

Automation of Molecular-Based Analyses: A Primer on Massively Parallel Sequencing

Lan Nguyen,¹ *Leslie Burnett^{1,2,3}

¹Pathology North, NSW Health Pathology Service, Royal North Shore Hospital, St Leonards, NSW 2065, Australia; ²SEALS, NSW Health Pathology Service, Prince of Wales Hospital, Randwick NSW 2031, Australia; ³Sydney Medical School, Royal North Shore Hospital – E25, University of Sydney, Sydney, NSW 2006, Australia.

*For correspondence: Professor Leslie Burnett, leslie.burnett@sydney.edu.au

Abstract

Recent advances in genetics have been enabled by new genetic sequencing techniques called massively parallel sequencing (MPS) or next-generation sequencing. Through the ability to sequence in parallel hundreds of thousands to millions of DNA fragments, the cost and time required for sequencing has dramatically decreased. There are a number of different MPS platforms currently available and being used in Australia. Although they differ in the underlying technology involved, their overall processes are very similar: DNA fragmentation, adaptor ligation, immobilisation, amplification, sequencing reaction and data analysis. MPS is being used in research, translational and increasingly now also in clinical settings. Common applications include sequencing of whole genomes, whole exomes or targeted genes for disease-causing gene discovery, genetic diagnosis and targeted cancer therapy. Even though the revolution that is occurring with MPS is exciting due to its increasing use, improving and emerging technologies and new applications, significant challenges still exist. Particularly challenging issues are the bioinformatics required for data analysis, interpretation of results and the ethical dilemma of ‘incidental findings’.

Introduction

For those who learnt about molecular-based techniques prior to 2005, it was a landscape dominated by bacterial cloning, polymerase chain reactions (PCR) and Sanger sequencing. Using these techniques, the Human Genome Project (HGP) was completed in 2003, taking 13 years at the cost of US\$2.7 billion.^{1,2} The publication of the 2.85 billion nucleotide sequence was heralded as a landmark event, laying the foundation for a new genomic era. At the time, a major contributor to the HGP published its ‘vision for the future of genomics research’ and one goal that seemed ‘so far off as to be almost fictional’, was the ability to sequence the human genome for less than \$1,000.³ Yet in less than 11 years, it appears this dream may be a reality, with a recent manufacturer’s claim that 18,000 whole human genome sequences could be produced in a year at a cost of less than \$1,000 each.⁴ What has driven such a rapid and dramatic decline in costs and time? The answer is massively parallel sequencing (MPS). Given how it is revolutionising practice in many areas of clinical medicine, we in laboratory medicine should at least have a basic understanding of the technology and its current and future applications.

What is Massively Parallel Sequencing?

MPS is also referred to as next-generation sequencing (NGS). In MPS, hundreds of thousands to millions of DNA sequencing reactions are performed simultaneously, or ‘in parallel’, with each sequencing step being coupled with detection. This is in contrast to traditional Sanger sequencing (or 1st generation sequencing), where sequencing of each separate DNA sequence with fluorescent-labelled dideoxy terminator nucleotides is completed before the labelled DNA fragments are separated and detected by capillary electrophoresis. The throughput of automated Sanger sequencing is limited by the capacity of thermal cyclers and capillary electrophoresis analysers, with the most advanced capillary analysers capable of an output of approximately 500 kilobases (kb) in 24 h,⁵ orders of magnitude less than MPS.

Though the various MPS platforms employ different technologies, their underlying workflow is similar:

1. **Fragmentation of DNA**- randomly by enzymatic digestion, nebulisation or sonication. Optimum fragment length is platform-dependent, ranging from 50 to 20,000 base pairs (bp).

Box 1. Glossary of terms.

Term	Definition
Adaptor	Platform specific oligonucleotides which are ligated to DNA fragments. May immobilise library to solid support or act as an amplification primer or sequencing primer.
Alignment	Process of mapping read sequences to a reference sequence.
Annotation	Process of adding biological information to a sequence, such as identifying known pathogenic variants and their clinical significance.
Base calling	Conversion of the sequencing signal produced to a nucleotide sequence.
Coverage	The percentage of target bases that are sequenced a given number of times. Often used interchangeably with depth of coverage.
Depth of coverage	The number of times a particular nucleotide has been sequenced. Usually refers to only high-quality aligned reads.
Filtering	Process of using bioinformatics tools to reduce the amount of data analysed, usually with user-defined criteria.
Library	The collection of adaptor-ligated DNA fragments to be sequenced.
Read	Nucleotide sequence of a single DNA fragment.
Read length	Number of nucleotides sequenced per DNA fragment.
Variant calling	Process of identifying differences between aligned read sequences and a reference sequence.

2. **Ligation to adaptor sequences**- these are platform-specific sequences that are ligated to the ends of DNA fragments, creating the sequencing library.
3. **Immobilisation**- through the adaptor sequence to a solid surface, such as a bead or a glass slide.
4. **Amplification**- most platforms require amplification to increase the signal for detection. This can be achieved through emulsion bead PCR or surface cluster PCR.
5. **Sequencing**- cycles of base incorporation by synthesis or ligation are followed immediately by signal detection. Signals are converted to base calls, from which a nucleotide sequence or 'read' is produced. The read length is platform-dependent but is usually shorter than Sanger sequencing reads, which may be 800 to 1000 bp. Each template DNA region is also sequenced a number of times (depth of coverage). Sequencing may also be paired-end or mate-paired. In paired-end sequencing, <1000 bp length fragments are sequenced from both ends. In mate-paired sequencing, >1000 bp length fragments are circularised by ligating the ends to a single adaptor. These are fragmented to create linear fragments with a central adaptor, another two adaptors are ligated to the ends and fragments are sequenced from both ends.⁶
6. **Data analysis**- due to the short read lengths, there is limited ability to reassemble a genome through overlapping sequences. Greater certainty in alignment can be achieved with longer read lengths or using paired-end or mate-

paired sequencing as paired reads are aligned together. Most applications of MPS involve 'resequencing', where sequencing is used to search for variations from 'normal' using a known reference genome to provide a template for read alignment. Differences between the alignment and reference sequence are identified ('variant calling'), filtered and annotated to identify those that may be clinically significant.⁷

The general advantages when using MPS for sequencing multiple genes compared to Sanger sequencing are the requirement for less template DNA, greater speed and lower cost per base.

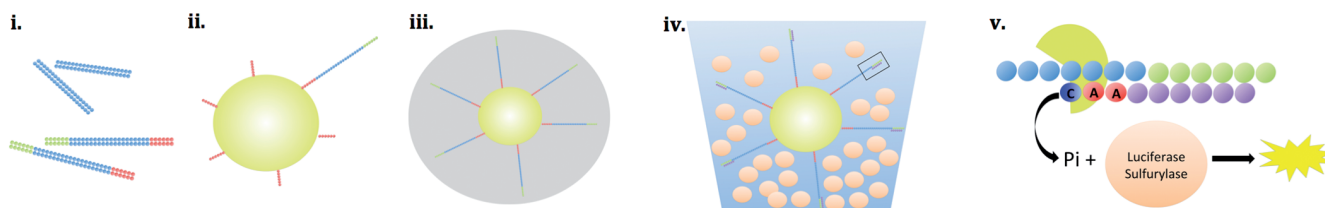
MPS Platforms Available

The MPS platforms available are still evolving, with improvements in current technologies and the development of even newer technologies. It is such a rapidly changing field that one of the pioneering instruments, the 454 by Life Sciences/Roche, is already in the process of being phased out and will not be supported past mid-2016.⁸ Following is a description of the major commercially available platforms worldwide.

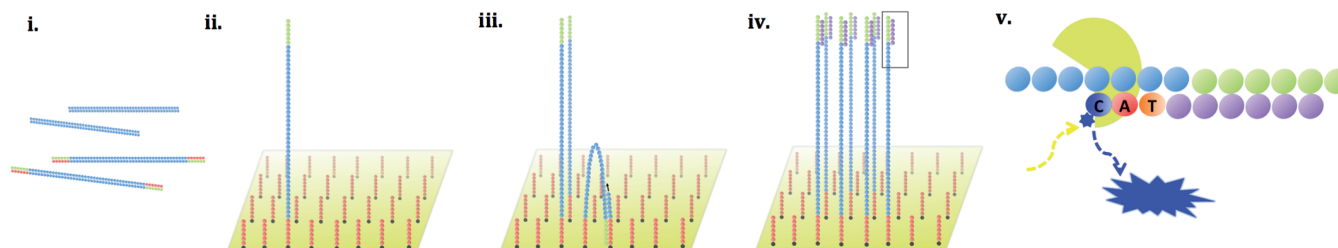
a. 454 Life Sciences/Roche (Genome Sequencer Junior/FLX) - Pyrosequencing

Double-stranded DNA (dsDNA) is fragmented into

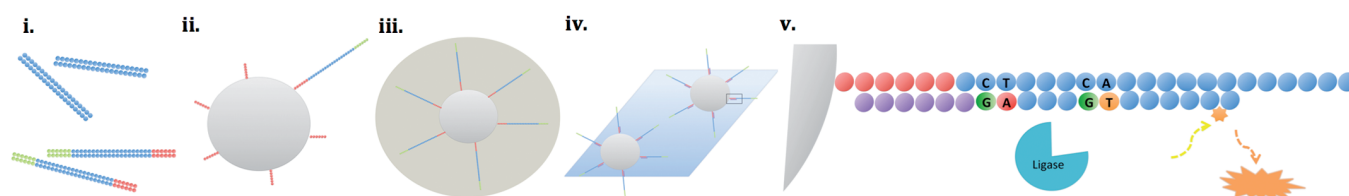
a. 454 Life Sciences/Roche- Pyrosequencing



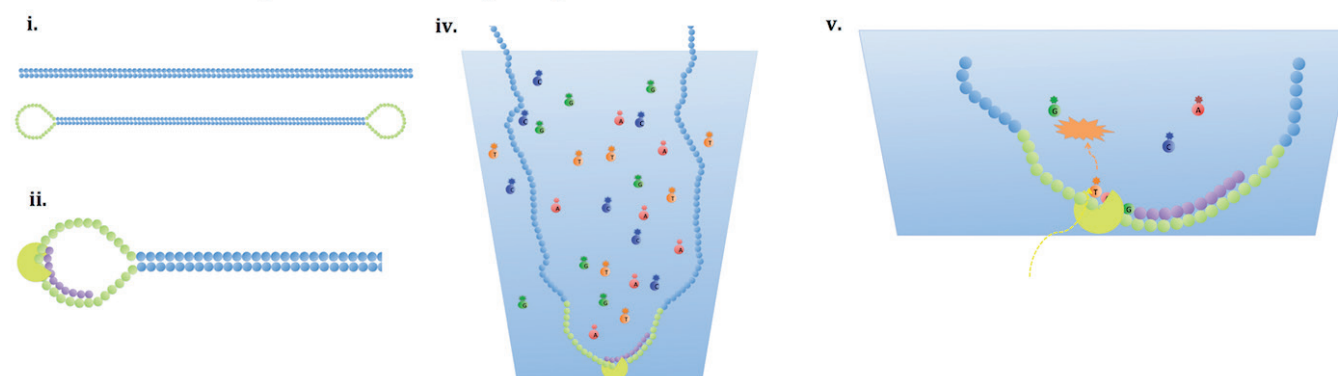
b. Illumina- Sequencing by synthesis



c. Applied Biosystems- Sequencing by ligation



d. Pacific Biosciences- Single-molecule real-time sequencing



e. Life Technologies- Sequencing by monitoring pH

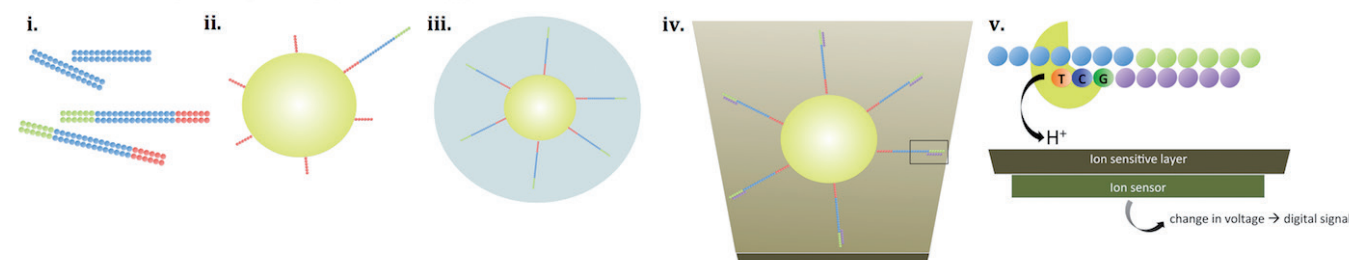


Figure 1. Workflow of different next-generation sequencing technologies. **i.** DNA fragmentation and ligation of adaptors. **ii.** Immobilisation to a solid surface. **iii.** Amplification. **iv.** Sequencing reaction. **v.** Signal detection.

400–600 bp lengths, adaptors are attached the ends, after which the dsDNA is denatured to produce single-stranded DNA (ssDNA) fragments with different adaptors at each end. This DNA library is mixed with capture beads, PCR reagents and an emulsion oil at a limiting dilution so as to create water droplets, each containing a single bead and a library fragment. PCR is performed within this droplet (emulsion PCR), creating millions of copies of amplified DNA fragments immobilised onto the bead. The beads are put in a plate comprised of 1.6 million, 75 picolitre wells, with each well designed so it can accommodate only one capture bead per well. In each sequencing cycle in turn, reagent containing only one of the four possible nucleotides (A, C, G or T) is washed over the plate and if the nucleotide is incorporated into the DNA template, there is release of a pyrophosphate. Within each well are also smaller beads coated with ATP sulfurylase and luciferase, which generate a chemiluminescent signal in the presence of pyrophosphate. This signal is detected by a camera, and is proportional to the number of bases incorporated. During a sequencing run, this cycle is repeated with the four possible different bases, always in a set order. Thus by monitoring the light signal generated by each well after a particular base is added, a ‘flowgram’ is created through which each DNA fragment sequence can be deduced (Figure 1a).⁹⁻¹¹

b. Illumina (MiSeq/NextSeq/HiSeq) - Sequencing by synthesis

DNA is fragmented into ssDNA fragments of less than 600 bp. The fragment ends are repaired, creating an ‘A’ overhang for ligation of different adaptors at each end. The DNA library is immobilised onto a lane of a flow cell, consisting of a glass surface coated with millions of primers complementary to the ligated adaptor sequences. Each library fragment is amplified by ‘bridge-PCR’, whereby different reagent washes initiate annealing of the free adaptor end with a complementary primer, extension by DNA polymerase, then denaturation of the double stranded ‘bridge’. Through repeating this cycle, a ‘cluster’ of thousands of copies of each library fragment is created, with up to 10 million clusters per square centimetre on a flow cell. Sequencing is initiated by the addition of fluorescent reversible terminator nucleotides, DNA polymerase and universal sequencing primers. In each cycle, only a single base is incorporated, a laser excites the fluorophore and the emitted light is detected by a camera. A cleavage reagent is added to remove the fluorophore and terminator, allowing the next cycle of nucleotide extension and detection (Figure 1b).¹²⁻¹⁴

c. Applied Biosystems (SOLiD 5500/5500xl W)- Sequencing by ligation

DNA is fragmented into 400–850 bp length fragments and different adaptors are ligated to each end. Emulsion PCR is used to immobilise and amplify the library fragment onto a paramagnetic bead, which is then immobilised on a glass slide. A universal sequencing primer is hybridised to an adaptor, then a ligation cycle commences with ligation of a dye-labelled probe complementary to the next two-base sequence, fluorescence detection, then removal of the label by cleavage. This ligation cycle is repeated a number of times, with each ligated probe assaying a two-base sequence every cycle, creating a sequence of fluorescent signals called a colour string. The extension product is then removed and a second universal sequencing primer offset by one base from the previous is bound and a second round of ligation cycles ensues. In total, 5 primer rounds are completed so that in theory, each base is interrogated by two independent ligation reactions. Four different fluorescent dyes encode for the 16 possible two-base combinations. Colour strings from each primer round are converted to all possible two-base sequence combinations and compared to deduce the correct read sequence (Figure 1c).¹⁵⁻¹⁷

d. Pacific Biosciences (PacBioRS II)- Single-molecule real-time (SMRT) sequencing

DNA is randomly fragmented into 10–20 kb double-stranded fragments. The ends are repaired and ligated to an adaptor with a ‘T’ overhang. A hairpin shaped universal adaptor is then ligated to each end, creating a circular template. Sequencing primers are annealed to the templates, which are then bound to DNA polymerase. Prepared template is loaded onto a SMRT cell, consisting of 150,000 zero-mode waveguides (ZMW), nanostructures measuring 100 nm in diameter in which a single primer bound template-DNA polymerase complex is immobilised. Fluorescently-labelled nucleotides are incorporated by DNA synthesis and the fluorescence signal is detected in each ZMW. Incorporation cleaves the fluorophore allowing the subsequent labelled nucleotide to be added. Smaller fragments may be sequenced in both the sense and antisense direction multiple times, increasing base calling accuracy (Figure 1d).¹⁸⁻²⁰

e. Life Technologies (Ion Torrent Ion Proton/Ion PGM)- Sequencing by monitoring pH

DNA is fragmented into 200–250 bp length fragment and different adaptors are ligated to each end. The sequencing library is amplified on capture beads by emulsion PCR. Beads now coated with amplified template are put onto a chip containing up to 7 million wells, each able to

accommodate a single coated bead. In each sequencing cycle, reagent containing one of the four possible nucleotides is washed over the chip and, if the base is incorporated, there is release of a proton. The change in pH (~0.02 pH units per base incorporation) within a well is detected by the ion sensor and converted to a digital signal. In successive cycles, the different nucleotides are washed over the chip in a set order to produce a read (Figure 1e).²¹⁻²³

Applications of MPS

MPS is being increasingly used in many areas of medicine, both in the research and clinical setting. Depending on the purpose, different nucleic acid targets may be sequenced:

- Whole genome
- Whole exome
- Targeted genes
- Transcriptome - RNA sequence including mRNA, tRNA
- Mitochondrial genome
- Metagenome-microbial DNA in biological samples
- Methylome - DNA sequence including cytosine methylation state
- Chromatin immunoprecipitation-Sequencing (ChIP-Seq): regions of DNA that bind to transcription factors, co-factors or chromatin proteins

Whole Genome Sequencing

MPS can be used for the de novo assembly of whole genomes, or to identify genetic variations in the genome that may be associated with or cause disease. The advantage of whole genome sequencing (WGS) is that it can identify a wide range of genetic variations such as single nucleotide polymorphisms (SNPs), small insertions or deletions (indels), copy number variations (CNVs) and structural variants. The 1000 Genomes Project, for example, is an ongoing international collaboration involving WGS of individuals of different ancestry with an aim of providing reference data for common SNPs in different population groups.²⁴ The public data can then be used in the analysis of genome-wide association studies or to help identify disease-causing variants. WGS is also being performed on cancer tissue to identify somatic mutations which may aid in cancer diagnosis, prognostication and targeted therapy.²⁵ The identification of CNVs is possible using specialised bioinformatics tools that for example estimate copy number based on the number of reads or distance between paired-end reads, although not all CNVs are currently detected. The better resolution and overall coverage of WGS combined with improvements in data analysis may see it supplant cytogenetic techniques as the preferred test for CNVs in the near future.²⁶

Whole Exome Sequencing

Only ~1–2% of the human genome consists of exons, the

protein-coding regions of DNA, yet this is where 85% of disease-causing variants are located.²⁷ Accordingly, whole exome sequencing (WES) has become the more favoured cost- and time-effective approach, particularly for discovering causative gene mutations for Mendelian disorders. To perform WES, exome-containing libraries are captured and enriched by kit-specific oligonucleotide baits before undergoing sequencing.²⁸ WES was first used to identify the cause of rare genetic disorders in 2009 and by the end of 2012, over 180 novel disease-causing genes had been described. It has been predicted that the remaining causes of Mendelian disorders will be identified before 2020.²⁹ To demonstrate its use in clinical practice, 250 patients with a suspected genetic disorder had WES performed and a genetic diagnosis was made in 62 patients. This diagnostic rate of 25% is generally higher than that of other genetic tests and is likely to improve as further novel disease-causing variants are identified.³⁰

Targeted Gene Sequencing

Where there is a specific genetic question in mind, it may be more efficient to sequence specific genes rather than the entire genome or exome. This approach could also be used to resequence areas of interest identified through WGS or WES as the overall and depth of coverage is usually greater. Similar to WES, libraries of interest may be captured and enriched by specific baits. Targeted genes may also be selected by multiplexed PCR followed by amplicon library preparation. Panels are commercially available for disease groups such as cancers and cardiomyopathies, and it is possible to design customised baits or primer pairs for specific targets.^{31,32} Another consideration is that targeted gene sequencing generates less data than WGS or WES. When one is interested only in a specific genomic region, having less data means a simpler bioinformatics analysis, and may also avoid completely the issue of incidental findings (see later) being generated.

While this approach is well suited for batched high volume testing, it is not efficient for low volume testing. Furthermore, with the speed at which disease-associated genetic variations are being discovered, the need to update panels or resequence a patient may also negate the initial efficiency of targeting sequencing.

Limitations of MPS

Errors and Incomplete Coverage

Errors and biases can be introduced during any of the steps involved in MPS. The different technologies are all prone to different sequencing errors to varying degrees. For example, the Ion PGM has difficulty accurately sequencing homopolymers greater than 8 bases long, while the Illumina MiSeq has difficulties with GC-rich motifs.^{33,34} Base calling errors are estimated by a base call quality score (Qscore).

Quality scores in MPS are a similar concept to the Phred quality score developed for Sanger sequencing.³⁵ If p is the estimated error probability, $Qscore = -10 \cdot \log_{10}(p)$. For example, if the error probability of a base call is 1/100, the base call Qscore is 20. Base calling algorithms are platform specific, and convert the signal detected to a base call with a Qscore. The probability of base call errors depends on parameters such as the signal to noise ratio, cluster/bead density and dephasing (loss of sequence cycle synchronicity due to incomplete or extra extension). These errors may be better identified (and therefore not be translated into errors in variant calling) by increasing depth of coverage, replicate sequencing or sequencing across different platforms,³⁶ but this increased volume of sequencing may significantly increase overall costs. Due to these limitations, Sanger sequencing, which has a lower raw error rate, is still considered the ‘gold standard’ and used to confirm variants identified through MPS.^{30,32}

The overall error rate is a determinant of the sensitivity and specificity for detecting low frequency variants. This is often the case when identifying somatic changes in cancers, where samples may be heterogeneous due to low amount of tumour tissue vs normal tissue or presence of multiple tumour subclones. A 30X depth of coverage is generally considered sufficient to identify SNPs, while for cancer genomes it may need to be much higher (500–1000X), depending on the mutation frequency and platform used.^{37,38} Incomplete coverage, where regions are not sequenced or poorly sequenced, is also a problem seen with all MPS technologies. This may be due to amplification or enrichment biases and inherent difficulty in sequencing particularly GC-rich and GC-poor regions.³⁴ It may be necessary to perform Sanger sequencing to fill-in these gaps in coverage and completely sequence areas of interest.

Bioinformatics and Interpretation

One of the major challenges of MPS is making sense of the data generated, which may be millions to billions of read sequences. It has been remarked that though a genome may cost \$1000 to sequence, its analysis and interpretation may cost over \$10,000.³⁹ Powerful computational and bioinformatics tools are required to store the data and perform the various steps involved in read alignment, variant calling and annotation. Filtering of data is important to ‘clean up’ and help to reduce the amount of data analysed and reported variants. For example, low-quality reads or reads that map to multiple regions may be excluded from further analysis. Variant filtering may be used to select for variants only in exonic regions, particular genes of interest or remove common SNPs. Numerous software tools and packages are available to analyse MPS data and depending on the preferences of the

end-user, are combined in an analysis workflow, sometimes referred to as a bioinformatics pipeline.

Both false positive and false negative variant calls may be made due to errors in alignment, alignment with an unsuitable reference sequence or too stringent or lenient filtering. Correct annotation is highly dependent on the accuracy of information obtained from interrogated databases, such as those containing known disease-causing mutations, common polymorphisms or mutations in cancer.⁷ Functional characterisation of novel variants is also difficult. The pathogenicity of a variant may be inferred by examining conservation across species, predicted changes in protein structure and function or by performing functional studies, but often the significance is uncertain.

Furthermore, with the fast pace at which genetic discoveries are being found, variants currently thought to be of unknown significance may later be found to disease-associated or associated with a polygenic condition; the converse may also apply, in that a variant currently thought to be pathogenic may turn out to be a harmless variant. This has led to currently active discussion on where the responsibility lies for possibly re-annotating variants based on new data.⁴⁰ While the testing laboratory is in the best position to re-annotate the known sequence, the requesting clinician in consultation with the patient may, or may not, wish this to occur. Conversely, leaving it to the patient or the clinician to instigate re-annotation may lead to delayed or overlooked reporting of a newly-recognised pathogenic variant that the laboratory could readily identify, but for which it may either lack consent or operational capability.

Incidental Findings

WGS and WES are unlike most diagnostic tests where the test performed is specific to the clinical indication. Thus, it is inevitable that WGS and WES identifies pathogenic or likely pathogenic alterations in genes that are unrelated to the clinical indication i.e. incidental findings. In 2013, the American College of Medical Genetics and Genomics (ACMG) published recommendations that incidental findings in 57 genes should be reported, regardless of patient age or preference, due to the benefit of early identification and intervention in the 24 conditions they could cause.⁴¹ These recommendations created controversy and were not endorsed by other national or supranational peak professional bodies, including the Human Genetics Society of Australasia (HGSA), as the ACMG recommendations conflicted with various local guidelines in other jurisdictions regarding patient autonomy and predictive testing in minors.⁴² Only recently have the ACMG recommendations been updated to allow patients to opt out of receiving incidental findings results.⁴³ If incidental findings are to be reported, then both HGSA and now also

ACMG recommend comprehensive pre-test counselling to obtain informed consent. With MPS being more widely adopted and possibly moving out of the specialist genetic laboratory arena, the ability of requesting clinicians to give this complex counselling or whether there are sufficient experts (e.g. genetic counsellors, clinical geneticists) to play this role, needs to be addressed.

The Future

Numerous new MPS technologies and applications are on the horizon. Oxford Nanopore Technologies have developed a single molecule, nanopore-based MPS technology that is currently being tested in laboratories worldwide. Their MinION system is a disposable, USB flash-drive sized device that can be plugged into the USB port of a computer and works directly with blood or serum.⁴⁴ The prospect of newborn screening by WGS or WES could also become a possibility with recent US government funding of a pilot study to investigate whether it improves the understanding and treatment of disorders in newborns and the ethical, legal and social implications.⁴⁵ In Australia, the state of genetic testing will change significantly with the increased adoption of MPS. For example, the Garvan Institute of Medical Research in Sydney is one of the first four sites worldwide to acquire the high-throughput 'USD\$1000 Genome' instruments and will soon be offering WGS for making genetic diagnoses and guiding cancer treatment.⁴⁶

Genomics has the potential to make major contributions to our understanding of disease pathogenesis, and will undoubtedly lead to changes in diagnostic pathology practice.

Acknowledgements: We wish to thank Dr Lisa Koe FRCPA, Ms Anné Proos FFSc(RCPA) and Mr Peter Ward FAACB for their review and feedback on this article.

Competing Interests: None declared.

References

1. National Human Genome Research Institute. The Human Genome Project Completion: Frequently Asked Questions. <http://www.genome.gov/11006943> (Accessed 15 February 2014).
2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931-45.
3. Collins FS, Green ED, Guttacher AE, Guyer MS; US National Human Genome Research Institute. A vision for the future of genomics research. *Nature* 2003;422:835-47.
4. Hayden EC. Is the \$1,000 genome for real? <http://www.nature.com/news/is-the-1-000-genome-for-real-1.14530> (Accessed 22 February 2014).
5. Applied Biosystems. Applied Biosystems® 3500 and 3500xL Genetic Analyzers -Specifications sheet.
6. Mardis ER. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)* 2013;6:287-303.
7. Dolled-Filhart MP, Lee M Jr, Ou-Yang CW, Haraksingh RR, Lin JC. Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. *Scientific World Journal* 2013;2013:730210.
8. GenomeWeb. Roche shutting down 454 sequencing business. <http://www.genomeweb.com/sequencing/roche-shutting-down-454-sequencing-business> (Accessed 15 February 2014).
9. Roche Diagnostics Corporation. 454 Sequencing. <http://www.454.com> (Accessed 15 February 2014).
10. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembien LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:376-80.
11. How genome sequencing is done. http://www.454.com/downloads/news-events/how-genome-sequencing-is-done_FINAL.pdf (Accessed 15 February 2014).
12. Illumina. <http://www.illumina.com> (Accessed 15 February 2014).
13. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008;456:53-9.
14. Illumina. Targeted Next-Generation Sequencing for Forensic Genomics. http://res.illumina.com/documents/products/appspotlights/app_spotlight_forensics.pdf (Accessed 18 February 2014).
15. SOLiD® Next-Generation Sequencing. <http://www.lifetechnologies.com/au/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing.html> (Accessed 15 February 2014).
16. SOLiD™ Analysis Tools (SAT) v3.5 Reference Guide. <https://tools.lifetechnologies.com/content/sfs/manuals/4443929.pdf> (Accessed 15 February 2014).
17. Breu H. A Theoretical Understanding of 2 Base Color Codes and its Application to Annotation, Error Detection, and Error Correction. https://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_058265.pdf (Accessed 18 February 2014).
18. Pacific Biosciences. <http://www.pacificbiosciences.com/> (Accessed 15 February 2014).
19. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science* 2009;323:133-8.
20. Pacific Biosciences. Template Preparation and Sequencing Guide. <https://na5.salesforce.com/sfc/p/70000000IVif/a/70000000PVYM/9Jo8G7kak2lGutgT6Q4IJMEmmlUb5c3L9lPwCqLbmRQ=> (Accessed 15 March 2014).
21. Life Technologies - Ion Torrent™. <http://www.lifetechnologies.com/au/en/home/brands/ion-torrent.html> (Accessed 15 February 2014).
22. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski

- W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011;475:348-52.
23. Life Technologies. Ion Torrent: Amplicon sequencing. http://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_094273.pdf (Accessed 25 February 2014).
24. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al; 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-73.
25. Wang L, Wheeler DA. Genomic sequencing for cancer diagnosis and therapy. *Annu Rev Med* 2014;65:33-48.
26. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics* 2013;14(Suppl 1):S1.
27. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J Med Genet* 2011;48:580-9.
28. Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* 2011;29:908-14.
29. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013;14:681-91.
30. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* 2013;369:1502-11.
31. Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, van den Berg MP, et al. Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Hum Mutat* 2013;34:1035-42.
32. Lepri FR, Scavelli R, Digilio MC, Gnazzo M, Grotta S, Dentici ML, et al. Diagnosis of Noonan syndrome and related disorders using target next generation sequencing. *BMC Med Genet* 2014;15:14.
33. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 2012;13:341.
34. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. *Genome Biol* 2013;14:R51.
35. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998;8:186-94.
36. Robasky K, Lewis NE, Church GM. The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet* 2014;15:56-62.
37. Mwenifumbo JC, Marra MA. Cancer genome-sequencing study design. *Nat Rev Genet* 2013;14:321-32.
38. Sims D, Sudbery I, Illott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014;15:121-32.
39. Mardis ER. The \$1,000 genome, the \$100,000 analysis? *Genome Med* 2010;2:84.
40. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, et al; Molecular Subcommittee of the ACMG Laboratory Quality Assurance Committee. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med* 2008;10:294-300.
41. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al; American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013;15:565-74.
42. Human Genetics Society of Australasia. HGSA commentary on ACMG recommendations. <http://www.hgsa.org.au/hgsanews/hgsa-commentary-on-acmg-recommendations> (Accessed 12 April 2014).
43. American College of Medical Genetics. ACMG Updates Recommendation on "Opt Out" for Genome Sequencing Return of Results. https://www.acmg.net/docs/Release_ACMGUpdatesRecommendations_final.pdf (Accessed 12 April 2014).
44. Oxford Nanopore Technologies. The MinION device. <https://www.nanoporetech.com/technology/the-minion-device-a-miniaturised-sensing-system/the-minion-device-a-miniaturised-sensing-system> (Accessed 12 April 2014).
45. NIH program explores the use of genomic sequencing in newborn healthcare. <http://www.nih.gov/news/health/sep2013/nhgri-04.htm> (Accessed 12 April 2014).
46. Garvan Institute. HiSeq X Ten. <http://www.garvan.org.au/research/clinical-genomics/hiseqXTen> (Accessed 12 April 2014).