What can exome sequencing do for you?

Jacek Majewski,[1,2] Jeremy Schwartzentruber,[1] Emilie Lalonde,[1,2]
Alexandre Montpetit,[1] Nada Jabado[2,3]

[1]McGill University and Genome
Quebec Innovation Centre,
Montreal, Canada
[2]Department of Human
Genetics, McGill University,
Montreal, Canada
[3]Department of Pediatrics,
McGill University Health Center
Research Institute, Montreal,
Canada

Correspondence to
Dr Nada Jabado, Montreal
Children's Hospital Research
Institute, 4060 Ste Catherine
West, PT-239, Montreal,
Quebec H3Z 2Z3, Canada;
nada.jabado@mcgill.ca

ABSTRACT

Recent advances in next-generation sequencing
technologies have brought a paradigm shift in how
medical researchers investigate both rare and common
human disorders. The ability cost-effectively to generate
genome-wide sequencing data with deep coverage in
a short time frame is replacing approaches that focus on
specific regions for gene discovery and clinical testing.
While whole genome sequencing remains prohibitively
expensive for most applications, exome sequencing—a
technique which focuses on only the protein-coding
portion of the genome—places many advantages of the
emerging technologies into researchers' hands. Recent
successes using this technology have uncovered genetic
defects with a limited number of probands regardless of
shared genetic heritage, and are changing our approach
to Mendelian disorders where soon all causative
variants, genes and their relation to phenotype will be
uncovered. The expectation is that, in the very near
future, this technology will enable us to identify all the
variants in an individual's personal genome and, in
particular, clinically relevant alleles. Beyond this, whole
genome sequencing is also expected to bring a major
shift in clinical practice in terms of diagnosis and
understanding of diseases, ultimately enabling
personalised medicine based on one's genome. This
paper provides an overview of the current and future use
of next generation sequencing as it relates to whole
exome sequencing in human disease by focusing on the
technical capabilities, limitations and ethical issues
associated with this technology in the field of genetics
and human disease.

INTRODUCTION

Discoveries made in the 20th century have helped
to completely reshape all fields of biomedical
studies as we know them. The revolution we are
currently witnessing was triggered by the discovery
of the DNA double helix in 1953[1 2] which has
enabled major advances in genetics and heredity.
Advances in knowledge have often been driven by
the advent of new technologies. PCR was discovered
in 1983[3] and revolutionised our approach
to the study of DNA, and that in turn revolutionised
the molecular analysis of mammalian genes. In
1977 two landmark articles describing methods for
DNA sequencing were published.[4 5] The approach
reported by Sanger and colleagues was further
refined and commercialised leading to its dissemination
throughout the research community and,
ultimately, into clinical diagnostics. In an industrial
high-throughput configuration, Sanger technology
was then used in the sequencing of the first human
genome which was completed in 2001 through
the Human Genome Project, a 13-year effort with

an estimated cost of $2.7 billion.[6–8] In 2008, by
comparison, a human genome was sequenced over
a 5-month period for approximately $1.5 million[9]
and now, in 2011, sequencing a whole genome is
done over a period of a few days and is soon
projected to cost less than $10 000. The latter
accomplishment was made possible by the commercial
launch of the first massively parallel pyrosequencing
platform in 2005, which ushered in the era
of high-throughput genomic analysis now referred to
as 'next-generation sequencing' (NGS). The time and
cost needed is expected to fall even further in the very
near future, making these technologies available to
researchers without large budgets. Vast amounts of
genotype and phenotype data are now being generated
by a growing number of research efforts. Interpreting
all these new data and translating the
findings to practical healthcare is a challenge. This
review focuses on what whole exome sequencing
(WES) using NGS can teach us about human disease,
starting with single gene disorders and moving on to
more complex genetic disorders including complex
traits and cancer. We overview the current state of
the technology, its practical current and future uses,
the reasons to use WES instead of whole genome
sequencing (WGS) and some of the ethical dilemmas
arising from the impact of the results on society and
on clinical practice.

NEXT-GENERATION SEQUENCING (NGS) TECHNOLOGIES AND EXPERIMENTAL APPROACHES FOR WHOLE EXOME SEQUENCING (WES)

NGS platforms share a common technological
feature—namely, massively parallel sequencing of
clonally amplified or single DNA molecules that are
spatially separated in a flow cell. This design is
a paradigm shift from that of Sanger sequencing
and has allowed scaling-up by orders of magnitude.
In NGS, sequencing is performed by repeated cycles
of polymerase-mediated nucleotide extensions or,
in the case of ABI SOLiD, by iterative cycles of
oligonucleotide ligation. As a massively parallel
process, NGS generates hundreds of megabases to
gigabases of nucleotide sequence from a single
instrument run, depending on the platform.
Targeted sequencing approaches have the general
advantage of increased sequence coverage of regions
of interest—such as coding exons of genes—at
lower cost and higher throughput compared
with random shotgun sequencing methods.[9–12]
Most large-scale methods for targeted sequencing
use a variation of a hybrid selection approach.
Complementary nucleic acid 'baits' are used to 'fish'
for regions of interest in the total pool of nucleic
acids, which can be DNA[12–15] or RNA.[16] Any
subset of the genome can be targeted, including

exons, non-coding RNAs, highly conserved regions of the genome, disease-associated LD blocks or other regions of interest. Because the exome represents only approximately 1% of the genome or about 30 Mb, vastly higher sequence coverage can be readily achieved using second-generation sequencing platforms with considerably less raw sequence and cost than WGS. For example, whereas 90 Gb of sequence is required to obtain 30-fold average coverage of the genome, 75-fold average coverage is achieved for the exome with only 3 Gb of sequence using the current state-of-the-art platforms for targeting.[17] [18] However, there are inefficiencies in the targeting process. For example, uneven capture efficiency across exons can result in exons with low sequence coverage, and off-target hybridisation means that at least 20% of reads come from genomic DNA outside the exome. In addition, exome capture is not complete. Indeed, the probes in sequence capture methods are designed based on information from gene annotation databases such as the consensus coding sequence (CCDS) Database and RefSeq Database. Therefore, unknown or yet-to-be-annotated exons, evolutionary conserved non-coding regions and regulatory sequences such as enhancers or promoters are not typically captured. Partly to address these issues of coverage, the latest commercially available capture kits provide nearly complete coverage of the well-annotated genes but also allow the user to add custom content by designing capture probes targeted to additional regions of interest such as promoters or highly conserved sequences. Newer kits have also expanded the regions captured to include micro-RNA sites and untranslated regions of genes, thus increasing the regions captured from ∼30 Mb to as high as 62 Mb (table 1).

An important consideration is that NGS technologies have higher base calling error rates than Sanger sequencing, although this can be remedied to some extent by increasing the depth of sequencing coverage to ensure minimal 'false calls'.[19] This makes the resequencing of mutant or variant genes using conventional sequencing techniques important for validation and increases the cost of the approach. All of these inefficiencies are likely to be ameliorated as sequencing and capture technology continue to improve. Importantly, the higher coverage of the exome that can be affordably achieved for a large number of samples makes exome sequencing highly suitable for mutation discovery and its use is becoming increasingly routine.

WES IN HUMAN DISEASE

"Why me?" "Why my child?" "What could I have done to avoid this?" "What can I do to be cured or to get better?" These questions routinely face clinicians caring for patients with a given disease, and they are even questions we ask ourselves as medical knowledge expands with the rapid advances in genome sequencing. They pertain to any human disease as all have a genetic component—major or minor—or, for some, they simply reflect the wish to know what genetic information they are born with. Until 2008 such questions usually remained unanswered because, even for some Mendelian disorders, we did not know how to identify causative mutations within our genome. Indeed, our individual genomes contain variants which may protect us against or increase our susceptibility to our environment and the multiple stressors we encounter during our lifespan. Knowing these variants can perhaps allow us to better prepare or to avoid the negative impacts they might have on our health, lifespan and offspring. In the short years since the first commercial platform became available, NGS has dramatically accelerated multiple areas of genomics research, enabling experiments that previously were not technically feasible or affordable. In this paper we describe the major ongoing applications of NGS as they pertain to WES.

Genetic variants identified using WES

Genetic variants induce a phenotype that can vary between individuals in penetrance or physiological effect and may depend on (1) environmental factors; (2) modifier genes and/or the epigenome; and (3) the additive/synergistic effect from another genetic variant (digenic inheritance).[20] High-penetrance variants induce a strong physiological effect and thus these alleles have usually been identified as causative for Mendelian monogenic disorders using linkage studies in families (see below). Low-penetrance variants have a weak phenotype and causative alleles are typically identified in large case/control cohorts as part of the study of complex trait disorders. Common high-penetrance disease variants are rare as they would normally be eliminated from breeding populations, except in cases of balancing selection where heterozygotes have an advantage over homozygotes (eg, sickle cell disease and resistance to cerebral malaria). The distinction between monogenic and complex diseases is therefore operational as all genetic variants are de facto transmitted in a Mendelian fashion and thus amenable to discovery using WES, giving this technique far wider applicability than simply the study of rare monogenic disorders (figure 1).

WES in characterising monogenic (Mendelian) disorders

Uncovering genetic defects underlying monogenic inherited disorders is one of the most obvious applications of WES. Single-gene disorders, while individually rare, are in aggregate numerous and have an enormous impact on the well-being of affected patients. To date, the gene responsible for the disease in more than 3000 Mendelian disorders has not yet been uncovered. Although rates of spontaneous mutation in the human genome have been estimated in various ways, it is clear that worldwide the entire human gene repertoire is bombarded with
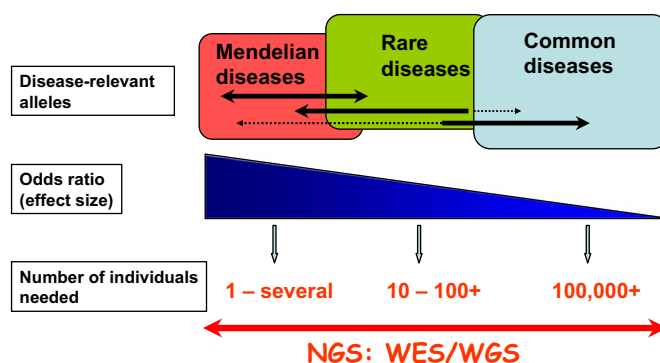
Table 1 Comparison of commercially available technologies for human whole exome capture (numbers taken from their respective data sheets)

Probe size	Agilent 120 bases	NimbleGen 55—105 bases	Illumina 95 bases
Target region size	50 Mb	44.1 Mb	62 Mb
Probe type	RNA	DNA	DNA
No of probes	561 824	2.1M	340 427
No of targeted exons	188 260	∼300 000	201 121
Reads on target (%)	>65	>65	>65
On targets ±200 bases (%)	77	83	80
Bases with >0.2× mean read coverage (%)	>80	>80	>80

Figure 1 Roadmap for the application of next generation sequencing technologies for the identification of disease-relevant genomic variations.

new pathogenic alleles each year.[21] The number of known mutations in human nuclear genes that underlie or are associated with human inherited disease now exceeds 100 000 in more than 3700 different genes (Human Gene Mutation Database). However, for a variety of reasons this figure probably represents only a small fraction of the clinically relevant genetic variants in the human genome.

NGS brings new ways of addressing monogenic disorders. Classical strategies involved using linkage analysis in families with known shared genetic heritage, identifying candidate genomic regions enclosing the gene with the causative mutation, narrowing the interval whenever possible with additional families/probands and thereafter either implementing a candidate gene approach or systematically sequencing the genes located within the interval. The advent of single nucleotide polymorphism (SNP) arrays, which can identify regions of homozygosity within a genome, helped significantly to hasten the linkage analysis by narrowing the regions of interest for further directed sequencing. However, these approaches are costly and time-consuming and their success in identifying causative genetic variants has been variable, mainly due to the small numbers of available affected individuals for a given Mendelian disease and possibly also due to locus heterogeneity.[22–25] Deep resequencing of all human genes for discovery of allelic variants could potentially identify the gene underlying any given rare monogenic disease where a shared genetic heritage is not readily available.[11]

Protein-coding genes constitute only approximately 1% of the human genome but harbour about 85% of the mutations with large effects on disease-related traits. Indeed, most Mendelian disorders are caused by exonic mutations or splice-site mutations that change the amino acid sequence of the affected gene. In contrast to the more laborious approach of SNP homozygosity mapping, the exome sequencing approach is faster, does not depend on shared allelic heritage and can be done in the presence of allelic heterogeneity. Instead, its success depends on the mutation being present in the captured portion of the genome and on our ability to identify it as a pathogenic variant among the many thousands of new variants detected in each exome (the 'background noise'). Strategies to identify these mutations using exome sequencing generally rely on certain assumptions. First, homozygous or heterozygous mutations in only a single gene are required to cause the disease and these mutations will be extremely rare (eg, only present in affected individuals). Second, these mutations have a large effect size, are highly penetrant and, as such, are assumed to affect the protein sequence (ie, non-synonymous SNPs, insertions/deletions or splice-site mutations). The main strategy employed to identify causative mutations is therefore to find all variants in the exome and apply various filters based on the assumptions above. Additional filters are also required to remove false positives caused by systematic, sequencing and misalignment errors. However, the development of bioinformatics analysis tools and the availability and rapidly decreasing cost of NGS technology render exome sequencing simpler and faster than homozygosity mapping. Both alternatives are still complementary as homozygosity mapping can allow us to focus rapidly on the variants most likely to be causative, which can now be identified in record time and with a small number of affected individuals using WES, a cost-effective, reproducible and robust strategy.

The reported successes using NGS have increased exponentially since its first application in 2008 and have frequently been achieved using a limited number of patients.[26–28] Identification of genetic defects in autosomal recessive diseases was performed using either unrelated individuals[10 29–34] and/or individuals from the same families[35–42] and was, in some instances, coupled to homozygosity mapping.[43–45] Similar success was achieved for autosomal dominant disorders.[46 47] WES has even been able to identify the causative mutation in diseases with genetic and phenotypic heterogeneity.[30 34 48 49] Such heterogeneity would make identification of the causative mutation very difficult, if not impossible, by traditional linkage-based approaches. In two recent publications WGS was used to identify the causative gene, but it should be noted that in both instances the investigators identified the genetic alterations in the exome.[50 51] Thus, in total since 2009, more than 20 causative genes have been identified and this number is growing exponentially. With the availability of this technology, several initiatives in North America through the National Institute of Health (NIH), Finding of Rare Disease Genes in Canada (FORGE) and Rare Disease Consortium for Autosomal Loci (RaDiCAL) have been launched which aim to characterise Mendelian disorders. The challenge is now how to validate, among the multiple variants that will be identified, the causative alteration and link it to disease and function. Moreover, it is likely that we will identify novel phenotypes for a known gene, or the reverse, as different hypomorphic mutations can lead to distinct phenotypes[52–54] and identical mutations in a given gene have been shown to induce distinct phenotypes.[55] In addition, in highly consanguineous probands, the compounded effect of added genetic alterations can affect the phenotype and lead to what has been identified as a novel disease.

Paradigm shift brought by WES in the identification of de novo mutations

A remarkable demonstration of how powerful WES using NGS can be in teaching us about human disease comes from a recent study that provides evidence that de novo single nucleotide variants may contribute substantially to mental retardation.[56] Investigators explored a major paradox in evolutionary theory—namely, that the per-generation mutation rate in humans is high despite the allelic loss due to reduced fertility. They postulated that these de novo mutations may compensate for allele loss in common neurodevelopmental and psychiatric diseases, and explain this paradox in evolutionary genetic theory. They used a family-based WES approach to test this de novo mutation hypothesis in 10 individuals with unexplained mental retardation. Following WES of their trios (parents and proband), they identified and validated unique non-synonymous de novo mutations in nine genes. While they further establish the power of NGS/WES in identifying the genetic basis of human diseases, their findings also provide strong experimental support for a de novo paradigm; de novo point mutations of large effect sizes together with de novo copy number variation could potentially explain the majority of all cases of mental retardation in the population.[56] In addition, in cases where the results from the analysis of WES are not conclusive in identifying a causative gene and/or in the scenario where only one affected individual within the family is available for studying a rare disorder, resequencing trios (parents and the patient) can help to pinpoint the causative genetic variant by excluding mutations shared with the parents.

WES in characterising complex trait disorders

Genome-wide association (GWA) studies of complex traits have been successful in identifying common variant associations but have failed to explain most of the heritability of these traits.[57] The field of complex trait genetics is shifting towards the study

of low-frequency (minor allele frequency (MAF) 0.01–0.05) and rare (MAF <0.01) variants, some of which are hypothesised to have larger effects. Indeed, GWA studies, which so far have focused on very common SNPs, have been completed for most common human diseases and many related traits.[58] [59] These studies have been designed based on the knowledge of most of the very common gene variants (MAF >~5%) in the human genome and have identified over 500 independent strong SNP associations ($p<1\times10^{-8}$) (see the National Human Genome Research Institute Catalog of Published GWA Studies). However, most of these associated SNPs have very small effect sizes and the proportion of heritability explained is at best modest for most traits.[57] [60] Furthermore, most GWA signals have yet to be tracked to causal polymorphisms. Fine mapping and functional evaluation of these loci is an ongoing process with several successful examples indicating that the causal variants must have subtle regulatory effects. Although the systematic identification of rare variants associated with common diseases has not yet been feasible, several rare variants have nevertheless been identified that confer a substantial risk of disease. For example, autism, mental retardation, epilepsy and schizophrenia have been shown to be influenced by rare structural variants that affect genes.[61] Additionally, it seems possible that some—perhaps even many—of the current GWA signals could reflect the effect of one or more rare variants that have been tagged by common variants.[62] Whatever underlies the GWA signals for common diseases, it is clear that GWA studies of common variants have limited value in disease prediction.

As discussed above, past results make a strong case for common diseases being more similar to Mendelian diseases than is postulated by the common disease—common variant model. It seems possible that much of the genetic cause of common diseases is due to rare and generally deleterious variants that have a strong impact on the risk of disease in individual patients, and which can now be identified thanks to NGS. Because the most obvious disease-influencing variants will be the clearly functional ones, WES has the potential to identify these rare variants and allow definitive connections to be rapidly established between specific genes and many important common diseases. However, there are drawbacks to this technique, the most important being that it almost entirely misses structural variation. WES, despite its name, also misses a certain set of exons: if causal variants lie within these exons that are not targeted (we found that 5—10% of RefSeq exons or ~3% of RefSeq coding exons have <5× coverage in the latest commercial capture kits), they will not be identified, as is also the case for monogenic disorders. Additionally, the capture methods currently used require the sequencing of a far greater number of bases than expected based on the size of the exome, which makes WES prices comparable to those of low-coverage WGS. However, low-coverage sequencing will miss many of the variants present in only a single individual. For these reasons, high-coverage WES is the method of choice for complex traits as it is becoming more affordable, and there is a rapid increase in the sequencing capacity of existing platforms as well as the development of new less-expensive platforms.

An essential point to improve the chances of success is to carefully choose cases and controls for each study as costs restrict sample size. Selecting cases with a strong family history will increase the probability of finding pathogenic variants with large effects. The availability of large control cohorts who can be recalled and phenotypically evaluated will also be crucial. Unlike variants with weak influences which could be expected to appear in the population at large without much phenotypic effect, a variant of strong effect is less likely to appear in an individual without any phenotypic consequence. Confirmation of such potentially causal variants will therefore often require the careful evaluation of the phenotype in any controls who are carriers. Several complex diseases are currently being investigated using NGS and include mental diseases, diabetes and autoimmune disorders such as lupus and inflammatory bowel diseases. The results from these studies have the potential to revolutionise the screening of these disorders and our therapeutic approach. The low-frequency and rare variants are likely to be population-specific at a much finer scale than common variants, so careful geographical and molecular matching of cases and controls is that much more important than in GWA studies. An alternative strategy is to do gene-based rather than variant-based analysis, which will require sequencing rather than genotyping in the replication cohorts. The pros and cons of these alternatives are still being debated.

WES in characterising cancer

The application of NGS technologies is allowing substantial advances in cancer genomics. Indeed, the development of massively parallel sequencing technologies makes it feasible to catalogue all classes of somatically-acquired mutations in a cancer.[63–65] It has become feasible to sequence all expressed genes (the transcriptomes),[66] [67] exomes and, more recently, complete genomes[64] [68–70] of cancer samples. WES with capillary sequencing allowed the analysis of all known coding genes in colorectal, breast and pancreatic carcinomas and glioblastoma.[71–73] These studies have led to the discovery of somatic mutations in isocitrate dehydrogenase 1 in glioblastoma[72] and of germline mutations in *PALB2* (the gene encoding partner and localiser of BRCA2) in patients with pancreatic carcinoma,[74] among other important findings. In addition, the hybrid selection approach will be particularly powerful for diagnostic analysis of the cancer genome; for diagnosis, there may be value in sequencing specific oncogenes and/or tumour suppressor genes at very high coverage in samples with a low percentage of tumour cells.[75] However, a major challenge of cancer genome analysis is to identify 'driver' mutations,[65] and several recent genome studies of leukaemias, myelomas and solid tumours including breast, lung and pancreatic cancer have concentrated their analysis on coding regions (exomes) to increase the likelihood of identifying driver mutations[76–79] or used integrative genomics approaches (mapping of structural variation, whole genome methylation and gene expression analysis) in association with NGS techniques.[72] Whole exome analysis on these distinct subtypes provides a better understanding of mechanisms underlying specific cancers and also identifies new biomarkers and/or drug targets, as recently reported, for example, in individuals with acute myeloid leukaemia.[68] [80] [81]

WES is thus opening new avenues towards understanding the molecular pathogenesis of cancers. For example, the discovery of DNMT3A (a gene involved in DNA methylation) mutations in acute myeloid leukaemia (AML) may imply that aberrant epigenetic regulation is critical for pathogenesis, but the exact link—whether it be altered gene expression or genome instability—has yet to be uncovered. Other key questions include what aspects of leukaemia biology can be attributed to mutations in this gene and why it is concentrated in specific AML subtypes and associated with a poor prognosis. Thus, additional genetic and/or epigenetic events can modulate this disease type and must be uncovered through further exploration of the genome and the epigenome. WGS is also being used to

investigate cancers, with an initial focus on very specific subgroups in certain cancers to minimise the confounding effect of genetic heterogeneity.[70]

Other innovative approaches are making use of NGS in targeted exome/genome sequencing for the design of cost-effective targeted sequencing methods to the benefit of personalised chemotherapy.[82] In another recently published study, targeted NGS detected point mutations, insertions, deletions and balanced chromosomal rearrangements and identified novel leukaemia-specific fusion genes in a single procedure combining 454 shotgun pyrosequencing with long oligonucleotide sequence capture arrays.[83] In yet another study, NGS was applied as a screening method to characterise a number of known genetic alterations in chronic myelomonocytic leukaemia and identified that a pattern of molecular mutations translated into distinct biological and prognostic categories.[84]

There are unique methodological considerations in NGS analyses of cancer samples (reviewed by Meyerson et al[85]). Cancer samples and cancer genomes have general characteristics that are distinct from other tissue samples and from genomic sequences that are inherited through the germ line. Cancers themselves may be highly heterogeneous and composed of different clones that have different genomes.[86] Cancer genomes are enormously diverse and complex and have major structural variability. They vary substantially in their sequence and structure compared with normal genomes and among themselves. To identify somatic alterations in cancer, comparison with matched normal DNA from the same individual is essential. This is largely because of our incomplete knowledge of the variations in the normal human genome; to date, each 'matched normal' cancer genome sequence has identified large numbers of mutations and rearrangements in the germ line that had not previously been described.

RELEVANCE FOR THE CLINICAL USE OF WES

Gene discovery is an essential starting point for both understanding the genetic mechanisms underlying diseases and for providing clues to therapeutic approaches. Gene-specific treatments are currently ongoing worldwide, and several successful gene therapy trials aim to correct inborn errors for diseases such as immune deficiencies, metabolic disorders and, more recently, thalassaemia.[87–93] Local delivery of the replacement gene is also being tested in human clinical trials for several forms of hereditary blindness such as Leber congenital amaurosis and retinitis pigmentosa.[94] Also, genetic testing for common mutations in recessive disorders such as Tay-Sachs disease has proved to be of benefit both for diagnosis and carrier detection.[5] For complex traits, understanding the genetic alterations in disease variability and resistance to treatment in a given individual could revolutionise care and may soon make the concept of personalised medicine a reality. WES is paving the way to identifying driver mutations in cancer as well as the genetic events leading to metastasis, the primary cause of cancer mortality, and which are potentially amenable to therapeutic targeting. These insights will provide improved means to prevent recurrence and to avoid therapeutic resistance. WES can also complement histopathological analysis by allowing for more accurate diagnosis and improved subgrouping of patients.[95] The clinical applications are thus enormous. With regard to personal genomics, numerous companies already use SNP arrays to offer predictions of common disease risk directly to consumers, which can influence lifestyle choices and decisions to use relatively non-invasive monitoring programmes (eg, imaging). Genome sequencing will greatly improve the specificity of such predictions and adds the ability to detect novel variants, and might lead to an expansion of fetal screening.
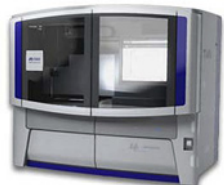
WHY NOT USE WGS?

WGS is being increasingly used based on its availability and improved cost efficiency. Indeed, in the future, WGS is predicted to be more economical than WES because the capture process is skipped entirely. This technique has the advantage of capturing all of the exome (as some can be missed by the exome capture process), and can provide information on variants in highly evolutionary conserved non-coding regions and other variants throughout the genome. In addition, WGS using a paired-end approach can be used to detect large structural variants such as large insertions or deletions, inversion and translocations. As the cost of WGS continues to decrease, it will become increasingly popular because of its ability to survey most of the genome as well as additional classes of mutations. However, the amount of data generated from WGS is 100 times more than the already overwhelming amount obtained by WES. The bioinformatics filtering techniques, storage facilities, software and hardware for data analysis will prove a challenge and most ongoing projects initially focus on the exome for the first analysis. It should be noted that WGS is not immune to some of the drawbacks of exome sequencing. There is significant variability in sequencing efficiency across the genome and the fluctuations in coverage will result in many regions of interest being missed. Also, repetitive regions—exonic and others—are difficult to align in either case and can result in missed variants or an excess of variant calls. These problems can be resolved in the future with more uniform library construction, higher sequencing depth and longer reads and paired end fragment sizes. However, in the foreseeable future we are likely to continue facing some gaps in genome coverage. Unless analyses will specifically focus on non-coding regions or on structural variation, WES provides most of the benefits of WGS but with lower costs, both for sequencing and for storage and analysis of the data.

SOME PRACTICAL CONSIDERATIONS

Large throughput genomic data analysis has traditionally been the domain of the bioinformatician and statistician. Laboratory researchers have gradually been embracing genomic technologies such as microarrays and applying them as 'hypothesis-free' discovery tools to be followed up by focused experimentation. This type of experimentation was usually accompanied by collaboration with skilled data analysts or the development of custom analysis software. NGS data will undoubtedly become a household item in the near future. What can a researcher or a clinician embarking on a WES or WGS adventure expect to obtain from the sequencing service provider?

A number of excellent commercially-available targeted sequence capture kits are available including kits from Agilent (Santa Clara, CA, USA), Illumina (San Diego, CA, USA) and Nimblegen (Madison, WI, USA). Once captured, the DNA fragments need to be sequenced. Currently, ABI (Carlsbad, CA, USA) and Illumina are the two major companies in the sequencing field, but a number of third-generation sequencers are being developed and may enter the field in the near future (table 2). After the sequencing process, individual sequence reads are typically aligned to the reference genome sequence. Next, variant positions are identified between the sample of interest and the reference genome. At this point a number of bioinformatic filtering steps are required to separate common benign polymorphisms from potentially deleterious mutations.

Table 2 Comparison of next-generation sequencing instruments

	Instrument	Amplification method	Sequencing method	Current max output
	Roche GS-FLX	Emulsion PCR	Pyrosequencing	1.5M reads/run (1 day) 500 bases/read
	Life Technologies Ion Torrent	Emulsion PCR	PostLight sequencing	100k reads/run (2 h) 100 bases/read
	Life Technologies SOLiD 5500XL	Emulsion PCR	Sequencing by ligation	4G reads/run (7 days) 75 bases/run
	Illumina HiSeq 2000	Bridge amplification	Sequencing by synthesis	6G reads/run (8 days) 100 bases/read
	Pacific Biosciences	Single molecule	Real-time sequencing	100k reads/run (2 h) >1000 bases/run

We believe that, while the choice of the optimal technology and analytical pipelines is important, it is secondary to the service provider's experience with the specific technology and willingness to engage in some back-and-forth dialogue with the researcher on custom analysis needs. As an example, at the McGill University and Genome Quebec Innovation Centre, we have to date processed the data from over 300 exomes. This experience is instrumental in identifying systematic false positive and false negative results. False positives most often arise from incorrect mapping and systematic sequencing errors—for example, certain words (combination of nucleotides) being systematically misread by the sequencer. Both of these errors can be removed by comparing each test sample against previously sequenced exomes. Systematic errors occur over and over again but, if they are present in a certain proportion of all sequenced samples, they can be easily removed from the final list of variants. False negative results can result from low overall coverage, poor capture efficiency of certain regions and difficulty in unambiguously aligning repetitive regions. Such missing regions can easily be flagged and reported to the researcher who may want to follow them up by targeted sequencing.

The final output for each sample is a list of variants that can be easily manipulated in a spreadsheet. In our experience, each sample produces roughly 500 potentially 'interesting' protein-coding variants—that is, those that have not been seen in more than 5% of other exomes. Our annotated output file contains basic information on the chromosomal position, nucleotide change, predicted protein change, gene name and gene description. Further annotation includes Online Mendelian Inheritance in Man (OMIM) entry (if available), Scale-invariant feature transform (SIFT)[96] prediction of how likely the change is to be damaging, interspecies conservation of each residue, dbSNP entry and allele frequency from the 1000 Genomes Project.[97] We also provide information on the sequencing quality of each variant and clickable links to the primary sequencing data visualised in the Integrative Genomics Viewer[98] and the graphic display of each position in the UCSC Genome Browser.[99] The end user who is interested in a recessive disease and studies a consanguineous family, for example, can then apply simple spreadsheet filtering functions to display only homozygous changes that have never been seen before or that have very low minor allele frequencies. In most cases this will limit the final list

to a manageable number of a dozen or fewer candidate variants that can then be followed up manually.

Of course, as the number of samples and complexity of a disorder increases, at some point a switch from a spreadsheet to a dedicated bioinformatician may be necessary. However, for Mendelian disorders a spreadsheet-savvy researcher should be quite successful in analysing exome sequencing results.

Some final rules of thumb: choose a friendly experienced sequencing centre; longer reads are better than shorter reads as they reduce false positives from mapping ambiguity; paired-end reads are better than single-end for the same reason; and 30× median coverage of the target may be sufficient, but 100× coverage is much safer as it ensures that variants can confidently be determined across a higher proportion of the exome.

ETHICAL ISSUES RAISED BY WES

The increased ability to share large amounts of individual-specific genetic information across borders puts a new twist on perennial ethical issues such as consent, feedback, protection of privacy and the governance of research.[100–103] Informed consent is needed from participants in research and has been a guiding principle of medical investigation since the mid-20th century. This conception of consent, along with the concomitant power to withdraw from research without prejudice, arose originally in the context of biomedical research. It had the aim of protecting participants from abuse and from potential physical harm, and focused on clinical interventions and the collection of samples rather than on data collection per se. From its inception, informed consent was strongly concerned with the protection of individuals.

Genomics research, however, moves away from these origins on several counts. The information that is derived from DNA is a powerful personal identifier and can provide information—not just on the individual but also on the individual's relatives and ethnic groups—in a format that is easy to share across international borders. Although samples and data have personal identifiers removed, individuals may still be re-identifiable because of the richness of the data derived from the analysis. The data produced are often shared informally among researchers, but more formal mechanisms have been put in place by funders to ensure the rapid sharing of NGS data, such as the requirements to deposit data sets in open access archives.[104] Examples are the European Genotype Archive and dbGaP (NIH-USA). The complexity of genomics research, together with the difficulty of providing precise specifications for future use of data, have prompted serious concerns about whether any consent to such research can be adequately 'informed'. There is a pressing need to learn from insights gained elsewhere, such as in genetic counselling and in family studies. Likewise, calls to involve the community in consent pose ethical issues about individual and group rights which may be different for communities across the globe.

Reporting findings back to the participants may be considered to be an important part of building and maintaining public trust in research.[102 105–107] Providing participants with information about the general findings of research, such as publications based on the research, is an uncontroversial and welcome practice. In contrast, informing a single individual of his or her results remains controversial in many areas of research and particularly in the area of WGS. There appears to be some agreement that, where there is a serious treatable condition, researchers have a moral obligation to feed this information back to research participants.[108] In cases where findings are of a less serious nature, are untreatable or of uncertain significance, the potential benefits for participants of being informed need to be balanced against the participant's right not to know. The thoughtful handling of such issues is of clear relevance to the maintenance of public trust in the research process and is the subject of ongoing studies in collaboration with some of the initiatives exploring WES/WGS in rare diseases and cancer in North America. Not surprisingly, an ethical investigation component has been added to each of these initiatives to try and assess the impact of the findings on families and to determine ways for appropriate return of information and protection of privacy.

CONCLUSIONS AND FUTURE DIRECTIONS

Vast amounts of clinical, biological and sequencing data are now being generated by an expanding number of research efforts on a scale that was only imagined just a few years ago. Interpreting these data and translating the findings to improve healthcare is a challenge in itself. In addition to developing locus-specific databases and large data warehouses for NGS datasets, there is a major need to create dedicated databases to enhance the clinical interpretation process. Indeed, the development of analysis techniques to cope with the millions of variants called per genome will be a high priority, as will the development of techniques that can combine data about different rare variants into one analysis. The advances we have described highlight the important implications that particular mutations, discovered through NGS and WES, can have for medical management and for tailoring therapy to the genetic background of a given individual in a vast array of diseases. Furthermore, definitive connections—for example, a clearly functional mutation in a single gene conferring a strongly increased risk of a disease—would provide validated therapeutic targets for the pharmaceutical industry and genetic discovery could be the most likely avenue for ameliorating the ongoing crisis in global drug development. This prediction assumes that rare variants will be found that have large influences on rare and common diseases, that their biological functions will be obvious and that locus and allelic heterogeneity will not prevent insights into the mechanisms of disease. How often these assumptions will hold is currently unknown and will largely determine the rate of discovery in the coming years.

REFERENCES
1. **Watson JD,** Crick FH. The structure of DNA. *Cold Spring Harb Symp Quant Biol* 1953;**18**:123–31.
2. **Watson JD,** Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953;**171**:737–8.
3. **Inoue T,** Orgel LE. A nonenzymatic RNA polymerase model. *Science* 1983;**219**:859–62.
4. **Maxam AM,** Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 1977;**74**:560–4.
5. **Sanger F,** Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;**74**:5463–7.
6. **Sachidanandam R,** Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L,

Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001;**409**:928—33.

7. **Lander ES,** Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860—921.

8. **McPherson JD,** Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen HC, Jang W, Green ED, Idol JR, Maduro VV, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, de Jong PJ, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, Trask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng JF, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen XN, Fujiyama A, Hattori M, Toyoda A, Yada T, Park HS, Sakaki Y, Shimizu N, Asakawa S, Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H. A physical map of the human genome. *Nature* 2001;**409**:934—41.

9. **Ng SB,** Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. *Hum Mol Genet* 2010;**19**:R119—24.

10. **Ng SB,** Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;**461**:272—6.

11. **Shendure J,** Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008;**26**:1135—45.

12. **Turner EH,** Lee C, Ng SB, Nickerson DA, Shendure J. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods* 2009;**6**:315—16.

13. **Albert TJ,** Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA. Direct selection of human genomic loci by microarray hybridization. *Nat Methods* 2007;**4**:903—5.

14. **Gnirke A,** Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009;**27**:182—9.

15. **Hodges E,** Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 2007;**39**:1522—7.

16. **Levin JZ,** Berger MF, Adiconis X, Rogov P, Melnikov A, Fennell T, Nusbaum C, Garraway LA, Gnirke A. Targeted next-generation sequencing of a cancer

17. **Bainbridge MN,** Wang M, Burgess DL, Kovar C, Rodesch MJ, D'Ascenzo M, Kitzman J, Wu YQ, Newsham I, Richmond TA, Jeddeloh JA, Muzny D, Albert TJ, Gibbs RA. Whole exome capture in solution with 3 Gbp of data. *Genome Biol* 2010;**11**:R62.

18. **Voelkerding KV,** Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 2009;**55**:641—58.

19. **Koboldt DC,** Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. *Brief Bioinform* 2010;**11**:484—98.

20. **Samuels ME.** Saturation of the human phenome. *Curr Genomics* 2010;**11**:482—99.

21. **Davies MA,** Samuels Y. Analysis of the genome to personalize therapy for melanoma. *Oncogene* 2010;**29**:5545—55.

22. **Laurier V,** Stoetzel C, Muller J, Thibault C, Corbani S, Jalkh N, Salem N, Chouery E, Poch O, Licaire S, Danse JM, Amati-Bonneau P, Bonneau D, Mégarbané A, Mandel JL, Dollfus H. Pitfalls of homozygosity mapping: an extended consanguineous Bardet-Biedl syndrome family with two mutant genes (BBS2, BBS10), three mutations, but no triallelism. *Eur J Hum Genet* 2006;**14**:1195—203.

23. **Nishimura DY,** Swiderski RE, Searby CC, Berg EM, Ferguson AL, Hennekam R, Merin S, Weleber RG, Biesecker LG, Stone EM, Sheffield VC. Comparative genomics and gene expression analysis identifies BBS9, a new Bardet-Biedl syndrome gene. *Am J Hum Genet* 2005;**77**:1021—33.

24. **Paisán-Ruiz C,** Scopes G, Lee P, Houlden H. Homozygosity mapping through whole genome analysis identifies a COL18A1 mutation in an Indian family presenting with an autosomal recessive neurological disorder. *Am J Med Genet B Neuropsychiatr Genet* 2009;**150B**:993—7.

25. **Strauss KA,** Puffenberger EG, Craig DW, Panganiban CB, Lee AM, Hu-Lince D, Stephan DA, Morton DH. Genome-wide SNP arrays as a diagnostic tool: clinical description, genetic mapping, and molecular characterization of Salla disease in an Old Order Mennonite population. *Am J Med Genet A* 2005;**138A**:262—7.

26. **Kuhlenbäumer G,** Hullmann J, Appenzeller S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum Mutat* 2011;**32**:144—51.

27. **Chen JM,** Férec C, Cooper DN. Revealing the human mutome. *Clin Genet* 2010;**78**:310—20.

28. **Patel K,** Larson C, Hargreaves M, Schlundt D, Wang H, Jones C, Beard K. Community screening outcomes for diabetes, hypertension, and cholesterol: Nashville REACH 2010 project. *J Ambul Care Manage* 2010;**33**:155—62.

29. **Hoischen A,** van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BB, Veltman JA. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 2010;**42**:483—5.

30. **Ng SB,** Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;**42**:790—3.

31. **Pierce SB,** Walsh T, Chisholm KM, Lee MK, Thornton AM, Fiumara A, Opitz JM, Levy-Lahad E, Klevit RE, King MC. Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault syndrome. *Am J Hum Genet* 2010;**87**:282—8.

32. **Choi M,** Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 2009;**106**:19096—101.

33. **Lalonde E,** Albrecht S, Ha KC, Jacob K, Bolduc N, Polychronakos C, Dechelotte P, Majewski J, Jabado N. Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. *Hum Mutat* 2010;**31**:918—23.

34. **Gilissen C,** Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M, van Reeuwijk J, Kant SG, Roepman R, Knoers NV, Veltman JA, Brunner HG. Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am J Hum Genet* 2010;**87**:418—23.

35. **Ng SB,** Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 2010;**42**:30—5.

36. **Johnson JO,** Gibbs JR, Van Maldergem L, Houlden H, Singleton AB. Exome sequencing in Brown-Vialetto-van Laere syndrome. *Am J Hum Genet* 2010;**87**:567—9; author reply 569—70.

37. **Musunuru K,** Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ, Fennell T, Banks E, Ambrogio L, Cibulskis K, Kernytsky A, Gonzalez E, Rudzicz N, Engert JC, DePristo MA, Daly MJ, Cohen JC, Hobbs HH, Altshuler D, Schonfeld G, Gabriel SB, Yue P, Kathiresan S. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med* 2010;**363**:2220—7.

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

mutations in hyperphosphatasia mental retardation syndrome. *Nat Genet* 2010;**42**:827—9.

39. **Anastasio N,** Ben-Omran T, Teebi A, Ha KC, Lalonde E, Ali R, Almureikhi M, Der Kaloustian VM, Liu J, Rosenblatt DS, Majewski J, Jerome-Majewska LA. Mutations in SCARF2 are responsible for Van Den Ende-Gupta syndrome. *Am J Hum Genet* 2010;**87**:553—9.

40. **Erlich Y,** Edvardson S, Hodges E, Zenvirt S, Thekkat P, Shaag A, Dor T, Hannon GJ, Elpeleg O. Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. *Genome Res* 2011;**21**:658—64.

41. **Glazov EA,** Zankl A, Donskoi M, Kenna TJ, Thomas GP, Clark GR, Duncan EL, Brown MA. Whole-exome re-sequencing in a family quartet identifies pop1 mutations as the cause of a novel skeletal dysplasia. *PLoS Genet* 2011;**7**:e1002027.

42. **Tsurusaki Y,** Osaka H, Hamanoue H, Shimbo H, Tsuji H, Doi H, Saitsu H, Matsumoto N, Miyake N. Rapid detection of a mutation causing X-linked leucoencephalopathy by exome sequencing. *J Med Genet* 2011;**48**:606—9.

43. **Bolze A,** Byun M, McDonald D, Morgan NV, Abhyankar A, Premkumar L, Puel A, Bacon CM, Rieux-Laucat F, Pang K, Britland A, Abel L, Cant A, Maher ER, Riedl SJ, Hambleton S, Casanova JL. Whole-exome-sequencing-based discovery of human FADD deficiency. *Am J Hum Genet* 2010;**87**:873—81.

44. **Caliskan M,** Chong JX, Uricchio L, Anderson R, Chen P, Sougnez C, Garimella K, Gabriel SB, dePristo MA, Shakir K, Matern D, Das S, Waggoner D, Nicolae DL, Ober C. Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13. *Hum Mol Genet* 2011;**20**:1285—9.

45. **Walsh T,** Shahin H, Elkan-Miller T, Lee MK, Thornton AM, Roeb W, Abu Rayyan A, Loulus S, Avraham KB, King MC, Kanaan M. Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *Am J Hum Genet* 2010;**87**:90—4.

46. **Johnson JO,** Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, Gibbs JR, Brunetti M, Gronka S, Wuu J, Ding J, McCluskey L, Martinez-Lage M, Falcone D, Hernandez DG, Arepalli S, Chong S, Schymick JC, Rothstein J, Landi F, Wang YD, Calvo A, Mora G, Sabatelli M, Monsurrò MR, Battistini S, Salvi F, Spataro R, Sola P, Borghero G, Galassi G, Scholz SW, Taylor JP, Restagno G, Chiò A, Traynor BJ; ITALSGEN Consortium. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* 2010;**68**:857—64.

47. **Wang JL,** Yang X, Xia K, Hu ZM, Weng L, Jin X, Jiang H, Zhang P, Shen L, Guo JF, Li N, Li YR, Lei LF, Zhou J, Du J, Zhou YF, Pan Q, Wang J, Wang J, Li RQ, Tang BS. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* 2010;**133**:3510—18.

48. **Isidor B,** Lindenbaum P, Pichon O, Bézieau S, Dina C, Jacquemont S, Martin-Coignard D, Thauvin-Robinet C, Le Merrer M, Mandel JL, David A, Faivre L, Cormier-Daire V, Redon R, Le Caignec C. Truncating mutations in the last exon of NOTCH2 cause a rare skeletal disorder with osteoporosis. *Nat Genet* 2011;**43**:306—8.

49. **Simpson MA,** Irving MD, Asilmaz E, Gray MJ, Dafou D, Elmslie FV, Mansour S, Holder SE, Brain CE, Burton BK, Kim KH, Pauli RM, Aftimos S, Stewart H, Kim CA, Holder-Espinasse M, Robertson SP, Drake WM, Trembath RC. Mutations in NOTCH2 cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss. *Nat Genet* 2011;**43**:303—5.

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

51. **Lupski JR,** Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 2010;**362**:1181—91.

52. **Notarangelo LD,** Giliani S, Mazza C, Mella P, Savoldi G, Rodriguez-Pérez C, Mazzolari E, Fiorini M, Duse M, Plebani A, Ugazio AG, Vihinen M, Candotti F, Schumacher RF. Of genes and phenotypes: the immunological and molecular spectrum of combined immune deficiency. Defects of the gamma(c)-JAK3 signaling pathway as a model. *Immunol Rev* 2000;**178**:39—48.

53. **de Villartay JP,** Lim A, Al-Mousa H, Dupont S, Déchanet-Merville J, Coumau-Gatbois E, Gougeon ML, Lemainque A, Eidenschenk C, Jouanguy E, Abel L, Casanova JL, Fischer A, Le Deist F. A novel immunodeficiency associated with hypomorphic RAG1 mutations and CMV infection. *J Clin Invest* 2005;**115**:3291—9.

54. **McCusker C,** Hotte S, Le Deist F, Hirschfeld AF, Mitchell D, Nguyen VH, Gagnon R, Mazer B, Turvey SE, Jabado N. Relative CD4 lymphopenia and a skewed memory phenotype are the main immunologic abnormalities in a child with Omenn syndrome due to homozygous RAG1-C2633T hypomorphic mutation. *Clin Immunol* 2009;**131**:447—55.

55. **Corneo B,** Moshous D, Gungor T, Wulffraat N, Philippet P, Le Deist FL, Fischer A, de Villartay JP. Identical mutations in RAG1 or RAG2 genes leading to defective V(D) J recombinase activity can cause either T-B-severe combined immune deficiency or Omenn syndrome. *Blood* 2001;**97**:2772—6.

56. **Vissers LE,** de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA. A de novo paradigm for mental retardation. *Nat Genet* 2010;**42**:1109—12.

57. **Cirulli ET,** Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010;**11**:415—25.

58. **Wang K,** Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010;**11**:843—54.

59. **Hakonarson H,** Qu HQ, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Eckert AW, Annaiah K, Lawson ML, Otieno FG, Santa E, Shaner JL, Smith RM, Onyiah CC, Skraban R, Chiavacci RM, Robinson LJ, Stanley CA, Kirsch SE, Devoto M, Monos DS, Grant SF, Polychronakos C. A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. *Diabetes* 2008;**57**:1143—6.

60. **Maher B.** Personal genomes: the case of the missing heritability. *Nature* 2008;**456**:18—21.

61. **Stankiewicz P,** Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med* 2010;**61**:437—55.

62. **Dickson SP,** Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010;**8**:e1000294.

63. **Ley TJ,** Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF, Wilson RK. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008;**456**:66—72.

64. **Pleasance ED,** Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordonez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, Beare D, Butler A, Carter RJ, Chen L, Cox AJ, Edkins S, Kokko-Gonzales PI, Gormley NA, Grocock RJ, Haudenschild CD, Hims MM, James T, Jia M, Kingsbury Z, Leroy C, Marshall J, Menzies A, Mudie LJ, Ning Z, Royce T, Schulz-Trieglaff OB, Spiridou A, Stebbings LA, Szajkowski L, Teague J, Williamson D, Chin L, Ross MT, Campbell PJ, Bentley DR, Futreal PA, Stratton MR. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 2010;**463**:191—6.

65. **Stratton MR,** Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009;**458**:719—24.

66. **Maher CA,** Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM. Transcriptome sequencing to detect gene fusions in cancer. *Nature* 2009;**458**:97—101.

67. **Maher CA,** Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtukova I, Barrette TR, Grasso C, Yu J, Lonigro RJ, Schroth G, Kumar-Sinha C, Chinnaiyan AM. Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc Natl Acad Sci U S A* 2009;**106**:12353—8.

68. **Mardis ER,** Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott RM, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, McMichael JF, Meyer RJ, Schindler JK, Pohl CS, Wallis JW, Shi X, Lin L, Schmidt H, Tang Y, Haipek C, Wiechert ME, Ivy JV, Kalicki J, Elliott G, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson MA, Baty J, Heath S, Shannon WD, Nagarajan R, Link DC, Walter MJ, Graubert TA, DiPersio JF, Wilson RK, Ley TJ. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 2009;**361**:1058—66.

69. **Pleasance ED,** Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordonez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA, Campbell PJ. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 2010;**463**:184—90.

70. **Lee W,** Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, Yue P, Zhang Y, Pant KP, Bhatt D, Ha C, Johnson S, Kennemer MI, Mohan S, Nazarenko I, Watanabe C, Sparks AB, Shames DS, Gentleman R, de Sauvage FJ, Stern H, Pandita A, Ballinger DG, Drmanac R, Modrusan Z, Seshagiri S, Zhang Z. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 2010;**465**:473—7.

71. **Jones S,** Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008;**321**:1801—6.

72. **Parsons DW,** Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008;**321**:1807—12.

73. **Sjöblom T,** Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006;**314**:268—74.

74. **Jones S,** Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, Lin JC, Palmisano E, Brune K, Jaffee EM, Iacobuzio-Donahue CA, Maitra A, Parmigiani G, Kern SE, Velculescu VE, Kinzler KW, Vogelstein B, Eshleman JR, Goggins M, Klein AP. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* 2009;**324**:217.

75. **Thomas RK,** Nickerson E, Simons JF, Jänne PA, Tengs T, Yuza Y, Garraway LA, LaFramboise T, Lee JC, Shah K, O'Neill K, Sasaki H, Lindeman N, Wong KK, Borras AM, Gutmann EJ, Dragnev KH, DeBiasi R, Chen TH, Glatt KA, Greulich H, Desany B, Lubeski CK, Brockman W, Alvarez P, Hutchison SK, Leamon JH, Ronan MT, Turenchalk GS, Egholm M, Sellers WR, Rothberg JM, Meyerson M. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med* 2006;**12**:852–5.

76. **Bignell GR,** Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM, Buck G, Chen L, Beare D, Latimer C, Widaa S, Hinton J, Fahey C, Fu B, Swamy S, Dalgliesh GL, Teh BT, Deloukas P, Yang F, Campbell PJ, Futreal PA, Stratton MR. Signatures of mutation and selection in the cancer genome. *Nature* 2010;**463**:893–8.

77. **Dalgliesh GL,** Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, Teague J, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Forbes S, Jia M, Jones D, Knott H, Kok CY, Lau KW, Leroy C, Lin ML, McBride DJ, Maddison M, Maguire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, O'Meara S, Pleasance E, Rajasingham A, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turrell K, Dykema KJ, Khoo SK, Petillo D, Wondergem B, Anema J, Kahnoski RJ, Teh BT, Stratton MR, Futreal PA. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* 2010;**463**:360–3.

78. **Morin RD,** Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, Kuchenbauer F, Yap D, Humphries RK, Griffith OL, Shah S, Zhu H, Kimbara M, Shashkin P, Charlot JF, Tcherpakov M, Corbett R, Tam A, Varhol R, Smailus D, Moksa M, Zhao Y, Delaney A, Qian H, Birol I, Schein J, Moore R, Holt R, Horsman DE, Connors JM, Jones S, Aparicio S, Hirst M, Gascoyne RD, Marra MA. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* 2010;**42**:181–5.

79. **Purow B,** Schiff D. Advances in the genetics of glioblastoma: are we reaching critical mass? *Nat Rev Neurol* 2009;**5**:419–26.

80. **Yan XJ,** Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* 2011;**43**:309–15.

81. **Ley TJ,** Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* 2010;**363**:2424–33.

82. **Wesolowska A,** Dalgaard MD, Borst L, Gautier L, Bak M, Weinhold N, Nielsen BF, Helt LR, Audouze K, Nersting J, Tommerup N, Brunak S, Sicheritz-Ponten T, Leffers H, Schmiegelow K, Gupta R. Cost-effective multiplexing before capture allows screening of 25 000 clinically relevant SNPs in childhood acute lymphoblastic leukemia. *Leukemia* 2011;**25**:1001–6.

83. **Grossmann V,** Kohlmann A, Klein HU, Schindela S, Schnittger S, Dicker F, Dugas M, Kern W, Haferlach T, Haferlach C. Targeted next-generation sequencing detects point mutations, insertions, deletions and balanced chromosomal rearrangements as well as identifies novel leukemia-specific fusion genes in a single procedure. *Leukemia* 2011;**25**:671–80.

84. **Kohlmann A,** Grossmann V, Klein HU, Schindela S, Weiss T, Kazak B, Dicker F, Schnittger S, Dugas M, Kern W, Haferlach C, Haferlach T. Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, and RUNX1. *J Clin Oncol* 2010;**28**:3858–65.

85. **Meyerson M,** Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 2010;**11**:685–96.

86. **Navin N,** Krasnitz A, Rodgers L, Cook K, Meth J, Kendall J, Riggs M, Eberling Y, Troge J, Grubor V, Levy D, Lundin P, Månér S, Zetterberg A, Hicks J, Wigler M. Inferring tumor progression from genomic heterogeneity. *Genome Res* 2010;**20**:68–80.

87. **Cavazzana-Calvo M,** Hacein-Bey S, de Saint Basile G, Gross F, Yvon E, Nusbaum P, Selz F, Hue C, Certain S, Casanova JL, Bousso P, Deist FL, Fischer A. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 2000;**288**:669–72.

88. **Hacein-Bey-Abina S,** Le Deist F, Carlier F, Bouneaud C, Hue C, De Villartay JP, Thrasher AJ, Wulffraat N, Sorensen R, Dupuis-Girod S, Fischer A, Davies EG, Kuis W, Leiva L, Cavazzana-Calvo M. Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *N Engl J Med* 2002;**346**:1185–93.

89. **Aiuti A,** Cattaneo F, Galimberti S, Benninghoff U, Cassani B, Callegaro L, Scaramuzza S, Andolfi G, Mirolo M, Brigida I, Tabucchi A, Carlucci F, Eibl M, Aker M, Slavin S, Al-Mousa H, Al Ghonaium A, Ferster A, Duppenthaler A, Notarangelo L, Wintergerst U, Buckley RH, Bregni M, Marktel S, Valsecchi MG, Rossi P, Ciceri F, Miniero R, Bordignon C, Roncarolo MG. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N Engl J Med* 2009;**360**:447–58.

90. **Bordignon C,** Notarangelo LD, Nobili N, Ferrari G, Casorati G, Panina P, Mazzolari E, Maggioni D, Rossi C, Servida P, Ugazio AG, Mavilio F. Gene therapy in peripheral blood lymphocytes and bone marrow for ADA-immunodeficient patients. *Science* 1995;**270**:470–5.

91. **Ott MG,** Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, Glimm H, Kuhlcke K, Schilz A, Kunkel H, Naundorf S, Brinkmann A, Deichmann A, Fischer M, Ball C, Pilz I, Dunbar C, Du Y, Jenkins NA, Copeland NG, Luthi U, Hassan M, Thrasher AJ, Hoelzer D, von Kalle C, Seger R, Grez M. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* 2006;**12**:401–9.

92. **Cartier N,** Hacein-Bey-Abina S, Bartholomae CC, Veres G, Schmidt M, Kutschera I, Vidaud M, Abel U, Dal-Cortivo L, Caccavelli L, Mahlaoui N, Kiermer V, Mittelstaedt D, Bellesme C, Lahlou N, Lefrere F, Blanche S, Audit M, Payen E, Leboulch P, l'Homme B, Bougneres P, Von Kalle C, Fischer A, Cavazzana-Calvo M, Aubourg P. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* 2009;**326**:818–23.

93. **Cavazzana-Calvo M,** Payen E, Negre O, Wang G, Hehir K, Fusil F, Down J, Denaro M, Brady T, Westerman K, Cavallesco R, Gillet-Legrand B, Caccavelli L, Sgarra R, Maouche-Chrétien L, Bernaudin F, Girot R, Dorazio R, Mulder GJ, Polack A, Bank A, Soulier J, Larghero J, Kabbara N, Dalle B, Gourmel B, Socie G, Chrétien S, Cartier N, Aubourg P, Fischer A, Cornetta K, Galacteros F, Beuzard Y, Gluckman E, Bushman F, Hacein-Bey-Abina S, Leboulch P. Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. *Nature* 2010;**467**:318–22.

94. **Maguire AM,** High KA, Auricchio A, Wright JF, Pierce EA, Testa F, Mingozzi F, Bennicelli JL, Ying GS, Rossi S, Fulton A, Marshall KA, Banfi S, Chung DC, Morgan JI, Hauck B, Zelenaia O, Zhu X, Raffini L, Coppieters F, De Baere E, Shindler KS, Volpe NJ, Surace EM, Acerra C, Lyubarsky A, Redmond TM, Stone E, Sun J, McDonnell JW, Leroy BP, Simonelli F, Bennett J. Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. *Lancet* 2009;**374**:1597–605.

95. **Aparicio SA,** Huntsman DG. Does massively parallel DNA resequencing signify the end of histopathology as we know it? *J Pathol* 2010;**220**:307–15.

96. **Ng PC,** Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucl Acids Res* 2003;**31**:3812–14.

97. **1000 Genomes Project Consortium.** A map of human genome variation from population-scale sequencing. *Nature* 2010;**467**:1061–73.

98. **Robinson JT,** Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6.

99. **Kent WJ,** Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res* 2002;**12**:996–1006.

100. **Kaye J,** Boddington P, de Vries J, Hawkins N, Melham K. Ethical implications of the use of whole genome methods in medical research. *Eur J Hum Genet* 2010;**18**:398–403.

101. **Lowrance WW,** Collins FS. Ethics. Identifiability in genomic research. *Science* 2007;**317**:600–2.

102. **Caulfield T,** McGuire AL, Cho M, Buchanan JA, Burgess MM, Danilczyk U, Diaz CM, Fryer-Edwards K, Green SK, Hodosh MA, Juengst ET, Kaye J, Kedes L, Knoppers BM, Lemmens T, Meslin EM, Murphy J, Nussbaum RL, Otlowski M, Pullman D, Ray PN, Sugarman J, Timmons M. Research ethics recommendations for whole-genome research: consensus statement. *PLoS Biol* 2008;**6**:e73.

103. **McGuire AL,** Caulfield T, Cho MK. Research ethics and the challenge of whole-genome sequencing. *Nat Rev Genet* 2008;**9**:152–6.

104. **Kaye J,** Heeney C, Hawkins N, de Vries J, Boddington P. Data sharing in genomics: re-shaping scientific practice. *Nat Rev Genet* 2009;**10**:331–5.

105. **Knoppers BM,** Brand AM. From community genetics to public health genomics — what's in a name? *Public Health Genomics* 2009;**12**:1–3.

106. **Lacroix M,** Nycum G, Godard B, Knoppers BM. Should physicians warn patients' relatives of genetic risks? *CMAJ* 2008;**178**:593–5.

107. **Knoppers BM,** Joly Y, Simard J, Durocher F. The emergence of an ethical duty to disclose genetic research results: international perspectives. *Eur J Hum Genet* 2006;**14**:1170–8.

108. **Wolf SM,** Lawrenz FP, Nelson CA, Kahn JP, Cho MK, Clayton EW, Fletcher JG, Georgieff MK, Hammerschmidt D, Hudson K, Illes J, Kapur V, Keane MA, Koenig BA, Leroy BS, McFarland EG, Paradise J, Parker LS, Terry SF, Van Ness B, Wilfond BS. Managing incidental findings in human subjects research: analysis and recommendations. *J Law Med Ethics* 2008;**36**:219–48, 1.