Next-generation sequencing in childhood disorders

Ricardo Parolin Schnekenberg,[1,2] Andrea H Németh[1,3,4]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
[2]School of Medicine, Universidade Positivo, Curitiba, Brazil
[3]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[4]Department of Clinical Genetics, Churchill Hospital, Oxford University Hospitals NHS Trust, Oxford, UK

Correspondence to
Dr Andrea H Nemeth, 6th Floor West Wing, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Headley Way, Oxford OX39DU, UK;
andrea.nemeth@eye.ox.ac.uk

ABSTRACT

Genetics has been revolutionised by recent technologies. The latest addition to these advances is next-generation sequencing, which is set to transform clinical diagnostics in every branch of medicine. In the research arena this has already been instrumental in identifying hundreds of novel genetic syndromes, making a molecular diagnosis possible for the first time in numerous refractory cases. However, the pace of change has left many clinicians bewildered by new terminology and the implications of next-generation sequencing for their clinical practice. The rapid developments have also left many diagnostic laboratories struggling to implement these new technologies with limited resources. This review explains the basic concepts of next-generation sequencing, gives examples of its role in clinically applied research and examines the challenges of its introduction into clinical practice.

INTRODUCTION

DNA sequencing refers to methods of determining the individual order of bases of the genetic code. In the early 1970s several techniques were developed, including that of Frederick Sanger,[1] known as dideoxy sequencing, which has been the gold standard in clinical laboratories ever since. Sanger sequencing was the basis for the Human Genome Project, which took 13 years of worldwide effort and cost nearly $3 billion to sequence the ~3.2 billion base pairs of human DNA. This achievement has fuelled the demand for large-scale sequencing that cannot be achieved using the dideoxy method, resulting in innovative new sequencing technologies capable of 'massively parallel' analysis.

MASSIVELY PARALLEL SEQUENCING

Sanger sequencing has been immensely successful due to its low error rate and cost effectiveness for small scale projects. However, it is labour intensive as it sequences one individually amplified DNA molecule at a time. In contrast, next-generation sequencing (NGS) simultaneously sequences many different molecules, providing a read-out of each. The major advantage of NGS is that it can perform high-throughput sequencing in a single (albeit large) experiment. Although initially used as a research tool to sequence bacterial genomes, the applications to clinical medicine quickly became apparent. Instead of single-gene analyses 'in series', analysis of multiple genes 'in parallel' is possible. In clinical medicine, phenotypical and genotypical heterogeneity is common and being able to simultaneously and quickly screen multiple genes has the potential to transform the diagnostic process.

However, the preparation of samples for NGS is not as straightforward as for Sanger sequencing.

Figure 1 shows the steps for NGS: patient DNA is extracted from nucleated cells and randomly fragmented usually using sonication or mechanical shearing. Adaptors are then ligated to the fragmented DNA; these adaptors are short oligonucleotides of known sequences that serve as universal priming sites during the amplification and sequencing steps (figure 1A). Commonly the fragments are *enriched* for specific genes of interest (*targeted sequencing*) or for all coding regions (*whole-exome sequencing* (WES)) in a physical *capture* step (figure 1B). In *whole-genome sequencing* (WGS) this capture step is skipped and all fragments are sequenced. Just prior to the sequencing cycles, the fragments are spatially separated and then clonally amplified by PCR in order to generate distinct clusters. The adaptors act as priming sites, directing sequencing inwards from each end. NGS platforms vary considerably, but common to all are multiple wash-and-scan cycles: nucleotides are added, a detectable signal is generated upon nucleotide incorporation to a growing chain and the unincorporated nucleotides are then washed away (figure 1C). Several thousand fragments are simultaneously analysed and decoded for their individual sequences without any information regarding the original position of each fragment in the genome.[2] A major advantage of NGS chemistries are the ability to perform these reactions many times in small volumes. However, a well recognised disadvantage is the poor representation of Guanine-cytosine content (GC)-rich regions.

The raw sequencing data consists of large computer text files (tens to hundreds of gigabytes) containing several million short (~35–400 bp) nucleotide sequences called *reads*. These cryptic files need to undergo complex computational processing in order to become meaningful information. To determine the position of the reads in the genome they must be aligned (*mapped*) to their most probable location on the reference human genome and possible mismatches or gaps must be taken into consideration (figure 1D). The alignment is based solely on their sequence; a complex task when dealing with short reads from a gigantic genome. Ideally, reads should overlap to *cover* each base several times (figure 1D). Following the alignment stage, each nucleotide is compared with its counterpart in the reference genome and recorded, in a process known as *variant calling*. Differences from the reference—mismatches, insertions or gaps—are regarded as variants. At any specific position, a homozygous change would be expected to differ from the reference genome in nearly all the reads, whereas a heterozygous change would be present in only ~50% of reads (figure 1D). Sequencing and mapping are not error-free processes; distinguishing real variants from background noise can be a challenge, hence

Editor's choice
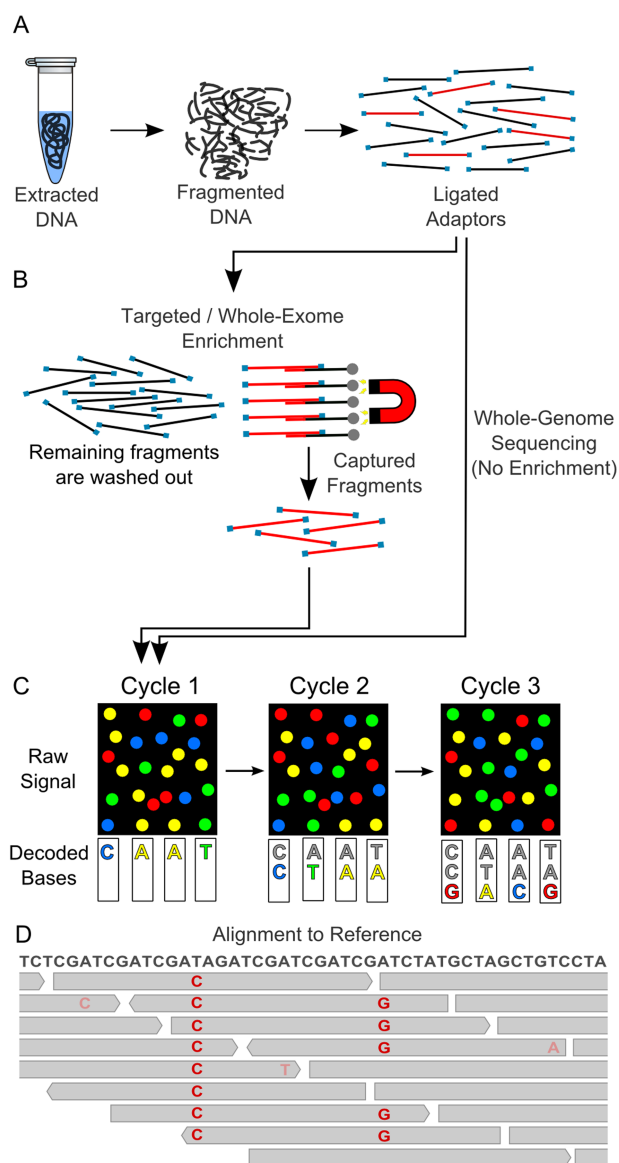Scan to access more free content

Figure 1 Steps in next-generation sequencing. (A) Extracted DNA is randomly broken into <1000 bp fragments. Known adaptor sequences are ligated to fragments. (B) In whole-genome sequencing, all fragments are sequenced, whereas in whole-exome and targeted sequencing only a subset of the original fragment pool is sequenced. (C) An example of a NGS platform (Illumina). It relies on spatial separation of fragments on a slide and clonal amplification by PCR to generate fragment clusters. Four fluorescently labelled nucleotides are added to the slide and compete to be incorporated to the growing chains. In each cycle, the clusters are excited by laser and the emitted fluorescence (colored circles) is recorded by an image-capturing device. As the position of each individual cluster remains fixed, the sequencer creates a 'time lapse' with the recorded images from all cycles, with each cluster generating a read. (D) Individual reads (gray rectangles) aligned to the reference genome. The coverage for each genomic position is the number of reads that overlap at that position. The first base in the figure (T) has 5× coverage; the last base (A) is covered nine times. Bases that match the reference sequence have been omitted. Examples of homozygous and heterozygous single-nucleotide variants are shown (left and right, respectively). Examples of sequencing and/or mapping errors are shown as faded bases.

a high *depth of coverage* (number of different reads that cover a specific base in the genome) is essential for accurate variant calling.

NEXT-GENERATION SEQUENCING STRATEGIES

The human genome is composed of 3.2 billion base pairs. Early in the development of NGS there was immediate recognition that *targeted capture*, in which only genes of interest are sequenced, could be applied in clinical practice to genetically heterogeneous disorders where tens or even hundreds of genes may be involved in a specific condition. Since then, rapid developments in capture designs have enabled all protein-coding regions to be sequenced (WES). However, the capture experiments themselves are time consuming and costly, and coupled with recent steep drops in sequencing costs (figure 2) the emphasis is gradually shifting from targeted sequencing to WES or WGS. The following sections outline the advantages and disadvantages of these different approaches and how they may impact on clinical practice.

Targeted sequencing

Fragments containing genes or chromosomal regions of interest are *captured* as shown in figure 1B. This is usually accomplished by using commercial kits of custom-made short complementary oligonucleotides (*baits*) that bind to fragments containing known target sequences, physically selecting them while unbound fragments are washed out. Designing these baits is done through commercial user-friendly web-based tools that accept a list of genes or regions. Capture efficiency can be variable and results in loss of target sequence information or off-target sequencing. Also, capture methods provide uneven coverage across target regions, requiring samples to be oversequenced to achieve adequate coverage in poorly captured regions. Nevertheless, targeted capture is a good strategy for sequencing a defined fraction of the genome. Its main advantages include:

- ▶ possibility of customisation and optimisation of the target regions;
- ▶ more affordable benchtop sequencers can be used;
- ▶ higher average depth of coverage;
- ▶ simpler Information Tecnology (IT) infrastructure for data processing and analysis;
- ▶ fewer variants to interpret;
- ▶ possibly shorter turnaround time in a diagnostic setting.

Whole-exome sequencing

The majority of disease-causing mutations are located in protein-coding regions of the genome (the *exome*), which represent less than 2% of the total. Thus, by capturing and sequencing only the exome, the focus is on regions most likely to harbour pathogenic mutations. This has proved extraordinarily successful in finding novel disease genes,[3–5] but it relies heavily on data filtering, as one patient's exome will output ~30 000 variants. Current commercial whole-exome enrichment kits capture between 200 000 exons and 300 000 exons, which corresponds to about 21 000 genes and 50–100 Mb genomic size, depending on the extent of extra information captured (eg, untranslated regions, flanking regions, microRNAs). WES was initially mainly reported in research projects, but the genetic and phenotypical heterogeneity of human disease make it attractive for clinical diagnostics, and some specialised laboratories already offer it commercially. However, routine analysis, storage and interpretation of such amounts of data are beyond the means of many clinical diagnostic laboratories without significant development of infrastructure
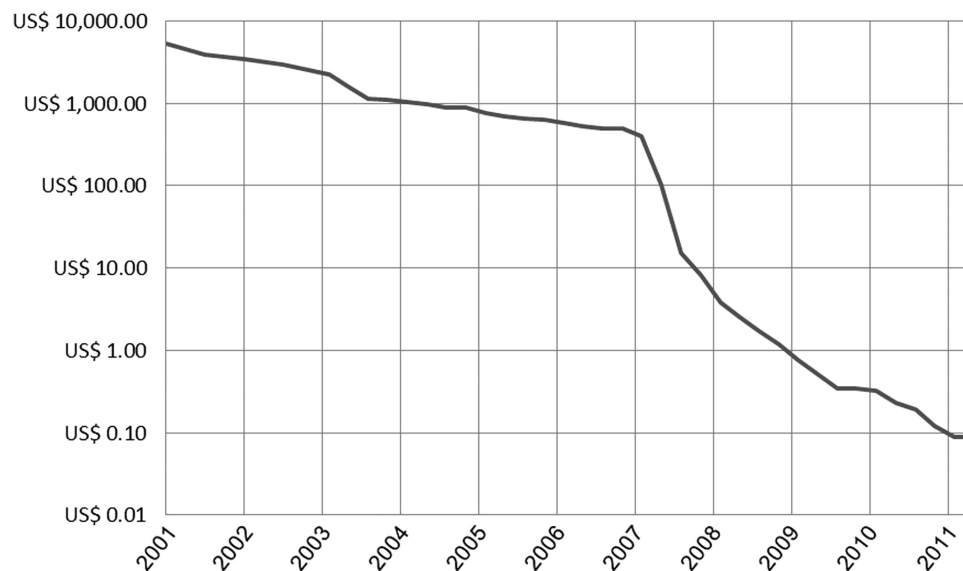
Figure 2 Costs per raw megabase of DNA sequence. Note the logarithmic scale of costs and steep drop after introduction of NGS in 2007. 1 megabase (Mb)=1 000 000 base-pairs. Data obtained from http://www.genome.gov/sequencingcosts.

and training. Furthermore, experience with WES suggests that, for a variety of reasons, there is incomplete capture of target regions, with as many as ~40% of targeted bases and ~20% of known disease-causing sites poorly covered (<20× and <10×, respectively).[6] [7] Therefore, without improvements, current enrichment methods for WES may limit its use in some diagnostic settings where false negative results can be disastrous.

Whole-genome sequencing
WGS aims to sequence all bases in the genome. An average of 30-fold coverage is desirable for downstream bioinformatics analysis, so WGS currently represents an expensive ultra-high-throughput option, since the total data produced is in excess of 100 Gb. Since there are over 3 million variants in an individual's genome, substantial IT infrastructure and staff are required to transfer and safely store this data, and bioinformatic analysis is slow and intricate. However, WGS offers a resolution of the genome that is unmatched by other sequencing methods. It allows the study of coding (<2%) and non-coding variation (>98%), and the latter is increasingly thought to be a rich source of disease-associated variation.[8–10] The absence of the capture step leads to uniform coverage, which reduces the average depth of coverage required for accurate and confident variant calling. This also facilitates searches for structural and copy-number variants, known disease-causing mechanisms. The scale of analysis in WGS produces numerous variants of uncertain significance and requires longer times for analysis. However, the advantages offered suggest that it will replace other sequencing methods in research and diagnostics within the next few years.

DATA INTERPRETATION
The vast amount of data generated by NGS creates major analysis and storage challenges, which greatly exceeds most desktop solutions and therefore requires dedicated storage facilities and IT expertise. Furthermore, bioinformatics is an emerging academic discipline in need of new training programs as supply of professionals currently falls short of demand. Coupled with our limited understanding of normal genetic variation, narrowing down several hundred thousand variants to a specific disease-causing one remains a significant challenge in research and clinical settings.

Filtering and interpretation of pathogenicity
Once reads are aligned and variants called, the data must be interpreted. This data-set contains a list of variants that can range from a few hundred, in small targeted-capture experiments, to many thousands (WES) or millions, in WGS. To determine which variants might be of clinical interest, the list must be filtered to produce a manageable number that can be inspected for causality.

The filtered variants are usually listed in large spreadsheets and annotated using information that provides evidence for pathogenicity. Each variant must be individually analysed to determine whether it is considered clearly benign, clearly pathogenic or unclassified. The most difficult variants to assess are missense mutations, which can have greatly varying effects on different proteins. Some pathogenicity criteria are shown in Box 1.

Box 1   Evidence for pathogenicity of filtered variants

▶ Previous reports of the mutation in curated mutation or literature databases (eg, Human Genome Mutation Database, Online Mendelian Inheritance in Man).
▶ Allele frequency data (eg, deposition in dbSNP, 1000 Genomes Project or Exome Variant Server): the more common an allele the less likely it is to be causal in a rare disease.
▶ Literature support (eg, animal models).
▶ Absence in ethnically matched controls.
▶ Cosegregation with the disease in a family.
▶ Identification of a de novo variant in a sporadic condition.
▶ Evolutionary conservation (nucleotide and amino acid residue).
▶ Large physicochemical distance in a missense amino acid change (Grantham score).
▶ In silico prediction of effect on splicing.
▶ In silico prediction of deleteriousness.

A pragmatic approach to pathogenicity determination of missense mutations includes assessment of the frequency of a variant using multiple variant databases such as DMuDB (https://secure.dmudb.net), Exome Variants Server,[11] 1000 Genomes,[12] dbSNP,[13] HGMD,[14] DECIPHER (http://decipher.sanger.ac.uk) and inhouse databases. These databases give frequency data: common variants are likely to be polymorphisms and therefore unlikely to be pathogenic. Specific pathogenicity assessment programs include: PolyPhen2,[15] SIFT,[16] MutPred,[17] and MutationTaster which use algorithms to predict possible functional effects of missense changes. However, it is important to recognise that filtering approaches reflect our current knowledge of benign and pathogenic genetic variation. Therefore it is possible that some truly causal mutations are filtered out because they defy established pathogenicity models.[5 18–20] It is also clear that the genome is more 'tolerant' of mutations than previously thought: a study using whole-genome sequence data from 185 individuals has estimated that healthy humans typically have ~100 loss-of-function variants, including ~20 homozygous variants leading to complete gene inactivation.[21] Therefore, establishing causality of a novel variant in an individual case may require functional studies, animal models and analysis of multiple patients, all of which are beyond the scope of most diagnostic laboratories.

Variants then need to be validated using Sanger sequencing. There is relatively little data comparing accuracy of NGS with dideoxy sequencing, and most data comes from genome-wide comparison of single nucleotide polymorphism (SNP) concordance across different platforms. However, the available evidence suggests that the error rate from NGS is low, but not negligible. As a result all NGS data must be confirmed using a different technology and at the moment dideoxy sequencing remains the most accurate, rapid and cost-effective means of doing this. Box 2 lists some well recognised causes for false positive and false negative results using NGS. Of particular note is that low depth of coverage (read numbers) can be related to capture inefficiency which can in turn lead to missing data and result in errors in mutation detection.

The point at which the filtered list of variants is small enough to spend time validating them is a significant issue, particularly for diagnostics laboratories (see below) as each variant needs specific primers to be designed and the variant sequenced in patient and controls. Once validation is completed, the final list of variants must be interpreted to produce a report. This is also a significant challenge and practice guidelines have been issued about this providing further detail about validation, data analysis and principles of reporting data.[27]

Box 2   Sources of uncertainty in next-generation sequencing data

▶ Capture efficiency.[6 7 22]
▶ Sequencing chemistry errors (high GC content, homopolymer tracts, short-reads, erroneous base incorporation).
▶ Alignment errors (short reads, errors in reference genome).
▶ Low depth of coverage (inefficient capture as above, platform capacity).
▶ Bioinformatic pipeline[23] (filtering algorithms).
▶ Pathogenic variants in benign-variation databases.[24]
▶ Benign variants flagged as pathogenic.[25 26]
▶ Filtering strategy.

HIGHLIGHTS OF NGS RESEARCH

The contribution of NGS to genetic disease research is indisputable. New Mendelian syndromes have been identified, new disease genes discovered and even new mechanisms of pathogenicity described. For the clinician, two main points recur in the NGS literature:

Phenotypical variability

Wide variations in phenotypes have been reported in NGS studies and sometimes overturned previous clinical diagnoses. For example, Pitt-Hopkins syndrome was diagnosed using WES when this diagnosis was previously dismissed because two of the most characteristic features—hyperventilation (86% of reported cases) and epilepsy (70%)—were lacking.[28] In another example, a patient initially diagnosed with ataxia with vitamin E deficiency (OMIM 277460) was found to have hereditary spastic paraplegia with thin corpus callosum (OMIM 604360) after WES found a homozygous mutation in SPG11.[29] In this case, a clinical review concluded that the clinical signs had been misinterpreted, imaging studies had missed the thin corpus callosum and the low vitamin E level had been a false-positive result. All experienced clinicians recognise the limitations of the diagnostic process and such studies illustrate that NGS may be an immensely useful tool in our quest for diagnostic accuracy.

De novo variants are a common cause of human disease

Unexpectedly, it has been shown recently that *de novo* mutations explain a significant proportion of sporadic disorders and are strongly related to the paternal age at conception.[30] Using WES, a study found 6/10 cases of mental retardation that were likely caused by *de novo* point mutations.[31] Another study found that 5/12 cases of various undiagnosed genetic conditions were also due to *de novo* mutations.[28] Other conditions that have been associated with or caused by *de novo* mutations include autistic-spectrum disorders,[32] schizophrenia[33] and Mendelian phenotypes including alternating hemiplegia of childhood,[34] Kabuki,[3] Weaver,[35] Baraitser-Winter[36] and Wiedemann-Steiner syndromes.[37] In conditions where the affected individual may reproduce this has significant implications, for example in retinitis pigmentosa, where *de novo* mutations have substantially altered the offspring risk for an affected individual from very low to 50%.[24 38]

NEXT-GENERATION DIAGNOSTICS

The classic approach to genetic diagnosis is based on clinical phenotyping followed by genetic testing, which is almost always performed on an individual gene basis, starting with the most likely gene to explain the phenotype, usually at the discretion of the clinician. For genetically heterogeneous conditions this approach is costly, time-consuming and inefficient. Next-generation sequencing allows a parallel sequencing strategy at a much lower cost per base and has the potential to increase diagnostic yield and reduce overall cost and time to diagnosis.

Gene panels have already been designed for targeted sequencing in several genetically heterogeneous disorders. For instance, hearing loss is associated with over 60 causal genes, whereas Sanger sequencing is generally offered for only a few. A recently published study used targeted NGS to sequence 34 autosomal recessive deafness genes, achieving a genetic diagnosis in 9/24 patients.[39] Other examples include retinitis pigmentosa,[24 38 40–42] Usher syndrome,[43 44] inherited arrhythmias,[45] congenital muscular dystrophy,[46] mitochondrial diseases[47 48] and ataxia.[49]

In cases of non-specific phenotypes or when serial testing or targeted panels have failed to reach a diagnosis, WES is quickly being developed as a diagnostic option. In addition, several reports of clinical applications and novel genes and syndromes being identified using WGS suggest that this will become a financially viable option in the near future.[50–53]

ETHICAL ISSUES

All genetic testing requires consideration of potential ethical issues and the consent process is designed to address these prior to testing. The types of issues have not been fundamentally altered with the advent of NGS, but the scale of likely problems has vastly increased. Ten specific issues have recently been identified and published in a study addressing informed consent for WGS studies.[54] Among the most important are the identification and management of a range of possible findings (table 1).

Consent and reporting

Although our understanding of the genome is increasingly sophisticated, unclassified variants and incidental findings are a common feature of NGS analysis. There is no consensus yet about how to report these and the consent process must address this prior to embarking on NGS in any setting where the results will be given to patients. Pragmatic solutions include restricting analysis to known genes and/or reporting only those variants with potentially medically actionable consequences (eg Groups 2–4). Specific care needs to be taken for children who should be entitled to an 'open future' and therefore be allowed, when an appropriate age for informed decision is reached, to choose *not to know* their genetic make-up.

FUTURE DIRECTIONS
Routine diagnostics

Although there are numerous published research examples of using NGS as a diagnostic tool, introducing NGS into routine diagnostic laboratories remains a challenge. Some difficulties have already been mentioned, such as error rates and

interpretation of variants. Other challenges for NGS diagnostics include difficulties with GC-rich genes, for example RPGR, which must be separately sequenced using standard methods[42] and trinucleotide repeat genes (eg, those causing Huntington's disease, Friedreich's ataxia and others) which are unsuitable for short read NGS at the moment. Strategic difficulties include development costs and infrastructure that can be prohibitive for many diagnostic laboratories, particularly those which are publicly funded. In addition, the rapid rate of change in technologies in such a short space of time has meant that establishing best practice for the diagnostic sector, which by definition requires accuracy, has not yet been straightforward. In particular, diagnostic laboratories, with stringent quality assurance must validate such tests before offering them as a clinical service and this requires clear evidence of accuracy, cost effectiveness, mechanisms for interpreting and reporting unclassified variants and investment in bioinformatics infrastructure. Although there are several reports of developing NGS for clinical diagnostics, only a small fraction of laboratories are currently offering this on a service basis, although it is likely to expand significantly in the next few years.

Data interpretation

Current analysis pipelines are computationally demanding, complex and not user-friendly. When raw sequencing files are analysed using different software or customised scripts, the list of variants produced is frequently different: there is an urgent need for reproducibility and standardisation for the data processing and analysis pipelines in the research and clinical settings.[23] In addition, current capabilities in calling small insertions or deletions are not yet ideal. The detection of structural (eg, inversions or complex rearrangements) and copy-number variants (increase or decrease from diploid genome) is theoretically possible with NGS,[55–57] but requires significant improvements before it can replace current diagnostic technology such as array-comparative genome hybridisation (CGH), SNP arrays, multiplex ligation-dependent probe amplification (MLPA) and fluorescent in situ hybridisation (FISH).

New platforms and chemistries

Third generation sequencing is now being developed, with improved chemistries and lower per base costs. Several companies have reported promising novel technologies, for example the PacBio RS uses Single Molecule Real Time (SMRT) technology to observe DNA polymerases actively incorporating fluorescent-tagged nucleotides to a single-stranded DNA template molecule.[58 59] Another promising technology is nanopore sequencing: molecules of DNA are moved through a biological nanopore and measurable changes in voltage are detected for each sequential nucleotide.[60]

However, accuracy of third generation sequencing has not yet been determined and this data will be crucial before introduction to a diagnostic setting. Can NGS ever entirely replace dideoxy sequencing? At the present time the answer is 'no'. This is because NGS chemistries are not yet as accurate as dideoxy sequencing. In addition, for small scale tests (eg in cases of a known family mutation for carrier testing, confirmatory diagnostic testing or presymptomatic testing), dideoxy sequencing remains faster, cheaper and more accurate. In the case of screening patients with known disorders for a panel of mutations, NGS is likely to gradually supersede other technologies as the overall preparation times and sequencing costs reduce. A major difficulty has been the identification of copy number variants or repetitive sequences using short sequence

Table 1	Possible findings of genome-wide sequencing
Group 1	Genetic findings useful for the original clinical or research question
Group 2	Any clinically relevant genetic findings, which may have immediate benefits for the patient related to present diseases or clinical conditions
Group 2A	Diseases for which possible treatments are available (eg, cardiovascular diseases predisposing to sudden cardiac death)
Group 2B	Diseases for which no available treatment exists (eg, Charcot–Marie–Tooth disease type 1A)
Group 3	Genetic mutations related to high risks for future Mendelian diseases
Group 3A	Information about risks of preventable or treatable diseases (eg, Lynch syndrome or BRCA1/2)
Group 3B	Information about risks of non-preventable, non-treatable future diseases (eg, Huntington's disease)
Group 4	Information about carrier status of mutations for a X linked or an autosomal recessive disorder impacting reproductive life decisions (eg, Tay-Sachs disease, cystic fibrosis)
Group 5	Information of variable risk for future diseases. Genetic traits that may be translated into high predisposition for certain complex diseases (eg, ApoE4 and Alzheimer's disease). Most pharmacogenetic variants (eg, β-blockers and β1-adrenergic receptor)
Group 6	Information of unknown significance

Adapted with permission from Ayuso *et al*.[54]

NGS and there are numerous research groups trying to address these issues to provide more comprehensive diagnostic solutions.[61]

SUMMARY

Next-generation sequencing offers the potential to profoundly alter diagnostics and investigation of the genomic contribution to human disease, but many challenges remain to ensure that it is used accurately and ethically in clinical practice. Although it is already being introduced, NGS will require significant changes to current delivery of diagnostic services including an understanding of it by all clinicians.

REFERENCES

1 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;74:5463–7.
2 Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet* 2010;11:31–46.
3 Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;42:790–3.
4 Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;42:30–5.
5 Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;12:745–55.
6 Sulonen AM, Ellonen P, Almusa H, et al. Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol* 2011;12:R94.
7 Asan , Xu Y, Jiang H, et al. Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol* 2011;12:R95.
8 Jarinova O, Ekker M. Regulatory variations in the era of next-generation sequencing: implications for clinical molecular diagnostics. *Hum Mutat* 2012;33:1021–30.
9 Qureshi IA, Mehler MF. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat Rev Neurosci* 2012;13:528–41.
10 Talkowski ME, Maussion G, Crapper L, et al. Disruption of a large intergenic noncoding RNA in subjects with neurodevelopmental disabilities. *Am J Hum Genet* 2012;91:1128–34.
11 Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. http://evs.gs.washington.edu/EVS/
12 Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
13 Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11.
14 Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2013 Sep 28. [Epub ahead of print]
15 Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
16 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073–81.
17 Li B, Krishnan VG, Mort ME, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 2009;25:2744–50.
18 Foo JN, Liu JJ, Tan EK. Whole-genome and whole-exome sequencing in neurological diseases. *Nat Rev Neurol* 2012;8:508–17.
19 Majewski J, Schwartzentruber J, Lalonde E, et al. What can exome sequencing do for you? *J Med Genet* 2011;48:580–9.
20 Gilissen C, Hoischen A, Brunner HG, et al. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 2012;20:490–7.
21 MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012;335:823–8.
22 Bell J, Bodmer D, Sistermans E, et al. Practice guidelines for the Interpretation and Reporting of Unclassified Variants (UVs) in Clinical Molecular Genetics. 2007. http://www.cmgs.org/BPGs/Best_Practice_Guidelines.htm (accessed 3 Feb 2013).
23 Wooderchak-Donahue WL, O'Fallon B, Furtado LV, et al. A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel-clinical diagnostics perspective. *BMC Med Genomics* 2012;5:50.
24 Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nat Rev Genet* 2012;13:667–72.
25 Shanks ME, Downes SM, Copley RR, et al. Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease. *Eur J Hum Genet* 2013;21:274–80.
26 Bell CJ, Dinwiddie DL, Miller NA, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* 2011;3:65ra4.
27 Davies WI, Downes SM, Fu JK, et al. Next-generation sequencing in health-care delivery: lessons from the functional analysis of rhodopsin. *Genet Med* 2012;14:891–9.
28 Need AC, Shashi V, Hitomi Y, et al. Clinical application of exome sequencing in undiagnosed genetic conditions. *J Med Genet* 2012;49:353–61.
29 Dixon-Salazar TJ, Silhavy JL, Udpa N, et al. Exome sequencing can improve diagnosis and alter patient management. *Sci Transl Med* 2012;4:138ra78.
30 Kong A, Frigge ML, Masson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 2012;488:471–5.
31 Vissers LE, de Ligt J, Gilissen C, et al. A de novo paradigm for mental retardation. *Nat Genet* 2010;42:1109–12.
32 Yu TW, Chahrour MH, Coulter ME, et al. Using Whole-Exome Sequencing to Identify Inherited Causes of Autism. *Neuron* 2013;77:259–73.
33 Xu B, Roos JL, Dexheimer P, et al. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* 2011;43:864–8.
34 Heinzen EL, Swoboda KJ, Hitomi Y, et al. De novo mutations in ATP1A3 cause alternating hemiplegia of childhood. *Nat Genet* 2012;44:1030–4.
35 Gibson WT, Hood RL, Zhan SH, et al. Mutations in EZH2 cause Weaver syndrome. *Am J Hum Genet* 2012;90:110–18.
36 Riviere JB, van Bon BW, Hoischen A, et al. De novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. *Nat Genet* 2012;44:440–4.
37 Jones WD, Dafou D, McEntagart M, et al. De novo mutations in MLL cause Wiedemann-Steiner syndrome. *Am J Hum Genet* 2012;91:358–64.
38 Neveling K, Collin RW, Gilissen C, et al. Next-generation genetic testing for retinitis pigmentosa. *Hum Mutat* 2012;33:963–72.
39 Schrauwen I, Sommen M, Corneveaux JJ, et al. A sensitive and specific diagnostic test for hearing loss using a microdroplet PCR-based approach and next generation sequencing. *Am J Med Genet A* 2013;161:145–52.
40 Audo I, Bujakowska KM, Leveillard T, et al. Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J Rare Dis* 2012;7:8.
41 Bowne SJ, Sullivan LS, Koboldt DC, et al. Identification of disease-causing mutations in autosomal dominant retinitis pigmentosa (adRP) using next-generation DNA sequencing. *Invest Ophthalmol Vis Sci* 2011;52:494–503.
42 O'Sullivan J, Mullaney BG, Bhaskar SS, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J Med Genet* 2012;49:322–6.
43 Le Quesne Stabej P, Saihan Z, Rangesh N, et al. Comprehensive sequence analysis of nine Usher syndrome genes in the UK National Collaborative Usher Study. *J Med Genet* 2012;49:27–36.
44 Licastro D, Mutarelli M, Peluso I, et al. Molecular diagnosis of Usher syndrome: application of two different next generation sequencing-based procedures. *PLoS One* 2012;7:e43799.
45 Ware JS, John S, Roberts AM, et al. Next generation diagnostics in inherited arrhythmia syndromes: a comparison of two approaches. *J Cardiovasc Transl Res* 2013;6:94–103.
46 Valencia CA, Rhodenizer D, Bhide S, et al. Assessment of target enrichment platforms using massively parallel sequencing for the mutation detection for congenital muscular dystrophy. *J Mol Diagn* 2012;14:233–46.
47 Calvo SE, Compton AG, Hershman SG, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci Transl Med* 2012;4:118ra10.
48 Cui H, Li F, Chen D, et al. Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. *Genet Med* 2013;15:1–7.
49 Németh AH, Kwasniewska AC, Lise S, et al. Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. *Brain* 2013;136(Pt 10):3106–18.
50 Lise S, Clarkson Y, Perkins E, et al. Recessive mutations in SPTBN2 implicate beta-III spectrin in both cognitive and motor development. *PLoS Genet* 2012;8:e1003074.
51 Bainbridge MN, Wiszniewski W, Murdock DR, et al. Whole-genome sequencing for optimized patient management. *Sci Transl Med* 2011;3:87re3.
52 Saunders CJ, Miller NA, Soden SE, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med* 2012;4:154ra35.

53 Talkowski ME, Ordulu Z, Pillalamarri V, *et al*. Clinical diagnosis by whole-genome sequencing of a prenatal sample. *N Engl J Med* 2012;367:2226–32.

54 Ayuso C, Millan JM, Mancheno M, *et al*. Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. *Eur J Hum Genet* 2013;21:1054–9.

55 Fromer M, Moran JL, Chambert K, *et al*. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet* 2012;91:597–607.

56 Krumm N, Sudmant PH, Ko A, *et al*. Copy number variation detection and genotyping from exome sequence data. *Genome Res* 2012;22: 1525–32.

57 Plagnol V, Curtis J, Epstein M, *et al*. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* 2012;28:2747–54.

58 Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum Mol Genet* 2010;19:R227–40.

59 Thompson JF, Milos PM. The properties and applications of single-molecule DNA sequencing. *Genome Biol* 2011;12:217.

60 Branton D, Deamer DW, Marziali A, *et al*. The potential and challenges of nanopore sequencing. *Nat Biotechnol* 2008;26:1146–53.

61 Loomis EW, Eid JS, Peluso P, *et al*. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. *Genome Res* 2013;23:121–8.