

# DNA SEQUENCING towards the third generation and beyond

Next-generation sequencing has expanded the frontiers of genomic research, opening up new avenues of enquiry and offering insights into diverse areas of biology. Rapid developments in this field are dramatically lowering costs and increasing speeds. With the arrival of third generation sequencing, this continually evolving technology is now being applied within clinical environments, where it has the potential to improve diagnosis and treatment outcomes of diseases such as cancer. There are still challenges facing sequencing today, in many ways magnified by the current economic climate, but these are being addressed by improvements across the sequencing ecosystem, but particularly in workflow, sequencing chemistry and analysis. Expected developments in sequencing technology, including the arrival of nanopore technology, offer the prospect of yet more advances in system performance and miniaturisation, going well beyond the present generation.

**T**he genetic code is composed of the sequence of nucleotide bases which form the fundamental repetitive units of DNA. Deciphering the code – or the sequence of bases – offers insights into a vast repository of genetic information governing all manner of biological phenomena. This can be achieved by DNA sequencing. In recent years, sequencing via the conventional capillary-based Sanger method has been superseded by the advent of next-generation DNA sequencing technology, overcoming some of the limitations associated with the previous method.

Next-generation DNA sequencing encompasses several technologies utilising distinct approaches to sequencing biochemistry. A defining feature of next-generation sequencing (NGS) is its ability to

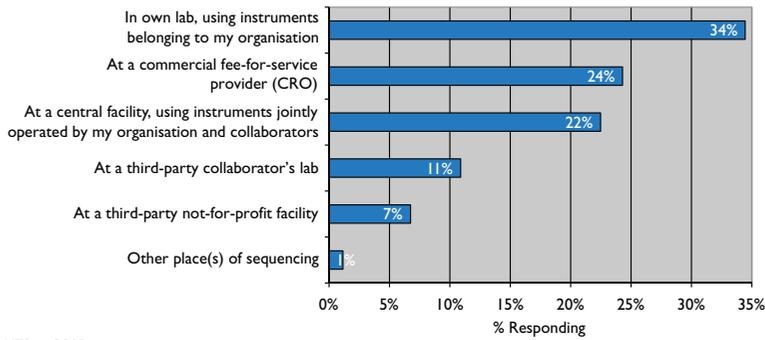
perform millions of sequencing reactions simultaneously. This attribute is referred to as massively parallel sequencing.

Despite the diversity in sequencing biochemistry, most NGS approaches retain a high degree of similarity in the steps preceding sequencing and data acquisition. Initially, genomic DNA is randomly fragmented. Fragments are subsequently ligated to adapter sequences, facilitating formation of arrayed clusters of PCR-amplified DNA. Each spatially distinct array feature forms the location of one of millions of ongoing sequencing reactions. The DNA sequence is finally determined as a result of enzyme-driven synthesis of the nucleic acid chain, using the cluster fragments as a template. A signal released upon incorporation of a specific

**By Andrew  
Szopa-Comley**

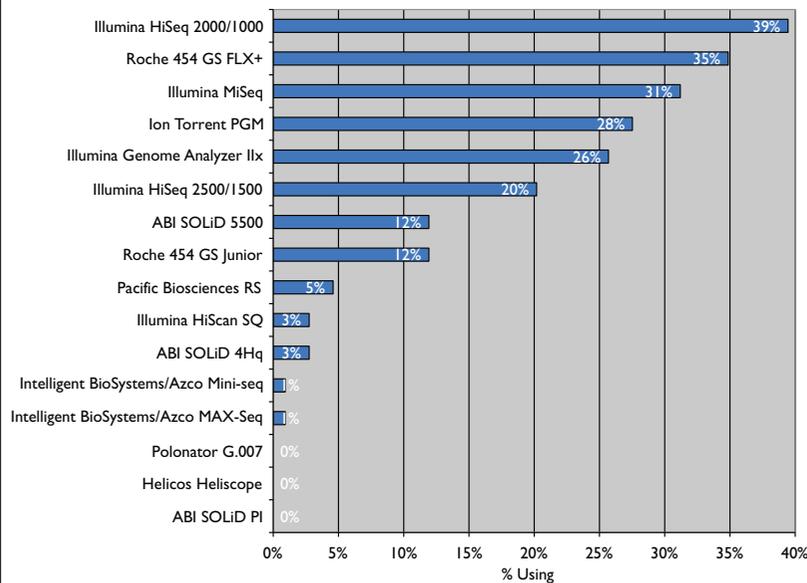
---

**Figure 1: Location of survey respondents' sequencing activities**



© HTStec 2012

**Figure 2: Use of different NGS instruments to generate sequencing data**



© HTStec 2012

base is then detected using an imaging-based system. Sequence data from each array feature is amassed through alternating cycles of enzyme-based biochemistry and imaging. Eventually the sequences corresponding to each array cluster (known as reads) are aligned to assemble the contiguous DNA sequence.

In the past decade, the rapid development of NGS technology has transformed the landscape of genomic research. Technological innovation has resulted in a dramatic reduction in the cost of sequencing. The ability to acquire affordable sequence data has enabled NGS technology to expand beyond the confines of large research centres into many smaller labs. When combined with the availability of the reference genomes for a growing number of organisms, this has precipitated an explosion in genomic research.

Researchers across many fields are using NGS technology to answer questions to diverse biological problems, ranging from analysis of genes commonly mutated in types of cancer to which gene *loci* promote speciation. A growing suite of applications of NGS has helped to reveal the intricacy of networks controlling gene expression. These approaches have provided insights into the epigenome, transcriptome and a multitude of protein-DNA interactions, hinting at the high levels of regulatory sophistication operating in cells. Recent research has even highlighted the importance of the expanses of non-coding DNA within the genome – which are far more extensive than coding regions, and were previously labelled as ‘junk’ DNA.

Although clearly a valuable tool for the investigation of a range of biological phenomena, the interpretation and analysis of the vast amounts of data being generated through NGS now represents a significant challenge. Similarly, improvements in read length and accuracy are desirable. In response to budgetary constraints in many research environments, efforts are also being focused on streamlining the sample preparation processes for numerous applications, and enhancing the efficiency of running a sequencing instrument.

A growing trend in sequencing has been the development of instruments capable of operating at higher speeds and producing longer read lengths. Some so-called ‘third generation’ sequencing platforms incorporate innovations such as single-molecule real-time sequencing, whereas others build upon existing approaches. It is expected that the further improvements offered by ‘third-generation’ platforms will facilitate transfer of sequencing into areas such as the clinical environment, where it could transform aspects of disease detection and

treatment. Many other potential approaches to sequencing (eg Oxford Nanopore Technologies) are at various stages of development towards full commercial release, generating considerable excitement and raising hopes of even more spectacular improvements. It seems that the emergence of 'third-generation' sequencing platforms and of novel approaches to sequencing chemistry will ensure that the sequencing revolution is likely to continue apace.

HTStec's Next-Generation Sequencing Trends 2012 survey and report<sup>1</sup> published in December 2012 set out to explore end-user experiences, practices, preferences and metrics in NGS and to understand future requirements. The report also details interest in purchasing new instruments, alternative purchasing scenarios and expectations for 'third generation' sequencing platforms. We now report on some of the survey findings and discuss them with reference to the latest developments in NGS.

### Where is sequencing being done?

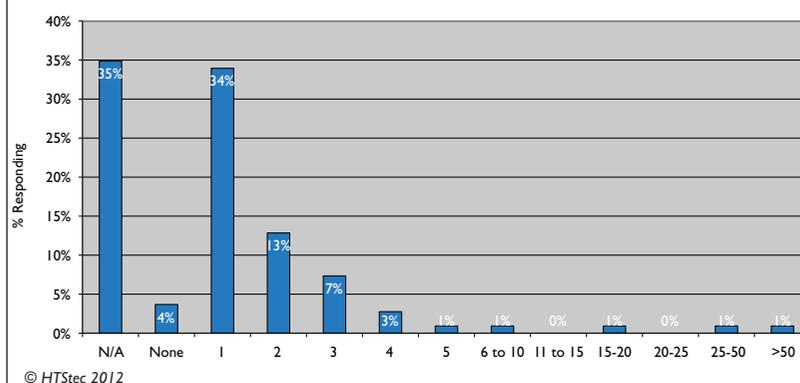
The location where most survey respondents' next generation sequencing (NGS) activities were undertaken was in their own lab, using instruments belonging to their organisation (34%). However, a substantial proportion of respondents carried out their NGS activities at commercial fee-for-service providers (CROs) (24%), or at a central facility using instruments jointly operated by their own organisation and collaborators (22%). Other locations for respondents' NGS activities included at a third-party collaborator's lab (11%), at a third-party not-for-profit facility (7%) and at other (undefined) places of sequencing (1%) (Figure 1).

### What sequencers are most used?

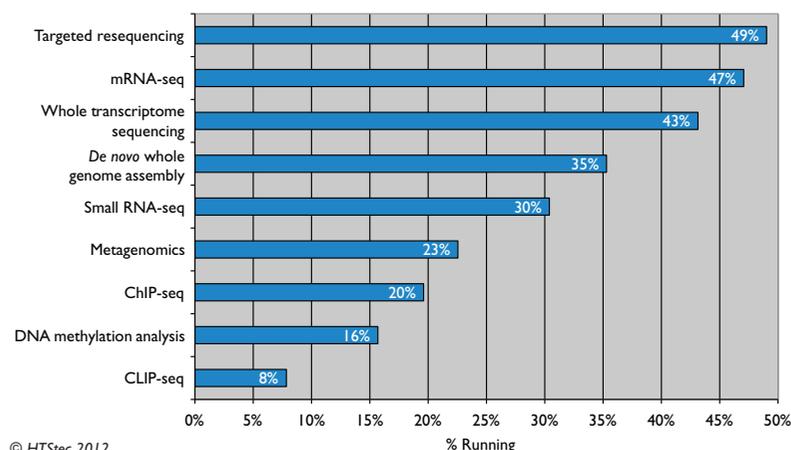
Survey respondents are using a wide range of NGS instrument platforms to generate sequencing data. The most used NGS instrument platform was the Illumina HiSeq 2000/1000 with 39% of respondents using. This was followed by 35% using the Roche 454 GS FLX+, 31% Illumina MiSeq, 28% Ion Torrent PGM, 26% Illumina Genome Analyser Ix, 20% Illumina HiSeq 2500/1500, 12% ABI SOLiD 5500, 12% Roche 454 GS Junior, 5% Pacific Biosciences RS, 3% Illumina HiScan SQ, 3% ABI SOLiD 4Hq, 1% Intelligent BioSystems/Azco MAX-Seq and 1% Intelligent BioSystems/Azco Mini-Seq (1%) (Figure 2).

Most survey respondents either had access to one sequencing unit in their lab (34%), or they used the sequencers at a collaborator or service facility and did not know the exact number of

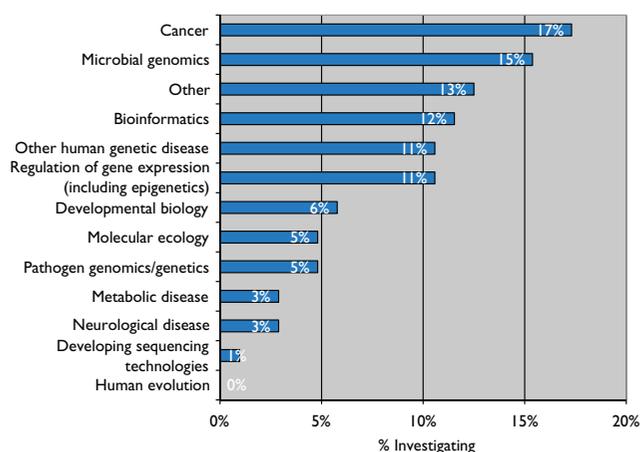
**Figure 3: Sequencing units available to survey respondents in their own lab/sequencing facility**

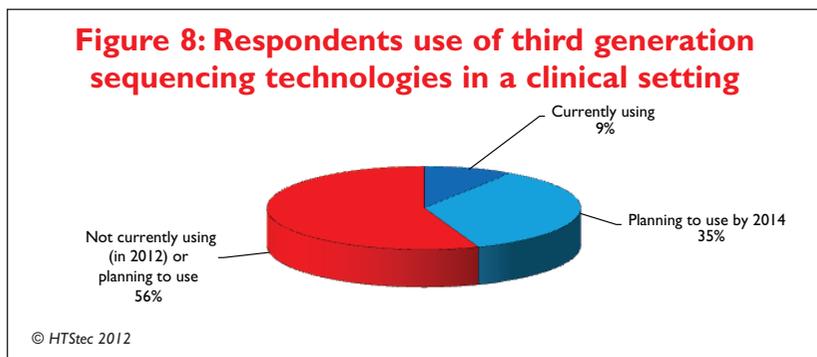
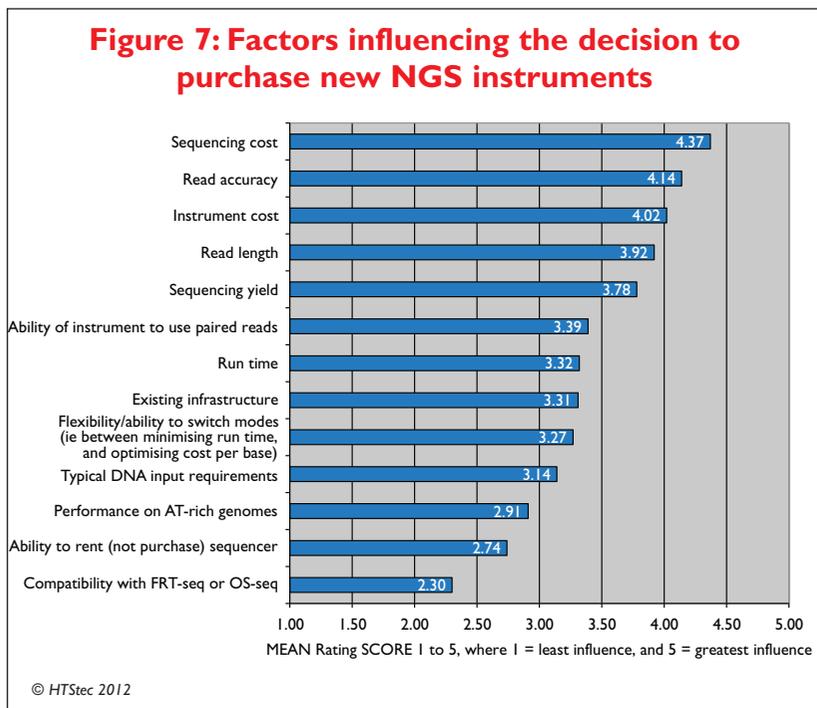
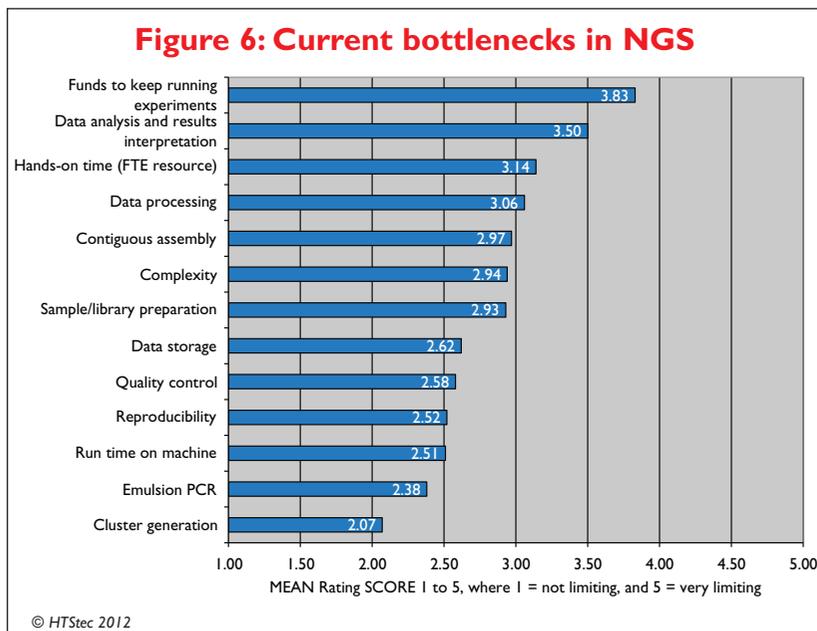


**Figure 4: Next-generation sequencing applications most run today**



**Figure 5: Primary focus of survey respondents NGS investigations**





units present (35%). Smaller proportions of respondents were able to access two sequencing units (13%), three units (7%), four units (3%), five units (1%), 6 to 10 units (1%), 15-20 units (1%), 25-50 units (1%) and >50 units (1%). The mean number of units per lab was 3.7 (Figure 3).

**NGS applications**

The three most widely investigated NGS applications today (2012) were targeted resequencing, mRNA-seq and whole transcriptome sequencing (49%, 47% and 43% of respondents ran these applications respectively). The least-run NGS application today (2012) was CLIP-seq (run by only 8% of respondents) (Figure 4).

Cancer was the most common primary focus of for survey respondents' NGS investigations. However, no single primary focus constituted more than one-fifth of all responses (Figure 5).

**NGS bottlenecks**

Funds to keep running experiments was rated by survey respondents the most limiting NGS bottleneck. This was followed by data analysis and results interpretation, and then hands-on time (FTE resource) (Figure 6).

**NGS purchasing factors**

Sequencing cost was rated by respondents the most important factor determining NGS instrument purchase. This was closely followed by read accuracy, instrument cost, read length and then sequencing yield (Figure 7).

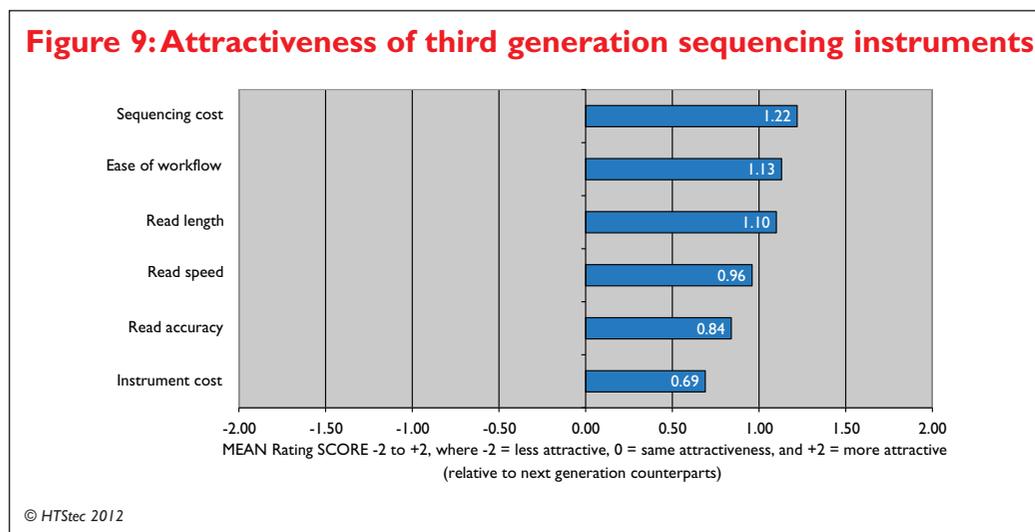
**'Third generation' sequencing**

Only 9% of survey respondents are currently using third generation sequencing in a clinical setting, and 35% intend to do so by 2014 (Figure 8).

Sequencing cost was rated by respondents the most attractive attribute of third generation sequencers relative to their next-generation counterparts. This was followed by ease of workflow and read length. For all attributes rated, the expectation is that third generation sequencers will be more attractive than their next-generation counterparts (Figure 9).

High raw error rate was ranked by survey respondents as the most serious potential issue affecting data quality in third generation sequencers. This was followed by difficulty in sequencing long homopolymeric regions and reductions in average sub-read lengths in comparison with quoted average read lengths. Coverage bias when sequencing AT-rich genomes was ranked the least serious potential issue (Figure 10).

**Figure 9: Attractiveness of third generation sequencing instruments**

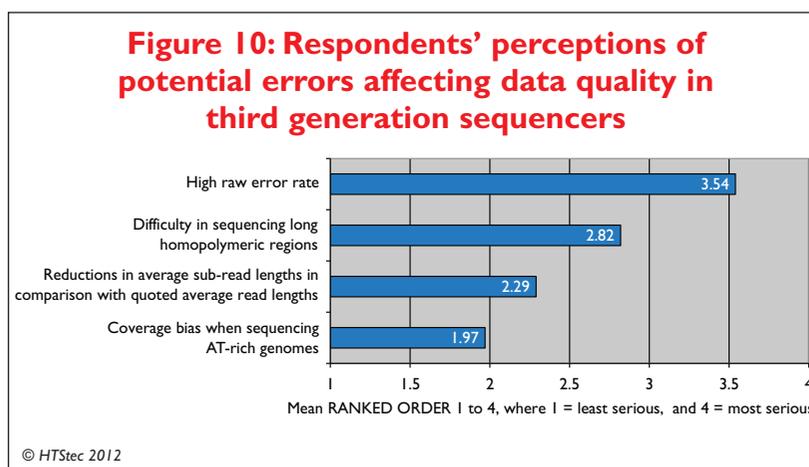


Read accuracy (irrespective of definition) was rated as the aspect of third generation sequencing respondents would most like to see improved. This was closely followed by sequencing cost, IT support and data handling, read length and then coverage. Run time was the aspect of third generation sequencing where respondents were least seeking improvement (Figure 11).

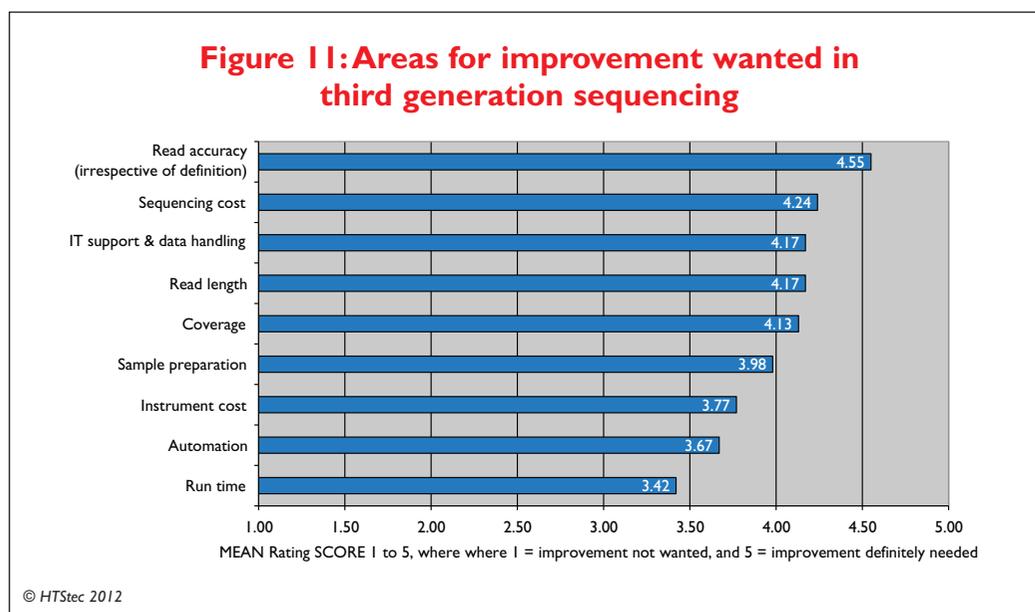
**Current status of the main sequencing platforms**

**Illumina** ([www.illumina.com](http://www.illumina.com)) continues to drive innovation across its entire sequencing ecosystem – from sample preparation to system enhancements to data analysis. Recently, the company announced

**Figure 10: Respondents' perceptions of potential errors affecting data quality in third generation sequencers**



**Figure 11: Areas for improvement wanted in third generation sequencing**





**Figure 12:** The MiSeq from Illumina is a fully integrated personal sequencer



**Figure 13:** The HiSeq 2500 from Illumina is ready for any application, any sample size, today

chemistry enhancements to HiSeq® 2500, its next-generation system capable of sequencing an entire genome in approximately 24 hours. With support of paired 250 base pair read lengths in rapid run mode, HiSeq 2500 will be able to generate up to 300Gb in rapid mode with sample to data in less than three days. Illumina also introduced a novel library prep method and analysis algorithm, enabled by its recent acquisition of Moleculo, Inc. It will produce synthetic read lengths up to 10Kb at an extremely low error rate, allowing for more comprehensive coverage and accurate genotyping

of clinically significant genes and new applications. Interest in MiSeq®, Illumina's benchtop sequencer, continues to grow as the company shared an updated product roadmap with a path to 15Gb of output. The company's cloud-based data analysis, storage and sharing platform, BaseSpace, also continues to expand. BaseSpace Apps, an applications store for BaseSpace, has grown to more than 100 independent developers who are bringing a wide range of analysis tools to the user base. Beyond these core updates, Illumina recently announced it has acquired Verinata Health, a leading provider of non-invasive prenatal tests for the early identification of foetal chromosomal abnormalities. This acquisition builds on Illumina's acquisition in 2012 of BlueGnome, a leader in cytogenetics and *in vitro* fertilisation, establishing the company as a leader in genomic-based diagnostics and advancing reproductive health (Figures 12 and 13).



**Figure 14:** The Ion PGM™ and Ion Proton™ Systems from Ion Torrent (Lifetech) deliver fast, simple, scalable next-generation sequencing that any lab can afford

**Ion Torrent** ([www.iontorrent.com](http://www.iontorrent.com)) has pioneered an entirely new approach to sequencing using semiconductor technology and simple chemistry. Ion Torrent sequencers deliver faster, simpler sequencing at a fraction of the cost of other sequencers on the market, making sequencing more accessible for all labs. Ion Torrent has recently released 400 base sequencing on the Ion PGM™ System for improved assembly of *de novo* microbial sequencing, making it the only benchtop sequencer to offer long read sequencing as a cost-effective option for routine use. The highly multiplexed Ion AmpliSeq™ Targeted Selection Technology, which offers targeted human DNA

sequencing with only 10ng of input DNA, has now expanded beyond the 2012 releases of fixed panels for cancer and inherited disease and the ability to design custom panels. The most recent Ion AmpliSeq™ technology launches in 2013 expand the DNA offering to community-designed DNA panels, including designs for colon/lung cancer and BRCA1/2, along with the ability to design custom mouse panels. The launch of the Ion AmpliSeq™ technology for targeted RNA sequencing includes panels for cancer and apoptosis, along with the ability to design custom human RNA panels. The Ion Proton™ System is capable of sequencing 1-2 exomes or 1-4 transcriptomes on the current Ion PI™ Chip. The Ion PII™ Chip, due to begin shipping in mid 2013 will scale to provide human-scale genome sequencing in a single day for \$1,000. Ion Torrent offers a comprehensive range of sample preparation kits and automation and seamless software integration for the fastest path to biological results (Figure 14).

Oxford Nanopore Technologies® ([www.nanoporetech.com](http://www.nanoporetech.com)) is the leading developer of nanopore sensing technology for the analysis of DNA, RNA, proteins and other single molecules. Oxford Nanopore's new generation of real-time sequencing technology uses nanopores to deliver ultra long read length single molecule sequence data, at competitive accuracy. Oxford Nanopore intends to commercialise 'strand sequencing' for DNA analysis on its scalable electronic GridION platform and offer a miniaturised version of the technology, MinION, which will make nanopore sequencing universally accessible. Oxford Nanopore's GridION system consists of scalable instruments (nodes) used with consumable cartridges that contain proprietary array chips for multi-nanopore sensing. Each GridION node and cartridge is initially designed for real-time sequencing by 2,000 individual nanopores at any one time and will deliver tens of Gb of sequence data per 24-hour period. Alternative configurations with more processing cores (more than 8,000 nanopores) will become available later. Nodes may be clustered in a similar way to computing devices, allowing users to increase the number of nanopore experiments being conducted at any one time if a faster time-to-result is required. For example, a 20-node installation using an 8,000 nanopore configuration would be expected to deliver a complete human genome in 15 minutes. Pricing will be competitive with other leading systems at launch. Oxford Nanopore has also miniaturised these devices to develop the MinION; a disposable DNA sequencing device the



Figure 15: The Oxford Nanopore Technologies GridION™ system

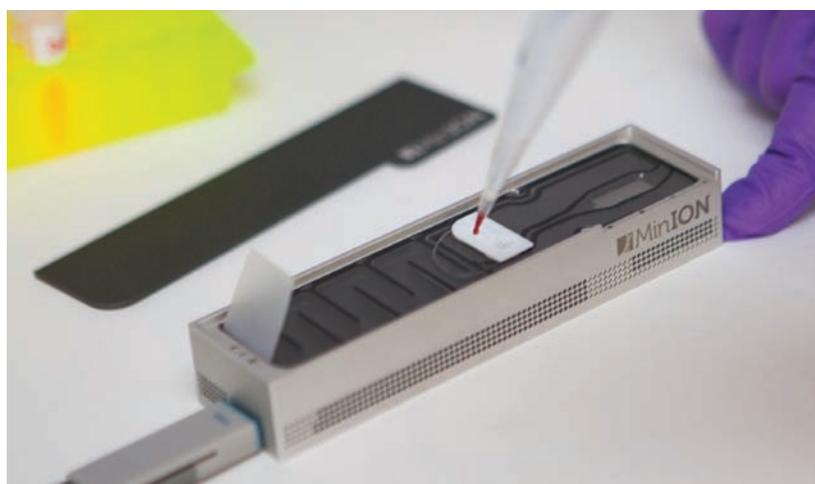


Figure 16: The Oxford Nanopore Technologies MinION™ device

size of a USB memory stick whose low cost, portability and ease of use are designed to make DNA sequencing universally accessible. A single MinION is expected to retail at less than \$900. Orders are not yet being taken for either system, however you may register your interest at its website. Oxford Nanopore has an intellectual property portfolio of more than 300 issued patents and



**Figure 17 (above)**  
PacBio SMRT cells and 8Pac  
format

**Figure 18 (right)**  
The PacBio® RS High  
Resolution Genetic Analyzer



patent applications in more than 80 patent families. The company is currently pursuing techniques for nanopore-based analysis using biological and solid-state nanopores, as well as hybrid versions of these, and also including a wide variety of adaptations and modifications (Figures 15 and 16).

**Pacific Biosciences** ([www.pacificbiosciences.com](http://www.pacificbiosciences.com)) has developed a third generation DNA sequencing system, the PacBio® RS High Resolution Genetic Analyzer, that incorporates novel, single molecule sequencing techniques and advanced real time analytics. PacBio calls this SMRT® (Single Molecule, Real-Time) technology. SMRT DNA sequencing is performed on SMRT Cells, each patterned with 150,000 zero mode waveguides or ZMWs. Each ZMW contains a single DNA polymerase, providing the window to observe DNA sequencing in

real-time. The PacBio RS system continuously monitors ZMWs in sets of 75,000 at a time. SMRT Cells are nanofabricated consumable substrates used in conjunction with the DNA Sequencing Kit for automated processing on the PacBio RS system. One SMRT Cell is consumed per sequencing reaction. SMRT Cells are packaged together in a streamlined 8Pac format. Experiments can be run on a single SMRT Cell or in batch mode to meet project needs. The instrument features high performance optics, automated liquid handling and an environmental control centre, all directed through an intuitive touchscreen interface. Also included is a state-of-the-art Blade Center, the computational brain responsible for primary data analysis. A comprehensive informatics suite completes the package. During the DNA sequencing process, the PacBio RS uses advanced collection optics to record light pulses emitted as a byproduct of nucleotide incorporation. These signals are delivered in real time to the primary analysis pipeline, housed entirely on the Blade Center. Proprietary algorithms translate each pulse into an A, C, G or T base call with its own set of quality metrics. As soon as the base call data is generated, it is available for secondary analysis through PacBio software or virtually any other secondary analysis pipeline. Long readlengths, intuitive operation and throughput flexibility combine to deliver the data faster than previously possible with high accuracy and the ability to detect real-time kinetic information. PacBio RS enables targeted sequencing to more comprehensively characterise genetic variations; *de novo* genome assembly to finish genomes in order to more fully identify, annotate and decipher genomic structures; and DNA base modification identification to help characterise epigenetic regulation and DNA damage (Figures 17 and 18).

At the AGBT Meeting held in Florida this February, **QIAGEN** ([www.qiagen.com](http://www.qiagen.com)) unveiled an innovative sample-to-result NGS workflow designed to enable the routine use of this breakthrough technology beyond life sciences research in areas such as clinical research and diagnostics. A key element of the workflow is GeneReader, a transformational NGS benchtop sequencer that offers many features essential for customers in clinical research and diagnostics to create routine laboratory processes. Unlike other platforms, which process only one flow cell at a time and often require sample pooling for cost-efficient runs, the sequencer has a turntable design that enables the continuous loading of up to 20 flow cells for independent and parallel sequencing. Individual patient

samples also can be handled cost-efficiently without the need for indexing or bar-coding, which means processing can occur at any time, and in any order, without delay or concerns about potential regulatory issues. QIAGEN is convinced that NGS is making a transformational impact on life science, but challenges are limiting more widespread adoption for clinical purposes. The development of the company's complete sample-to-result workflow is a key achievement in its initiative to offer a seamless integration of new NGS platforms with high-quality reagents, molecular testing content and services. QIAGEN is planning to begin placing NGS workflows with selected customer groups during 2013 and expects NGS to complement established molecular technologies, particularly real-time PCR. The adoption of NGS in clinical research and diagnostics has been hampered for various reasons, particularly workflow challenges that become more pronounced in clinical settings due to the increased number of samples being processed. Other challenges include manual sample preparation, delays caused by batching samples

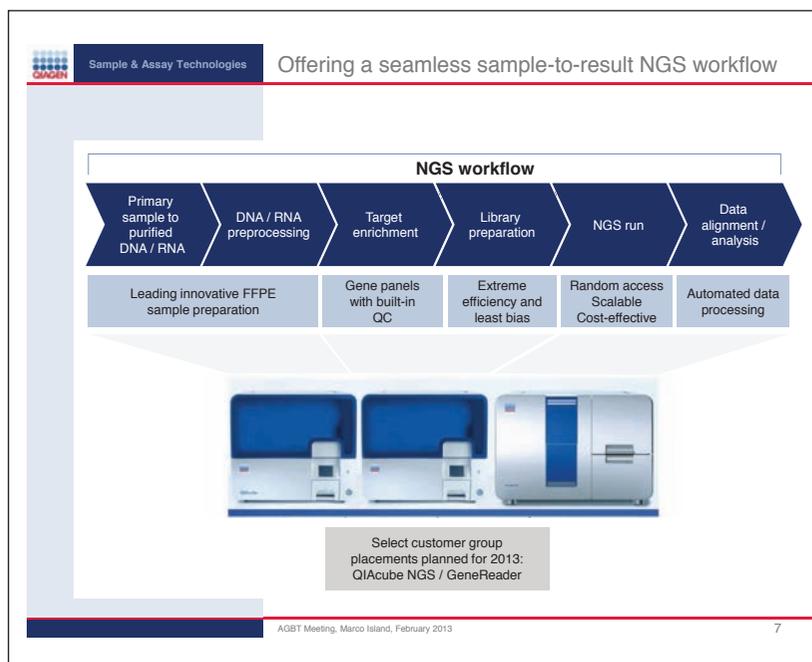
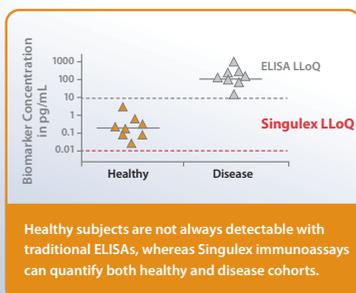


Figure 19: Complete and automated NGS workflow solution from QIAGEN, addressing clinical research and diagnostic settings



**Imagine** what you can accomplish by **quantifying** previously **undetectable biomarkers.**

Uncover clinically relevant biomarkers faster and with more confidence. Powered by Singulex digital technology, you can measure low-abundance biomarkers accurately and reliably, even in healthy individuals.

Want to see the data? It's waiting for you at [www.singulex.com/imagine](http://www.singulex.com/imagine)



**Figure 20 (right)**  
The Roche 454 GS FLX+ (left)  
and GS Junior Benchtop  
Systems (right)

**Figure 21 (below)**  
The Roche NimbleGen  
SeqCap EZ Exome Library for  
amplicon assays



to achieve cost-efficient runs, and the speed and quality of data analysis. QIAGEN's highly automated workflow addresses these challenges by offering an ecosystem of automated products and services from primary sample to digital result (Figure 19).

Roche ([www.454.com](http://www.454.com)) is investing in three main areas in its NGS portfolio: 1) significant performance and ease-of-use improvements to the 454 GS Junior Benchtop System and the GS FLX+ Systems; 2) an expanding menu of amplicon assays and SeqCap EZ target enrichment panels for translational research applications; and 3) innovative technologies that will enable the next leap in performance, cost and scalability for the future of sequencing. The GS Junior System brings the power of NGS directly to the laboratory benchtop. Upcoming improvements will extend read length beyond 400bp to deliver more Sanger-like reads. New automation solutions will reduce hands-on time, drive lab efficiency and deliver powerful benchtop sequencing. The GS FLX+ System offers the unique combination of high throughput and long Sanger-like reads, with lengths up to 1,000bp and beyond. New developments will enable extra-long read amplicon sequencing on the GS FLX+ System for targeted gene sequencing and 16S metagenomics applications. Roche's ready-to-run

amplicon assay menu is expanding to include new sequence-based assays for infectious disease (ie HIV drug resistance). These complement already available assays for leukaemia and human leukocyte antigen (HLA) typing research. In addition, Roche is investing in improvements to its portfolio of proven NimbleGen target enrichment products, including the SeqCap EZ Exome Library, SeqCap EZ Choice Library and SeqCap EZ Designs. Future sequencing technologies in development include semi-conductor-based sequencing system in collaboration with DNA Electronics. This development builds on the current 454 pyrosequencing-based chemistry and will enable seamless evolution from optical detection to inexpensive, highly scalable electrochemical detection. Also in development is a single molecule nanopore-based sequencer in collaboration with IBM and Arizona State University, which directly reads and decodes human DNA quickly and efficiently. This novel technology offers true single molecule sequencing by sequencing molecules of DNA as they are threaded through a nanometer-sized pore in a solid-state silicon chip (Figures 20 and 21).

### Discussion

NGS is an area renowned for its dynamism and the last two years have been no exception. Sequencers and their capabilities are in an almost constant

state of flux. Continual technological evolution is driving the transfer of sequencing into new environments, while allowing sequencing to become routine in many areas of traditional research.

Around two-thirds of sequencing activity is no longer carried out in the labs of survey respondents, indicating a shift away from conventional routes of access to sequencing technology. Respondents are able to access a multiplicity of sequencing platforms and purchasing decisions are based on multiple factors. These survey findings reflect how individual respondents match their own specific requirements to the unique characteristics of each instrument. This diversity is paralleled by the wide range of applications currently being employed to conduct research in a variety of areas in biology.

One noticeable trend is the shift to targeted resequencing, which has overtaken *de novo* sequencing as the most widely-investigated NGS application. In contrast to *de novo* sequencing, in which the genome sequence of an organism is discovered for the first time, targeted resequencing involves comparison with a previously generated reference sequence. This allows the variation within the sample sequence to be determined, enabling detection of genetic variants known to play a role in disease, a prerequisite for genomics-based diagnostics.

Sequencing vendors are directing their efforts towards improving system performance and extending read lengths. More chemistry kits are available to support target enrichment, an expanding menu of amplicon assays (Roche 454), targeted human DNA sequencing and targeted RNA sequencing (Ion Torrent). Furthermore, some vendors (Illumina) are developing novel IT platforms to expand the range of analytical tools available to users.

Many labs are currently operating under considerable financial constraints, with the economic downturn affecting all aspects of biomedical research. These unrelenting pressures are leading to demands for cheaper instruments and reduced running costs. As a result, there has been increased emphasis on improving aspects of sample prep and increasing ease of workflow in order to realise greater efficiency (QIAGEN).

The launch of desktop personal sequencers represents another response of the industry to the financial situation and aims to take NGS beyond the centralised facility to a wider diversity of smaller labs. The advent of third-generation sequencing instruments, able to reach rapid test times while maintaining precision and low costs, has the potential to greatly improve diagnosis of cancer and many inherited diseases.

Sequencing is set to become even more accessible, with single molecule real-time sequencing, semi-conductor technology and simpler chemistry already in use. Nanopore technology is expected to deliver the first fully miniaturised systems, as well as the next big leap in sequencing performance and cost. The launch of the Ion PIITM chip from Ion Torrent raises the possibility that the milestone figure of \$1,000 for human-scale genome sequence in a single day will become a reality.

These exciting advances suggest that the current trajectory of technological progress in DNA sequencing is likely to continue, extending the sequencing revolution from the lab into the clinic. **DDW**

## References

1 Next Generation Sequencing Trends 2012 Report, published by HTStec Limited, Cambridge, UK, December 2012.

---

*Andrew Szopa-Comley prepared the NGS Trends 2012 market report discussed in this article while gaining work experience as a market analyst at HTStec. A recent graduate in Natural Sciences from Cambridge University, Andrew is hoping to pursue a research career in biological research and is particularly interested in how genomics can provide insights into questions in evolutionary biology. HTStec Limited is an independent market research consultancy whose focus is on assisting clients delivering novel enabling platform technologies (liquid handling, laboratory automation, detection instrumentation and assay reagent technologies) to drug discovery and the life sciences. Since its formation nine years ago, HTStec has published more than 80 market reports on enabling technologies and more than 40 review articles in Drug Discovery World. Please contact [info@htstec.com](mailto:info@htstec.com) for more information about HTStec reports.*