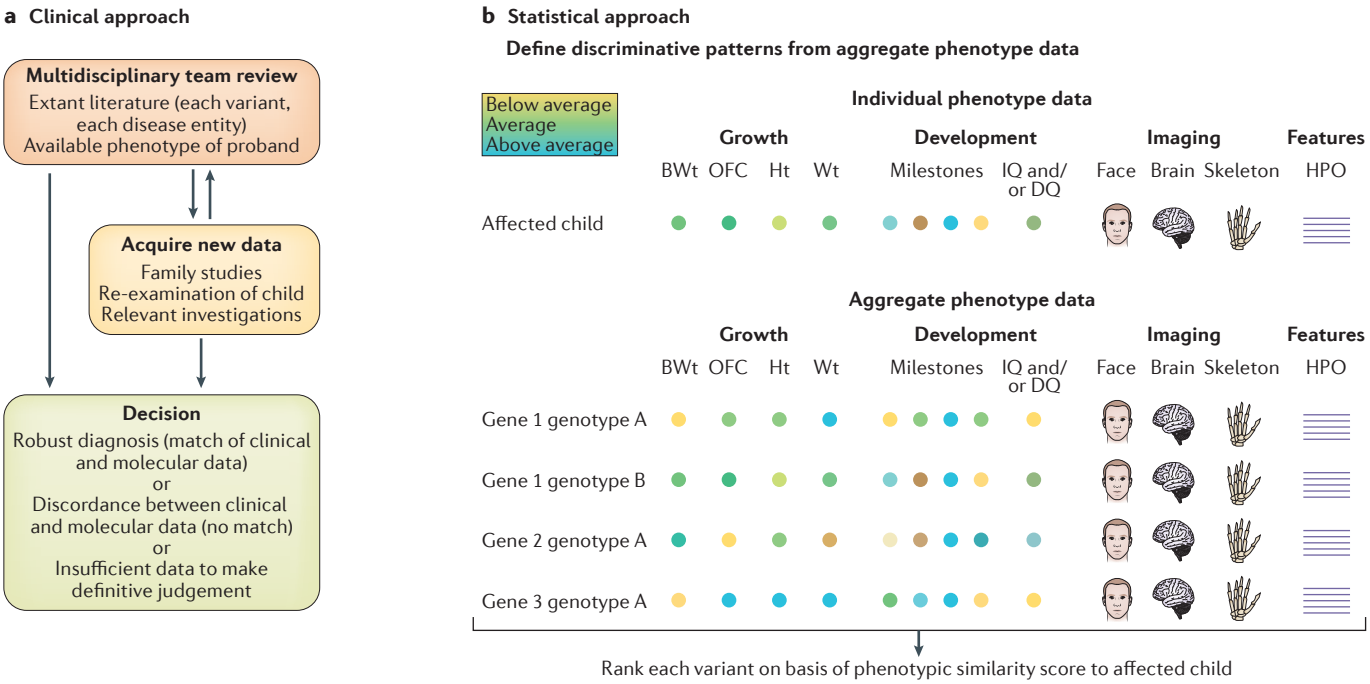


Figure 4 | **Multidisciplinary team review with phenotype-based variant ranking and interpretation.** Two complementary approaches (clinical and statistical) can be used for phenotype-based ranking. **a** | In the clinical approach, variants that survive clinical filtering for consequence, frequency and inheritance should be assessed with reference to the phenotype observed in the affected child. This is best done with the support of an expert multidisciplinary team, who will review the patient's presenting features, family history, evolving clinical story and results of other investigations (such as biochemical testing and imaging) and refer to available aggregate data on the clinical features associated with a given gene (as in FIG. 4b) and to the published literature. Based on the suggested molecular diagnosis, new data (such as family studies, further clinical examination or reinvestigation of the child) may be required to make a definitive decision and to establish a robust genetic diagnosis or refute a molecular finding that does not explain the patient's clinical features or inheritance pattern observed in the family. In a proportion of cases, no definitive decision can be made on whether a variant is causative or not. **b** | A statistical approach to phenotype matching requires that the categorical, quantitative and/or image-based phenotype information is available both in the affected child and in a considerable number of other individuals with causative variants in the same gene. These data will allow a statistical approach to quantify the phenotype matching between the affected child and the disease model compared with the match to other unrelated disease models. Such approaches become more feasible as the number of affected children with rare and ultra-rare molecularly defined genetic disorders rises. BWt, body weight; DQ, developmental quotient; HPO, Human Phenotype Ontology; Ht, height; OFC, occipitofrontal circumference; Wt, weight.

stored in standardized Variant Call Format (VCF) files, which are commonly modified for clinical analysis to include annotation of each variant. Some annotations are crucial for clinical interpretation, particularly the minor allele frequencies (MAFs) and variant consequence predictions. All annotations, but MAF data in particular, vary with the data and software used at the point of annotation. As a result, VCF files derived from the same data can vary substantially among providers and between versions despite having a standardized structure[68,69].

Variant filtering. Decisions about which variants to include in diagnostic evaluation or to exclude from further analysis are heavily dependent on the resources used and the date they were accessed. Variants are typically excluded if they have no predicted or known functional consequence, if they are deemed to be too common to be the cause of a particular disease or because they do not fit with the observed pattern of inheritance for the disease[70] (FIG. 3). If using a trio approach, care needs to be taken during phenotypic evaluation of the parents to ensure that they do not have a mild, subclinical phenotype relevant to the disorder in their child; otherwise, the wrong inheritance pattern will be specified, potentially confounding the analysis. Filtering out variants with no known function requires accurate prediction of the effect of the genetic variant by software such as Ensembl VEP[71] or SNPeff[72]. However, these predictions may not reflect the real biological consequence of the variant[73,74]. Moreover, the same genomic variant may have different predicted consequences owing to the presence of alternative transcripts, and the lack of a simple selection method to determine the clinically relevant transcript for most genes is a significant source of variation between diagnostic labs[67]. Identification of common variants requires access to large-scale population resources, such as ExAC and gnomAD[75]. At the date of writing, these databases contained variants from more than 120,000 exomes and 15,000 genomes from a range of adult study cohorts, though many ethnic groups remain under-represented. Genetic variants that cause severe childhood diseases are rarely present

Minor allele frequencies (MAFs). Measurements of how often the less common allele occurs at a given polymorphic locus.

in a disease-causing configuration in adult disease cohorts or their control groups, and lifestyle factors and long-term environmental exposures are unlikely to lead to substantive disease variability in early-onset disease. As a result, very stringent allele frequency thresholds can be used to select only very rare variants for further analysis in children. However, the ethnicity of the individual being sequenced and how well their ethnicity is represented in reference data sets can affect the effectiveness of variant filtering, and there continues to be a European bias in genomic analysis[76].

Assigning pathogenicity. A thorough clinical workup remains crucial for interpreting the candidate variants identified by genomic sequencing and for maintaining good clinical practice (FIGS 3,4). Assigning pathogenicity to any of the rare variants selected by an automated pipeline and establishing a genetic diagnosis can be extremely challenging. Variants in genes previously associated with the disease in question are often prioritized using virtual gene panels[77], and the ability to identify individuals who do not have one of the well-known causes for a disease often facilitates the identification of novel disease determinants[78]. Unfortunately, spurious gene–disease associations are commonplace in the literature, which results in enormous differences between gene panels and a lack of clear consensus regarding the evidence level required for a gene to be included in a gene panel[79,80]. Moreover, clinically useful databases of pathogenic variation, such as the Human Gene Mutation Database, contain numerous errors[81,82], leading to benign variants being incorrectly picked out of the data and assigned as plausible diagnoses. There is potential for this situation to improve as more genomes are sequenced, but the importance of large data sets comprising populations of genomes from both healthy individuals and individuals with rare diseases, such as gnomAD and ClinVar, cannot be overstated[75,83]. Large paediatric cohorts such as the Deciphering Developmental Disorders Study in the UK have facilitated the development of statistically robust methods for establishing pathogenicity solely on the basis of genomic data[9,84]. Such studies have also been aided by the acquisition of both categorical and quantitative phenotypic data. Categorical approaches use structured ontologies, such as the Human Phenotype Ontology[85] (HPO), to capture important clinical features, while quantitative data in paediatric genetics include growth and developmental milestones. The publication of the American College of Medical Genetics and Genomics–Association for Molecular Pathology (ACMG–AMP) guidelines[86] was a major step towards establishing a common framework for variant classification, which continues to be refined[87].

Platforms such as DECIPHER allow standardized phenotype capture by clinicians and its integration with genotype as a patient-level record within a single resource, enabling holistic exploration of relevant data. DECIPHER[88,89] displays phenotype information, imaging data, genotype information (including variant type (such as SNV or CNV), consequence and mode of inheritance) and 2D and 3D protein structure (where available), as well as reference and population resources. It is supplemented by ACMG variant pathogenicity classification support[86] and an assessment tool to facilitate multidisciplinary team decisions based on integration of molecular and clinical expertise in order to achieve a robust genetic diagnosis. The systematic capture of categorical and quantitative phenotypic data for individual children with rare and ultra-rare genetic disorders will also facilitate statistical approaches to identify discriminative patterns for each disorder. This will enable a confident assessment of the phenotypic similarity between the observed phenotype in the tested individual and the expected phenotype associated with each of the plausible causative genotypes that have been generated by the genetic test (FIG. 4b).

Data sharing. Interpretation of genomic data is highly dependent on access to databases of variants from both control and affected individuals. However, widespread data sharing has logistical challenges because of the sheer volume of data, and ethical challenges because it is impossible to truly anonymize genome sequence data[90,91] (BOX 1). Sharing curated or suspected phenotype-variant associations through resources such as DECIPHER[89] and Phenome Central[92] facilitates this process and minimizes the number of variants shared per individual. Initiatives such as the GA4GH Matchmaker Exchange[93] federate these and other resources, such as GeneMatcher[94], which increases the scope of searches. The ACMG recently issued a position statement that laboratory and clinical genomic data sharing was crucial to improving genetic health care[95]. The opportunities provided by WES and WGS to enable diagnosis at the edge of our knowledge can be underpinned by close review of the current literature, a working knowledge of the appropriate interpretation platforms and knowledge repositories and interface with research studies. Ideally, this framework should be complemented by a collaborative infrastructure to support functional and model organism studies, which bring a broader expertise to achieving a robust molecular and clinical diagnosis[96].

Considerations for clinical application

The expertise to clinically evaluate the diagnostic importance of genomic variants and to communicate noteworthy results to the family and to other health professionals is an essential prerequisite to the introduction of NGS testing in a health-care system[97]. The protracted search for a diagnosis, which can extend over many years, is often resolved when new technology (such as NGS) and new knowledge (such as the discovery of new disease genes) enable diagnoses to be made today that were not possible several years ago. Appropriate use of WES and WGS has the potential to dramatically shorten the 'diagnostic odyssey' for patients with rare and ultra-rare paediatric disorders[98]. It can also often prevent the need for invasive and expensive investigations, such as muscle biopsy and lumbar puncture. However, especially with

Human Phenotype Ontology
(HPO). A standardized vocabulary of phenotypic abnormalities encountered in human disease. Each term in the HPO describes a phenotypic abnormality, such as atrial septal defect. The HPO currently contains ~11,000 terms.

Causative genotypes
Genotypes in which a single locus is perturbed and that have a high positive predictive value for a restricted pattern of morphological, biochemical or physiological features of clinical significance.

very young patients, investigations that may consolidate a clinical diagnosis should continue in parallel with WES or WGS, as the results may be helpful in establishing the validity (or otherwise) of diagnoses suggested by candidate variants emerging from the diagnostic pipeline. Follow-up investigations (such as enzyme assays or cranial magnetic resonance imaging (MRI) scans) may be needed after NGS to confirm whether features normally found in the candidate diagnosis are present and whether the suggested diagnosis is clinically plausible. The less specific the presentation (for instance, neonatal hypotonia) is, the greater the problem of determining whether candidate variants are pertinent findings. With careful patient selection in a neonatal intensive care unit (NICU), it has been possible to establish a genetic diagnosis for critically ill infants in as little as 26 hours[99], but this time frame remains unattainable in most health-care settings because of a lack of relevant expertise and appropriate logistics and because of the currently prohibitive cost of delivering such expedited results. However, it establishes that expedited diagnosis is feasible and provides a strong drive to improve the speed of current pipelines that are more typically measured in months than in days[100].

Clinical expertise. Although it is possible for a few conditions (such as achondroplasia, Genetic Testing Registry MIM 100800) to make a definitive molecular diagnosis from the genotype or a confident clinical diagnosis from the phenotype, establishing a robust genetic diagnosis for most rare diseases requires the integration and correlation of molecular genetic or genomic data (genotype) with the clinical features (phenotype) of the patient (FIG. 3). The depth of phenotype required to do this accurately is not yet established for most conditions, and specialist assessment by a clinician with expertise in rare disease is generally recommended to ensure that a molecular diagnosis of severe genetic disease in a young person is well founded. Similarly, whether the molecular diagnosis represents a full explanation for the clinical presentation or whether the child has a blended phenotype also merits expert assessment. Making a lifelong genetic diagnosis in a young patient with a severe disorder is a crucial step in enabling safe and appropriate ongoing management for that patient and their family. Depending on the genetic variant or variants, this process often requires review of recent primary publications, additional clinical evaluation and investigation to determine the clinical fit and to corroborate or refute a suggested molecular diagnosis, and segregation studies to determine whether the suggested molecular diagnosis fits with the inheritance pattern observed in the family. Furthermore, for patients with rare or ultra-rare disorders, time should be invested to adequately investigate the availability of research studies and treatment trials. This package of care is time-consuming but essential to safe practice in genomic medicine; the time and expertise necessary to deliver it may not be easy to accommodate in a general paediatric consultation, and referral to a clinical genetics service may be appropriate.

Blended phenotype
A mixed phenotype that results from causal variants in two or more genes. The phenotypes may either be distinct, with discrete (composite) manifestations, or overlapping, with similar phenotypic manifestations that are impossible to disentangle.

Where best to focus clinical genomics expertise, whether at the point of test request, at the point of evaluation of candidate variants or both, is not yet clear. Experience with the implementation of genomic arrays[101] suggests that this expertise is best employed once candidate variants have been identified that require expert clinical evaluation to determine whether they represent robust genetic diagnoses. Patients for whom an initial WES or WGS test report is normal but a genetic basis for disease remains very likely will also benefit from expert assessment by a geneticist. As WGS becomes more widely deployed, the clinically focused reanalysis of existing WGS data is likely to emerge as a major focus for clinical geneticists as phenotype-informed and patient-centred processes improve diagnostic yield[17]. If a possible clinical diagnosis can be identified, existing NGS data can be re-evaluated with less stringency and with additional analytical strategies to try to identify variants that have been overlooked by the standard diagnostic analysis pipeline.

Who to test. Patient selection is crucial for ensuring that WES or WGS is deployed primarily for patients with a significant chance of a monogenic cause for their presentation (BOX 2). In this context, using WES or WGS as part of a battery of investigations with the intention to exclude a genetic cause for disease is particularly problematic. Many diagnostic pathways, especially in paediatric neurology, use a wide variety of biochemical investigations to exclude rare conditions, such as inborn errors of metabolism, where the prior probability of each individual condition causing the clinical presentation is very low. This is part of long-established practice and may be appropriate where normal ranges are well documented. However, the wealth of uncertain variants generated by NGS and the lack of confidence with which such variants can be definitively assigned as benign or pathogenic mean that it is not appropriate to use WES or WGS as a generic exclusion test for a genetic cause of disease; NGS investigations should therefore currently be reserved for patients in whom there is a high index of clinical suspicion for an underlying genetic disorder. Indeed, with appropriate case selection, an NGS diagnosis can be transformative and alter management. A recent study that used WES to establish a molecular diagnosis for patients in the NICU with a likely monogenic disorder showed that clinical care was altered by the diagnosis in 23 of 32 such patients (72%)[102].

Test selection. The choice of test is influenced by many factors, not least the confidence in and specificity of the clinical diagnosis and the cost and availability of NGS assays, analysis and interpretation (FIG. 2c). For example, the most appropriate test for an infant who is failing to thrive with fat malabsorption, a sweat chloride measurement of 88 mmol per litre and a clinical diagnosis of cystic fibrosis is a single-gene test of the *CFTR* gene, not WES or WGS. In general terms, the broader the locus heterogeneity of the condition is, the greater

Box 2 | **Indications for paediatric genomic analysis using WES, WGS or large panel approaches**

1. Neurodevelopmental disorder: for example, developmental delay and/or learning disability (of a level requiring or likely to require a statement of special educational needs), epileptic encephalopathy or severe cerebral palsy.
2. Congenital anomalies: multiple congenital anomalies (two or more major anomalies) or a single major anomaly together with a neurodevelopmental disorder, aberrant growth, dysmorphic features or unusual behaviour.
3. Abnormal growth parameters (height, weight, occipitofrontal circumference): two or more parameters >3 s.d. above or below the mean or a single parameter >4 s.d. above or below the mean (except for obesity where the threshold for isolated obesity is >4.5 s.d. together with a strong suspicion of a genetic aetiology).
4. Dysmorphic features.
5. Unusual behavioural phenotype in conjunction with one or more of the above features or extreme behavioural phenotype strongly suspected to have a genetic basis.
6. Disorder of considerable impact for which a simple genetic basis is thought likely with the following: several affected family members; one other affected family member with a rare, consistent and distinctive phenotype; or a single case that is associated with a particularly severe phenotype.

the utility of WES or WGS. However, the broader the scope of the test (that is, the number of genes covered) is, the greater the expertise required to interpret the result, particularly for the many disorders for which simple confirmatory investigations (such as enzyme assays) are not available.

WES outperforms panel testing in expert hands for research and diagnostic purposes[103], but whether it better enables accurate diagnosis in a routine clinical setting is a different question and one where there is currently little published research to inform practice. Modern well-designed WES includes efficient capture of promoters and splice regions as well as coding sequence, enabling the great majority of SNVs causing rare disease to be detected[104]. In 2016, WES cost less than a third of the cost of WGS (internal data), so WES of the family trio could be performed for approximately the same cost as WGS of the child. The additional power of trio analysis for detecting *de novo* mutations (which are the most common cause of disease in childhood developmental disorders)[9] coupled with our current inability to interpret non-coding variants means that, where cost is a limiting factor, trio WES will outperform proband-only WGS in paediatric rare disease diagnosis. Nonetheless, the benefits of WGS in terms of data quality across the exome and access to non-coding and structural variants may ultimately outweigh the current cost benefits of WES[105].

Complexity. Some patients present diagnostic challenges because they have two or more genetic diagnoses that lead to a blended phenotype[106]. Multiple genetic diagnoses in a single patient have been estimated to occur at a rate of approximately 5% (REFS 8,19). Clinicians therefore need to consider whether the candidate diagnosis identified by a genetic test represents a full explanation for their patients' clinical features or only a partial one, with the main diagnosis still to be identified (BOX 3). Even for patients with a single diagnosis, the clinical features observed in an individual patient are rarely the result of a single variant acting in isolation. A variant in a gene of major effect may act in concert with variants in genes of modest effect and a large number of variants of minor effect (often in non-coding parts of the genome). The combined effects of these variants, together with environmental exposures, may influence penetrance and expressivity to determine the clinical presentation, thus explaining some of the variability seen between individuals with the same genetic diagnosis. Such complexity likely underlies many behavioural and developmental traits in the normal population, and for some individuals, even in the absence of a single variant of major effect, a cumulative burden of adverse variants of modest and minor effects may result in a diagnosis of, for example, an autism spectrum disorder[107]. Over time, our understanding of this complexity is likely to improve, and it may become possible to integrate data from highly penetrant single genes with genetic risk scores to better stratify patients and personalize prognosis and management[108,109].

Value. The benefits of an accurate genetic diagnosis include a better understanding of prognosis, more tailored management and improved surveillance. A precise genetic diagnosis enables the provision of accurate genetic advice to individuals and their families and may provide them with increased reproductive choice, for example, by enabling pre-implantation diagnosis, non-invasive prenatal testing or prenatal diagnosis. It also facilitates improved access to education, health and social care and to information and support from patient support groups[2]. Once a robust molecular and clinical diagnosis is made, resources such as GeneReviews[110] and Orphanet provide expert advice on management for clinicians, and resources such as Unique and Genetics Home Reference provide succinct, patient-friendly summaries for patients and their families. Deployment of NGS at an early stage in specialist evaluation offers a cost-effective way to both improve the chance of identifying a diagnosis and significantly shorten the time to diagnosis[111–114]. In some cases, diagnosis can also lead to improved or more personalized treatments, and many of the most striking advances in gene discovery and diagnosis and reports of life-changing case studies from exome and genome sequencing have been made in children[115–118]. For example, in one study of individuals

Genetic risk scores
Quantitative measures of genetic predisposition to a trait that are calculated from data for multiple (usually low-risk) genetic variants, which are usually obtained from genome-wide association studies.

Box 3 | **Case study illustrating the need for expert review of candidate variants**

A genomic microarray analysis for a child presenting with infantile epileptic encephalopathy detects a *de novo* recurrent ~900 kb pathogenic duplication of 1q21.1. An inexperienced clinician may well consider this to be the diagnosis for the child's condition as approximately 15–35% of children with this copy number variation will develop seizures. However, seizures in 1q21.1 duplication syndrome generally respond well to first-line therapy and usually have their onset in childhood rather than in infancy, so it is more likely that this finding is a minor contributor to the clinical presentation, and the main diagnosis has not been identified by this test. When family trio whole-exome sequencing is performed, the major determinate of the clinical presentation is found to be a *de novo* protein truncating c.910G>T, p.Glu304Ter variant in the *CHD2* gene. Expert review by a clinician familiar with rare disease diagnosis is thus essential to determine whether a variant identified by a genetic test is the full cause of a child's presenting features, a contributory factor or an unrelated finding.

with a neurometabolic disorder, WES led to a diagnosis in 68% of patients, identified 11 candidate genes and suggested a change in treatment for 44% (REF. 119).

Future perspectives

Paediatric genomics is a rapidly developing field, and a greater understanding of the molecular basis of disease is already beginning to transform the quality of care that can be delivered today. The future is likely to bring improved workflows that will find a wider range of clinical applications and lead to new treatment options.

Improved implementation of paediatric genomics workflows. Paediatric genomics is still in its infancy; to draw parallels with child development, it is perhaps in the toddler phase, where exploration and acquisition of new skills combine with frequent trips and tumbles. Genome diagnostics will increasingly be applied across many of the paediatric subspecialities, facilitated by multidisciplinary team meetings and expert clinical assessment to support variant interpretation[120,121]. However, optimizing implementation is challenging when nearly every aspect of the analytical pipeline required for an effective genomic medicine service is evolving. Clinicians are learning where best to position genomic tests in patient pathways of investigation; sequencing companies are developing longer-read technology; bioinformaticians are improving algorithms for detecting and prioritizing different variant types; population databases are growing, both in depth and representation of diverse ethnicities; gene discovery continues apace, both in the identification of new genes associated with disease and increasingly in recognizing pleiotropy; and variant databases and the literature continue to grow and improve, although they remain heavily polluted with incorrect pathogenicity assignments. Each of these parameters has the potential to affect test performance, particularly sensitivity and specificity. A key goal will be to optimize variant detection and filtering such that this process maximizes the chance of true diagnosis while minimizing the opportunity for misdiagnosis or overdiagnosis[122,123] for any given test; success will require focused attention on patient ascertainment, phenotyping, comprehensive detection of relevant variant types, gene panel selection and appropriate expertise. Thus, bioinformaticians, clinical scientists and specialist clinicians all have important roles to play in the safe and effective practice of genetic medicine.

Integrating multi-omic data to improve diagnostic ability. NGS technologies have paved the way for the use of other genome-wide technologies, such as transcriptomics, epigenomics, metabolomics and proteomics, to investigate the functional impact of genetic variation on specific tissues. Although these approaches are not yet routinely implemented in a diagnostic setting, they are increasingly being used to determine the pathogenicity of genomic variants[124]. For instance, combined WGS and transcriptomic analysis of muscle biopsy samples from patients with paediatric neuromuscular disease enables validation of candidate splice-disrupting mutations and identification of splice-altering variants in both exonic regions and non-coding regions of genes and yields an overall diagnosis rate of 35%. Furthermore, using transcriptomic data to guide genomic reanalysis, it was possible to diagnose 21% of patients with no strong candidates from WGS or WES[124].

In many cases, deeper phenotyping or standard imaging or biochemical and electrophysiological assays are crucial for determining the functional effect of a particular genomic variant. Even in 2017, most of the ~2,000 variants identified in the *CFTR* gene are rare or private despite the fact that the sequence of *CFTR* was determined in 1989 (REF. 125), and cystic fibrosis and CFTR-related diseases are some of the most prevalent among rare disorders. Ongoing endeavours to comprehensively evaluate genotypic, phenotypic and functional aspects of *CFTR* variants at scale indicate the enormity of the task ahead[126,127].

Application of genomics to diagnose fetal, neonatal and adult disease. Paediatrics has led the field in genetic medicine, and genomics has had a larger impact on it than other specialities for a number of reasons. For instance, genomics-based diagnosis of disease in fetal and neonatal life presents challenges because the phenotypic features are fewer and less specific than at later stages of development; phenotypic assessment is therefore generally weaker, and the ability to discriminate between candidate variants based on phenotype is diminished, which adversely affects the diagnostic yield. Applying genomics to adult medicine and elderly care is also challenging but for different reasons. In adults, the genotypic signal is generally more difficult to interpret than in children because incomplete and age-dependent penetrance often leads to causal variants being present at considerable frequency in population data sets. In addition, polygenic causes may be responsible for a greater proportion of the burden of disease in adults than in children, especially in common diseases. Furthermore, environmental exposures are likely to be more important in adult

Pleiotropy
The phenomenon whereby variants in a single gene may cause multiple phenotypic expressions or disorders.

Transcriptomics
A global approach for looking at gene expression patterns. This can involve measurements of thousands of genes simultaneously with microarrays or measurements of small numbers of genes that are facilitated by global sequence information from expressed sequence tag or genome-sequencing projects.

Epigenomics
A global approach for looking at the complete collection of epigenetic marks, such as DNA methylation and histone modifications, and other molecules that can transmit epigenetic information, such as non-coding RNAs, that exist in a cell at any given point in time.

Metabolomics
A global approach using quantitative analytical methods to look at the entire metabolic content of a cell or organism at a given time.

Proteomics
A global approach for looking at the complete collection of proteins in a cell or tissue at a given time.

disorders than childhood disorders, and age-related somatic variation is likely to be an important player in late-onset disease. For these reasons, the paediatric population is likely to remain at the forefront of genomic medicine. However, as our knowledge increases, less highly penetrant monogenic forms of disease will become more tractable, and clinical application of genomics to other age groups will become more feasible.

From diagnosis to treatment. Achieving a secure genetic diagnosis of a rare disease is the main goal of paediatric genomics at present, but there is an increasing number of inspiring examples where specific knowledge of the genetic basis of disease is leading to direct therapeutic intervention. WGS inspired a rational approach to therapy for twins with severe dystonia who showed only modest improvement when treated with L-dopa; when sequence analysis revealed that they carried biallelic variants in the *SPR* gene that contributed to reduced synthesis of the neurotransmitters dopamine and serotonin, their L-dopa treatment was supplemented with 5-hydroxytryptophan, and their condition improved dramatically[128]. Another exciting example is provided by SMA. For the great majority (>95%) of patients with SMA, a homozygous deletion of exon 7 of the *SMN1* gene leads to a lack of functional survival motor neuron (SMN) protein. A second gene, *SMN2*, also encodes SMN protein, but incorrect splicing of *SMN2* transcripts results in a non-functional truncated SMN protein. The antisense drug nusinersen has been specifically designed to alter splicing of *SMN2* pre-mRNA to increase the amount of functional SMN protein[129]. A recent trial showed that 51% of infants receiving intrathecal nusinersen had a motor-milestone response versus none in the sham control group[130]. Genome sequencing can also be used to stratify patients for therapy; for instance, in patients homozygous for the *F508del-CFTR* mutation, combination therapy with lumacaftor and ivacaftor is associated with improved lung function and a 42% slower rate of decline of ppFEV1 (percentage predicted forced expiratory volume in 1 second; a measure of lung function) than in matched registry controls[117]. Stratified therapy for rare paediatric disease based on a specific genetic diagnosis and development of effective and affordable therapies through drug repurposing and innovation are exciting applications of paediatric genomics that are likely to become routine in the near future. Over the longer term, corrective therapies for Mendelian disease using CRISPR–Cas9 genome editing technologies, which are now feasible using *ex vivo* approaches, may become a reality for patients[131]. The potential and scope of CRISPR–based technologies are further improved by the development of recent variations, such as Cas9-mediated adenine base editing[132] and Cas13-mediated RNA-editing[133].

Conclusions

Paediatric genomics as currently practised is focused primarily on establishing genetic diagnoses to explain rare paediatric disorders using the substantially increased diagnostic power of genomic technologies. It has broad applicability across a range of paediatric phenotypes, including neurodevelopmental disorders, multiple congenital anomalies, infantile epileptic encephalopathy and extreme obesity, among others. Rare genetic diseases are usually severe, lifelong and sometimes life-limiting conditions, so proper investment in expert clinical assessment is appropriate to ensure that the diagnosis suggested by genomic analysis is clinically sound. The diagnostic yield of genomic sequencing in previously unsolved paediatric cases is already around 40% (REF. 134) (though this figure varies with clinical indication)[77] and will continue to increase as knowledge grows.

There remains much to discover, even about the fundamental causes of disease, and currently ~70% of protein-coding genes have no established human disease phenotype[2,75]. However, our understanding of the genomic architecture of rare paediatric disease and the multitude of mechanisms by which variants, either singly or in combination, can cause disease continues to improve. Further insight into the molecular basis of most rare diseases will also yield therapeutic benefits: treatment options will be streamlined, existing drugs may be re-purposed, and novel targeted therapeutics will be developed. Indeed, recent studies on cystic fibrosis have shown that stratifying disease by its molecular genetic basis enables more rigorous clinical trials to evaluate new therapies[135]. Newly published discoveries can be implemented remarkably quickly into a WGS or WES clinical test through a simple software update that adds a new gene to a virtual panel, which obviates the need to design, test and market a new testing kit. Moreover, the comprehensive nature of genomic data could see it become an integral part of routine care for all children. Genomic data could be used to provide improved newborn screening[136,137], to target immunization by identifying susceptibility to specific infections[138] and to help stratify risk and personalize treatment not only for rare disease but also for common paediatric disorders[139,140].

Although it is clear that *in silico* analysis can undoubtedly improve variant interpretation, the inherent biological complexity by which genotype gives rise to phenotype largely confounds attempts to rely solely on automated predictions. Efforts to document clinical genotype–phenotype associations remain of paramount importance to support the safe practice of genomic medicine. Further investigation is needed to evaluate clinical utility and factors that affect the reliability of genomic diagnosis. Additional research is also needed to determine the amount and type of phenotypic information required to support safe genomic diagnosis and where in the diagnostic pathway this information is best incorporated. Nevertheless, with continued progress in these areas, and as sequencing costs continue to drop and knowledge grows, we can look forward to a future where almost every child with a serious rare genetic disease could have access to an accurate, specific genetic diagnosis.

1. European Organisation for Rare Diseases. *Rare Diseases: Understanding this Public Health Priority*. (Eurodis, 2005).
2. Boycott, K. M. *et al.* International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.* **100**, 695–705 (2017).
3. Quintana-Murci, L. Understanding rare and common diseases in the context of human evolution. *Genome Biol.* **17**, 225 (2016).
4. Amberger, J., Bocchini, C. A., Scott, A. F. & Hamosh, A. Mckusick's online mendelian inheritance in man (OMIM). *Nucleic Acids Res.* **37**, D793–D796 (2009).
5. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
6. 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).
7. Dodge, J. A. *et al.* The importance of rare diseases: from the gene to society. *Arch. Dis. Child* **96**, 791–792 (2011).
8. Wright, C. F. *et al.* Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* **385**, 1305–1314 (2015).
This is an important paper outlining a prototype clinical bioinformatics pipeline for the diagnosis of developmental disorders.
9. Deciphering Developmental Disorders Study. Prevalence and architecture of *de novo* mutations in developmental disorders. *Nature* **542**, 433–438 (2017).
This is a landmark paper highlighting the major contribution of *de novo* mutation to developmental disorders.
10. Austin, C. P. *et al.* Future of rare diseases research 2017-2027: an IRDiRC perspective. *Clin. Transl Sci.* https://doi.org/10.1111/cts.12500 (2017).
11. Grozeva, D. *et al.* Targeted next-generation sequencing analysis of 1,000 individuals with intellectual disability. *Hum. Mutat.* **36**, 1197–1204 (2015).
12. Kochinke, K. *et al.* Systematic phenomics analysis deconvolutes genes mutated in intellectual disability into biologically coherent modules. *Am. J. Hum. Genet.* **98**, 149–164 (2016).
13. Torgerson, P. R. & Mastroiacovo, P. The global burden of congenital toxoplasmosis: a systematic review. *Bull. World Health Organ.* **91**, 501–508 (2013).
14. Del Campo, M. & Jones, K. L. A review of the physical features of the fetal alcohol spectrum disorders. *Eur. J. Med. Genet.* **60**, 55–64 (2017).
15. Chavali, P. L. *et al.* Neurodevelopmental protein Musashi-1 interacts with the Zika genome and promotes viral replication. *Science* **357**, 83–88 (2017).
16. Firth, H. V. & Wright, C. F. & DDD Study. The Deciphering Developmental Disorders (DDD) study. *Dev. Med. Child Neurol.* **53**, 702–703 (2011).
17. Baynam, G. *et al.* The rare and undiagnosed diseases diagnostic service — application of massively parallel sequencing in a state-wide clinical service. *Orphanet J. Rare Dis.* **11**, 77 (2016).
18. Doherty, E. S. *et al.* Muenke syndrome (FGFR3-related craniosynostosis): expansion of the phenotype and review of the literature. *Am. J. Med. Genet. A* **143A**, 3204–3215 (2007).
19. Posey, J. E. *et al.* Resolution of disease phenotypes resulting from multilocus genomic variation. *N. Engl. J. Med.* **376**, 21–31 (2017).
This interesting paper focuses on diagnosing individuals with several independent rare genetic conditions.
20. Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nat. Rev. Genet.* **16**, 333–343 (2015).
21. Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat. Rev. Genet.* **14**, 681–691 (2013).
22. Samocha, K. E. *et al.* A framework for the interpretation of *de novo* mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
This is a useful paper outlining a model for predicting the number of *de novo* mutations expected by chance across the genome, which is essential for robust discovery of genes that cause novel dominant *de novo* disorders.
23. Weiner, D. J. *et al.* Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk of autism spectrum disorders. *Nat. Genet.* **49**, 978–985 (2017).
24. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
25. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
26. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
This study presents a comprehensive account of the wealth of variation in the genomes of normal individuals.
27. Katsanis, S. H. & Katsanis, N. Molecular genetic testing and the future of clinical genomics. *Nat. Rev. Genet.* **14**, 415–426 (2013).
28. Brewington, J. & Clancy, J. P. Diagnostic testing in cystic fibrosis. *Clin. Chest Med.* **37**, 31–46 (2016).
29. Aartsma-Rus, A., Ginjaar, I. B. & Bushby, K. The importance of genetic diagnosis for Duchenne muscular dystrophy. *J. Med. Genet.* **53**, 145–151 (2016).
30. Speicher, M. R. & Carter, N. P. The new cytogenetics: blurring the boundaries with molecular biology. *Nat. Rev. Genet.* **6**, 782–792 (2005).
This is a useful Review of microarray technologies and their use for the diagnosis of rare paediatric syndromes.
31. Shaw-Smith, C. *et al.* Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J. Med. Genet.* **41**, 241–248 (2004).
32. Crespi, B. J. & Procyshyn, T. L. Williams syndrome deletions and duplications: genetic windows to understanding anxiety, sociality, autism, and schizophrenia. *Neurosci. Biobehav. Rev.* **79**, 14–26 (2017).
33. Sagoo, G. S. *et al.* Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects. *Genet. Med.* **11**, 139–146 (2009).
34. Mardis, E. R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **24**, 133–141 (2008).
35. Ansorge, W. J. Next-generation DNA sequencing techniques. *N. Biotechnol.* **25**, 195–203 (2009).
36. Moorthie, S., Mattocks, C. J. & Wright, C. F. Review of massively parallel DNA sequencing technologies. *Hugo J.* **5**, 1–12 (2011).
37. Vissers, L. E. L. M. *et al.* A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genet. Med.* **19**, 1055–1063 (2017).
This is an excellent study of the impact of NGS in clinical practice.
38. Shashi, V. *et al.* The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet. Med.* **16**, 176–182 (2014).
39. Weiss, M. M. *et al.* Best practice guidelines for the use of next-generation sequencing applications in genome diagnostics: a national collaborative study of Dutch genome diagnostic laboratories. *Hum. Mutat.* **34**, 1313–1321 (2013).
40. Sun, Y. *et al.* Next-generation diagnostics: gene panel, exome, or whole genome? *Hum. Mutat.* **36**, 648–655 (2015).
41. Ece Solmaz, A. *et al.* Targeted multi-gene panel testing for the diagnosis of Bardet Biedl syndrome: Identification of nine novel mutations across BBS1, BBS2, BBS4, BBS7, BBS9, BBS10 genes. *Eur. J. Med. Genet.* **58**, 689–694 (2015).
42. Schrijver, I. Hereditary non-syndromic sensorineural hearing loss. *J. Mol. Diagn.* **6**, 275–284 (2004).
43. Myers, C. T. & Mefford, H. C. Advancing epilepsy genetics in the genomic era. *Genome Med.* **7**, 91 (2015).
44. Mastrangelo, M. Novel genes of early-onset epileptic encephalopathies: from genotype to phenotypes. *Pediatr. Neurol.* **53**, 119–129 (2015).
45. Cheng, A. Y., Teo, Y.-Y. & Ong, R. T.-H. Assessing single nucleotide variant detection and genotype calling on whole-genome sequenced individuals. *Bioinformatics* **30**, 1707–1713 (2014).
46. Beck, T. F. & Mullikin, J. C., NISC Comparative Sequencing Program & Biesecker, L. G. Systematic evaluation of Sanger validation of next-generation sequencing variants. *Clin. Chem.* **62**, 647–654 (2016).
47. Telenti, A. *et al.* Deep sequencing of 10,000 human genomes. *Proc. Natl Acad. Sci. USA* **113**, 11901–11906 (2016).
48. Li, W. *et al.* Identifying human genome-wide CNV, LOH and UPD by targeted sequencing of selected regions. *PLoS ONE* **10**, e0123081 (2014).
49. de Ligt, J. *et al.* Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum. Mutat.* **34**, 1439–1448 (2013).
50. Noll, A. C. *et al.* Clinical detection of deletion structural variants in whole-genome sequences. *npj Genomic Med.* **1**, 16026 (2016).
51. Suzuki, T. *et al.* Precise detection of chromosomal translocation or inversion breakpoints by whole-genome sequencing. *J. Hum. Genet.* **59**, 649–654 (2014).
52. Ellingford, J. M. *et al.* Validation of copy number variation analysis for next-generation sequencing diagnostics. *Eur. J. Hum. Genet.* **25**, 719–724 (2017).
53. Budworth, H. & McMurray, C. T. A brief history of triplet repeat diseases. *Methods Mol. Biol.* **1010**, 3–17 (2013).
54. Nowak, K. J. & Davies, K. E. Duchenne muscular dystrophy and dystrophin: pathogenesis and opportunities for treatment. *EMBO Rep.* **5**, 872–876 (2004).
55. Singh, N. N., Seo, J., Rahn, S. J. & Singh, R. N. A multi-exon-skipping detection assay reveals surprising diversity of splice isoforms of spinal muscular atrophy genes. *PLoS ONE* **7**, e49595 (2012).
56. Halvorsen, M. *et al.* Mosaic mutations in early-onset genetic diseases. *Genet. Med.* **18**, 746–749 (2016).
57. Rios, J. J. & Delgado, M. R. Using whole-exome sequencing to identify variants inherited from mosaic parents. *Eur. J. Hum. Genet.* **23**, 547–550 (2015).
58. Saudi Mendeliome Group. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. *Genome Biol.* **16**, 134 (2015).
59. van El, C. G. *et al.* Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. *Eur. J. Hum. Genet.* **21**, 580–584 (2013).
60. Green, R. C. *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15**, 565–574 (2013).
This is a controversial paper advocating routine opportunistic screening of genomic sequence data in adults and children.
61. Matthijs, G. *et al.* Guidelines for diagnostic next-generation sequencing. *Eur. J. Hum. Genet.* **24**, 2–5 (2016).
62. Boycott, K. *et al.* The clinical application of genome-wide sequencing for monogenic diseases in Canada: position statement of the Canadian College of Medical Geneticists. *J. Med. Genet.* **52**, 431–437 (2015).
63. Ormondroyd, E. *et al.* "Not pathogenic until proven otherwise": perspectives of UK clinical genomics professionals toward secondary findings in context of a Genomic Medicine Multidisciplinary Team and the 100,000 Genomes Project. *Genet. Med.* https://doi.org/10.1038/gim.2017.157 (2017).
64. Goldstein, D. B. *et al.* Sequencing studies in human genetics: design and interpretation. *Nat. Rev. Genet.* **14**, 460–470 (2013).
65. Wright, C. F. *et al.* Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1133 families withdevelopmental disorders. *Genet. Med.* https://doi.org/10.1038/gim.2017.246 (2018).
66. Moorthie, S., Hall, A. & Wright, C. F. Informatics and clinical genome sequencing: opening the black box. *Genet. Med.* **15**, 165–171 (2013).
67. Yen, J. L. *et al.* A variant by any name: quantifying annotation discordance across tools and clinical databases. *Genome Med.* **9**, 7 (2017).
68. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
69. Endrullat, C., Glökler, J., Franke, P. & Frohme, M. Standardization and quality management in next-generation sequencing. *Appl. Transl Genom.* **10**, 2–9 (2016).
70. Salgado, D., Bellgard, M. I., Desvignes, J.-P. & Béroud, C. How to identify pathogenic mutations among all those variations: variant annotation and filtration in the genome sequencing era. *Hum. Mutat.* **37**, 1272–1282 (2016).

71. McLaren, W. *et al.* The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
72. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
73. Desmet, F.-O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).
74. Soemedi, R. *et al.* Pathogenic variants that alter protein code often disrupt splicing. *Nat. Genet.* **49**, 848–855 (2017).
75. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
This is a landmark paper describing the ExAC database.
76. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
77. Retterer, K. *et al.* Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* **18**, 696–704 (2016).
This is a useful paper comparing the diagnostic yield of WES across different clinical indications.
78. Ansari, M. *et al.* Genetic heterogeneity in Cornelia de Lange syndrome (CdLS) and CdLS-like phenotypes with observed and predicted levels of mosaicism. *J. Med. Genet.* **51**, 659–668 (2014).
79. Chambers, C., Jansen, L. A. & Dhamija, R. Review of commercially available epilepsy genetic panels. *J. Genet. Couns.* **25**, 213–217 (2016).
80. Strande, N. T. *et al.* Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* **100**, 895–906 (2017).
81. Biesecker, L. G. Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project. *Genet. Med.* **14**, 393–398 (2012).
82. Ghouse, J. *et al.* Numerous Brugada syndrome-associated genetic variants have no effect on J-point elevation, syncope susceptibility, malignant cardiac arrhythmia, and all-cause mortality. *Genet. Med.* **19**, 521–528 (2017).
83. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–868 (2016).
84. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2015).
85. Köhler, S. *et al.* The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* **42**, D966–D974 (2014).
86. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
87. Nykamp, K. *et al.* Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet. Med.* **19**, 1105–1117 (2017).
88. Bragin, E. *et al.* DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res.* **42**, D993–D1000 (2014).
89. Chatzimichali, E. A. *et al.* Facilitating collaboration in rare genetic disorders through effective matchmaking in DECIPHER. *Hum. Mutat.* **36**, 941–949 (2015).
90. Kaye, J. The tension between data sharing and the protection of privacy in genomics research. *Annu. Rev. Genom. Hum. Genet.* **13**, 415–431 (2012).
91. Gymrek, M., McGuire, A. L., Golan, D., Halperin, E. & Erlich, Y. Identifying personal genomes by surname inference. *Science* **339**, 321–324 (2013).
92. Buske, O. J. *et al.* PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum. Mutat.* **36**, 931–940 (2015).
93. Philippakis, A. A. *et al.* The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum. Mutat.* **36**, 915–921 (2015).
94. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928–930 (2015).
95. ACMG Board Of Directors. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. *Genet. Med.* **19**, 721–722 (2017).
96. Ramoni, R. B. *et al.* The Undiagnosed Diseases Network: accelerating discovery about health and disease. *Am. J. Hum. Genet.* **100**, 185–192 (2017).
97. Bowdin, S. *et al.* Recommendations for the integration of genomics into clinical practice. *Genet. Med.* **18**, 1075–1084 (2016).
98. Thevenon, J. *et al.* Diagnostic odyssey in severe neurodevelopmental disorders: toward clinical whole-exome sequencing as a first-line diagnostic test. *Clin. Genet.* **89**, 700–707 (2016).
99. Petrikin, J. E., Willig, L. K., Smith, L. D. & Kingsmore, S. F. Rapid whole genome sequencing and precision neonatology. *Semin. Perinatol* **39**, 623–631 (2015).
100. Saunders, C. J. *et al.* Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci. Transl Med.* **4**, 154ra135 (2012).
101. Miller, D. T. *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* **86**, 749–764 (2010).
102. Meng, L. *et al.* Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. *JAMA Pediatr.* **171**, e173438 (2017).
103. Hartley, T. *et al.* Whole-exome sequencing is a valuable diagnostic tool for inherited peripheral neuropathies: outcomes from a cohort of 50 families. *Clin. Genet.* https://doi.org/10.1111/cge.13101 (2017).
104. Warr, A. *et al.* Exome sequencing: current and future perspectives. *G3* **5**, 1543–1550 (2015).
105. Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl Acad. Sci. USA* **112**, 5473–5478 (2015).
106. Boycott, K. M. & Innes, A. M. When one diagnosis is not enough. *N. Engl. J. Med.* **376**, 83–85 (2017).
107. Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
108. Patel, K. A. *et al.* Type 1 diabetes genetic risk score: a novel tool to discriminate monogenic and type 1 diabetes. *Diabetes* **65**, 2094–2099 (2016).
This study provides a demonstration of the application of a genetic risk score derived from genome-wide association studies to discriminate between common complex and rare monogenic disease.
109. Oram, R. A. *et al.* A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diabetes Care* **39**, 337–344 (2016).
110. Adam, M. P. *et al. GeneReviews®* (University of Washington, Seattle, 2017).
111. Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat. Rev. Genet.* **17**, 9–18 (2016).
112. Tan, T. Y. *et al.* Diagnostic impact and cost-effectiveness of whole-exome sequencing for ambulant children with suspected monogenic conditions. *JAMA Pediatr.* **171**, 855–862 (2017).
This study presents evidence that cost-effectiveness is maximized by early application of WES in the diagnostic pathway of children with suspected monogenic conditions.
113. Stark, Z. *et al.* Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. *Genet. Med.* **19**, 867–874 (2017).
114. Stark, Z. *et al.* A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet. Med.* **18**, 1090–1096 (2016).
115. Soden, S. E. *et al.* Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci. Transl Med.* **6**, 265ra168 (2014).
116. Willig, L. K. *et al.* Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *Lancet Respir. Med.* **3**, 377–387 (2015).
117. Konstan, M. W. *et al.* Assessment of safety and efficacy of long-term treatment with combination lumacaftor and ivacaftor therapy in patients with cystic fibrosis homozygous for the F508del-CFTR mutation (PROGRESS): a phase 3, extension study. *Lancet Respir. Med.* **5**, 107–118 (2017).
118. Worthey, E. A. *et al.* Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* **13**, 255–262 (2011).
This study provides the first published example where WES was successfully used in the clinic to diagnose and treat a child suffering from a severe rare disease.
119. Tarailo-Graovac, M. *et al.* Exome sequencing and the management of neurometabolic disorders. *N. Engl. J. Med.* **374**, 2246–2255 (2016).
120. Stalke, A. *et al.* Diagnosis of monogenic liver diseases in childhood by next-generation sequencing. *Clin. Genet.* https://doi.org/10.1111/cge.13120 (2017).
121. Ormondroyd, E. *et al.* Insights from early experience of a Rare Disease Genomic Medicine Multidisciplinary Team: a qualitative study. *Eur. J. Hum. Genet.* **25**, 680–686 (2017).
122. Moynihan, R., Doust, J. & Henry, D. Preventing overdiagnosis: how to stop harming the healthy. *BMJ* **344**, e3502 (2012).
123. Newman-Toker, D. E. A unified conceptual model for diagnostic errors: underdiagnosis, overdiagnosis, and misdiagnosis. *Diagnosis* **1**, 43–48 (2014).
124. Cummings, B. B. *et al.* Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl Med.* **9**, eaal5209 (2017).
This is an excellent paper demonstrating the improved diagnostic power of combining transcriptome analysis with NGS for the diagnosis of rare neuromuscular disease.
125. Riordan, J. R. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066–1073 (1989).
126. Castellani, C. & CFTR2 Team. CFTR2: How will it help care? *Paediatr. Respir. Rev.* **14** (Suppl. 1), 2–5 (2013).
127. Claustres, M. *et al.* CFTR-France, a national relational patient database for sharing genetic and phenotypic data associated with rare CFTR variants. *Hum. Mutat.* **38**, 1297–1315 (2017).
128. Bainbridge, M. N. *et al.* Whole-genome sequencing for optimized patient management. *Sci. Transl Med.* **3**, 87re3 (2011).
129. Finkel, R. S. *et al.* Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *Lancet* **388**, 3017–3026 (2016).
130. Finkel, R. S. *et al.* Nusinersen versus sham control in infantile-onset spinal muscular atrophy. *N. Engl. J. Med.* **377**, 1723–1732 (2017).
131. Dever, D. P. *et al.* CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature* **539**, 384–389 (2016).
132. Gaudelli, N. M. *et al.* Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
133. Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017).
134. Lionel, A. C. *et al.* Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet. Med.* https://doi.org/10.1038/gim.2017.119 (2017).
135. Taylor-Cousar, J. L. *et al.* Tezacaftor-ivacaftor in patients with cystic fibrosis homozygous for Phe508del. *N. Engl. J. Med.* **377**, 2013–2023 (2017).
136. Friedman, J. M. *et al.* Genomic newborn screening: public health policy considerations and recommendations. *BMC Med. Genom.* **10**, 9 (2017).
137. Berg, J. S. *et al.* Newborn sequencing in genomic medicine and public health. *Pediatrics* **139**, e20162252 (2017).
138. Dale, A. P. & Read, R. C. Genetic susceptibility to meningococcal infection. *Expert Rev. Anti Infect. Ther.* **11**, 187–199 (2013).
139. Bønnelykke, K. & Ober, C. Leveraging gene-environment interactions and endotypes for asthma gene discovery. *J. Allergy Clin. Immunol.* **137**, 667–679 (2016).
140. Michels, A. *et al.* Prediction and prevention of type 1 diabetes: update on success of prediction and struggles at prevention. *Pediatr. Diabetes* **16**, 465–484 (2015).
141. Burke, W. *et al.* The translational potential of research on the ethical, legal, and social implications of genomics. *Genet. Med.* **17**, 12–20 (2015).

142. Hercher, L. & Jamal, L. An old problem in a new age: revisiting the clinical dilemma of misattributed paternity. *Appl. Transl Genom.* **8**, 36–39 (2016).
143. Jackson, L., Goldsmith, L., O'Connor, A. & Skirton, H. Incidental findings in genetic research and clinical diagnostic tests: a systematic review. *Am. J. Med. Genet. A* **158A**, 3159–3167 (2012).
144. Botkin, J. R. *et al.* Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. *Am. J. Hum. Genet.* **97**, 6–21 (2015).
145. Clarke, A. J. Managing the ethical challenges of next-generation sequencing in genomic medicine. *Br. Med. Bull.* **111**, 17–30 (2014).
146. Wright, C. F., Middleton, A. & Parker, M. in *Genomic Medicine Principles and Practice* (eds Kumar, D. & Eng, C.) 250–258 (Oxford Univ. Press, 2014).
147. Anderson, J. A. *et al.* Parents perspectives on whole genome sequencing for their children: qualified enthusiasm? *J. Med. Eth.* **43**, 535–539 (2016).
148. Horn, R. & Parker, M. Opening Pandora's box?: ethical issues in prenatal whole genome and exome sequencing. *Prenat. Diagn.* **23**, 34–39 (2017).
149. Newson, A. J. Whole genome sequencing in children: ethics, choice and deliberation. *J. Med. Eth.* **43**, 540–542 (2017).
150. Committee on Bioethics *et al.* Ethical and policy issues in genetic testing and screening of children. *Pediatrics* **131**, 620–622 (2013).
151. Burstein, M. D., Robinson, J. O., Hilsenbeck, S. G., McGuire, A. L. & Lau, C. C. Pediatric data sharing in genomic research: attitudes and preferences of parents. *Pediatrics* **133**, 690–697 (2014).
152. Wright, C. F., Hurles, M. E. & Firth, H. V. Principle of proportionality in genomic data sharing. *Nat. Rev. Genet.* **17**, 1–2 (2016).
153. Muddyman, D., Smee, C., Griffin, H. & Kaye, J. Implementing a successful data-management framework: the UK10K managed access model. *Genome Med.* **5**, 100 (2013).
154. Wilfond, B. S. & Carpenter, K. J. Incidental findings in pediatric research. *J. Law Med. Eth.* **36**, 332–340 (2008).
155. Eldomery, M. K. *et al.* Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome Med.* **9**, 26 (2017).
156. Carrieri, D. *et al.* Recontacting in clinical genetics and genomic medicine? We need to talk about it. *Eur. J. Hum. Genet.* **25**, 520–521 (2017).
157. Gliwa, C. & Berkman, B. E. Do researchers have an obligation to actively look for genetic incidental findings? *Am. J. Bioeth* **13**, 32–42 (2013).
158. Crawford, G., Foulds, N., Fenwick, A., Hallowell, N. & Lucassen, A. Genetic medicine and incidental findings: it is more complicated than deciding whether to disclose or not. *Genet. Med.* **15**, 896–899 (2013).
159. Clayton, E. W. Incidental findings in genetics research using archived DNA. *J. Law Med. Eth.* **36**, 286–291 (2008).
160. Wolf, S. M. *et al.* Managing incidental findings and research results in genomic research involving biobanks and archived data sets. *Genet. Med.* **14**, 361–384 (2012).
161. Kalia, S. S. *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet. Med.* **19**, 249–255 (2017).
162. Mand, C., Gillam, L., Delatycki, M. B. & Duncan, R. E. Predictive genetic testing in minors for late-onset conditions: a chronological and analytical review of the ethical arguments. *J. Med. Eth.* **38**, 519–524 (2012).
163. Shkedi-Rafid, S., Fenwick, A., Dheensa, S. & Lucassen, A. M. Genetic testing of children for adult-onset conditions: opinions of the British adult population and implications for clinical practice. *Eur. J. Hum. Genet.* **23**, 1281–1285 (2015).
164. Caga-anan, E. C. F., Smith, L., Sharp, R. R. & Lantos, J. D. Testing children for adult-onset genetic diseases. *Pediatrics* **129**, 163–167 (2012).
165. Claustres, M. *et al.* Recommendations for reporting results of diagnostic genetic testing (biochemical, cytogenetic and molecular genetic). *Eur. J. Hum. Genet.* **22**, 160–170 (2014).
166. Wright, C. F. *et al.* Policy challenges of clinical genome sequencing. *BMJ* **347**, f6845 (2013).
167. FitzPatrick, D. R. Resequencing at scale in neurodevelopmental disorders. *Nat. Genet.* **49**, 488–489 (2017).

Acknowledgements
The authors are all members of the management team of the Deciphering Developmental Disorders (DDD) Study, which undertakes independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003], a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute [grant number WT098051]. The views expressed in this publication are those of the authors and not necessarily those of the Wellcome Trust or the Department of Health. The study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC). The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network. This study makes use of DECIPHER, which is funded by the Wellcome Trust. H.V.F. is supported by The Wellcome Trust award 200990/Z/16/Z 'Designing, developing and delivering integrated foundations for genomic medicine'.

Author contributions
C.F.W. and H.V.F. drafted and revised the manuscript. D.R.F. provided editorial input, designed some of the figures and contributed to the ideas and discussions on which the article is based.

Competing interests statement
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

DATABASES

ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/
DECIPHER: https://decipher.sanger.ac.uk
GeneReviews: https://www.ncbi.nlm.nih.gov/books/NBK1116
Genetics Home Reference: https://ghr.nlm.nih.gov/
ExAC: exac.broadinstitute.org
gnomAD: gnomad.broadinstitute.org
Human Gene Mutation Database: https://www.hgmd.cf.ac.uk
OMIM: https://www.omim.org/
Orphanet: https://www.orpha.net
Unique: https://www.rarechromo.org

FURTHER INFORMATION

Deciphering Developmental Disorders Study: https://www.ddduk.org
Matchmaker Exchange: https://www.matchmakerexchange.org/
Phenome Central: https://www.phenomecentral.org/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF