Bridging genomics technology and biology.

Branco, MR

For additional information about this publication click this link.
http://qmro.qmul.ac.uk/xmlui/handle/123456789/17895

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk

Genome **Biology**

MEETING REPORT

Bridging genomics technology and biology

Miguel R Branco[1,2,3]

Abstract

A report on the UK Genome Science Meeting, held at the University of Nottingham, UK, 2–4 September 2013.

This year's newly named UK Genome Science Meeting was the fourth edition of what was previously known as the UK Next Generation Sequencing Meeting. The re-naming reflects technological developments that continue to redefine the meaning of next generation sequencing, and the fact that high-throughput sequencing is now a common tool in the scientific community. Indeed, an enormous diversity of topics was presented at this compact 3-day meeting, ranging from evolving technologies and bioinformatics, through to evolutionary genomics, metagenomics and clinical applications, amongst others. The meeting succeeded in portraying how, in relatively few years, new sequencing technologies have revolutionized the way we do basic research in just about every subject area, and are quickly making their way through translational research, and into the clinic and field.

Pushing the technological boundaries

At the base of current genomics research lie the powerful technologies that continuously aim to deliver more data at lower cost. The 'sequencing boom' has driven the creation of an assortment of new companies, as well as heavy investment from existing companies, to provide sequencing solutions, library preparation reagents and bioinformatics support. This was evidenced by the 20 or so companies represented at the meeting.

Scott Brouilette (Illumina) introduced a clever genome phasing solution based on work from Stephen Quake's lab (Howard Hughes Medical Institute, Stanford University, USA) that relies on sequencing of large DNA fragments (6 to 8 kb) aliquoted at limiting dilution. This allows for segregation of haplotypes and resolution of cases of compound heterozygosity, as well as improved assembly of repetitive regions.

Clive Brown (Oxford Nanopore, UK) provided a much-anticipated update that highlighted the technical hurdles that Oxford Nanopore has had to overcome to produce nanopore arrays that not only work but are also shippable. Oxford Nanopore has now put more emphasis on the release of the portable MinION, for which field-testing is ongoing, while it continues to develop the modular GridION system. The technology can deliver multi-kilobase reads, and can process 50 kb reads with consistent data quality throughout. Much of the interest in nanopores goes beyond their potential as DNA sequencers and, according to Brown, any given nanopore type can distinguish about one-third of all known DNA modifications, such that a combination of pores could potentially resolve them all. The future may also see RNA and protein sequencing through nanopores, but for now we await the arrival of the first commercial nanopore DNA sequencer.

One technology that can already deliver robust detection of DNA modifications, including 6-methyladenine (6mA), is the PacBio system - Paul Coupland (Wellcome Trust Sanger Institute, UK) shared his experience in using it in a variety of contexts. The system has undergone significant upgrades at several levels and Coupland reported sequencing, on average, 2 to 3 kb library inserts (although the polymerase can deliver longer reads). Apart from 6mA detection, some of the most interesting applications of the technology that were touched on were the sequencing of single-cell cDNA, where whole transcripts are visible and isoforms resolved easily, and the direct sequencing of viral and bacterial genomes without the need for library preparation.

As genomics research moves into diagnostics, targeted technologies that can deliver in-the-field results quickly and cheaply are necessary. Jonathan O'Halloran (QuantuMDx Group, UK) presented an impressive hand-held device (Q-POC) that performs accurate multiple diagnostics in

Correspondence: m.branco@qmul.ac.uk
[1]Centre for Trophoblast Research, University of Cambridge, Cambridge, UK
[2]Epigenetics Programme, Babraham Institute, Cambridge, UK
Full list of author information is available at the end of the article

less than 20 minutes from the point of sample collection (and aiming to get under 10 minutes). The device incorporates a microfluidics chamber for rapid DNA amplification and a nanowire-based hybridization system for multiplexed detection of up to 100 pathogens or genetic variants. The company is hoping the technology will change medical diagnostics in developing countries where the current need for expensive lab-based analyses is hampering efficient treatment.

Managing variability

While generating sequencing data is now a routine undertaking, extracting meaningful results often remains a challenge. Bioinformatic analyses are not only vital for dissecting out key results but can also feedback useful information for experimental planning. Indeed, Geoff Barton (University of Dundee, UK) and his team (talks also by Marek Gierlinski and Nick Schurch, University of Dundee, UK) performed careful analysis of RNA-seq data to answer a basic question in differential expression experiments: how many replicates are enough? In collaboration with Mark Blaxter (University of Edinburgh, UK), the same RNA-seq experiment was performed 48 times and several commonly used algorithms for differential expression analysis were tested to gauge how variability affects the calling of differentially expressed genes. While variability is bound to be different between experimental set-ups, the study nonetheless establishes a baseline of expected performance from replicate data (for example, using edgeR and a twofold change cut-off yielded an 80% true positive rate). Gierlinski pointed out that badly correlated (or outlier) replicates are present even in such a technically uniform experiment as this one, and that inclusion of these outliers invalidates the basic assumptions of differential expression algorithms. It was estimated that with three replicates there is a 24% chance of having one badly correlated replicate. One way that technically poor replicates can be detected is by using spike-in control RNAs, as was highlighted by Schurch. These can also be used as powerful data normalization controls, especially in cases where a significant portion of genes is differentially expressed in one preferred direction.

If variability in a standard RNA-seq experiment already requires substantial attention, managing it in single-cell transcriptomes is crucial. John Marioni (EMBL-EBI, UK) and Xiuwei Zhang (EMBL-EBI and Wellcome Trust Sanger Institute, UK) presented their solutions to the problem, which again are based on the inclusion of spike-in RNA controls. Marioni described how such spike-in data can be used to model technical variability, and therefore ensure that true inter-cellular variability is being detected. Zhang demonstrated how the carefully designed External RNA Control Consortium

(ERCC) sequences can impart information not only on inter-cellular technical variability, but also on intracellular technical variability, allowing one to get a firm handle on the main problem of single-cell RNA-seq and serving as a metric for the evaluation of potential protocol developments.

From lab bench to clinic

Genomics research has amply demonstrated its invaluable use in translational research, mainly by identifying variants associated with phenotypic traits and genes implicated in rare diseases. Ultimately, genome sequencing aims to provide personalized treatment solutions that are most effective for a given individual. Zamin Iqbal (Wellcome Trust Centre for Human Genetics, University of Oxford, UK) showed how sequencing of regional *Plasmodium falciparum* strains can help to design more effective drug treatments. In a striking example, Iqbal explained why a particular vaccine had failed in Africa - whilst the allele targeted by the drug was present in the reference strain, its frequency in African strains was only 16%. Prostate cancer treatment may also benefit from sequencing approaches, as Ian Sudbery (University of Oxford, UK) explained that patients that acquire resistance to androgen deprivation treatment (ADT) commonly show upregulation of genes involved in Wnt signaling. Using cell line models, Sudbery showed that inhibition of Wnt signaling may be a useful strategy for reversal of ADT resistance.

If these encouraging studies are representative examples of how genomics research may improve patient care, then the ambitious plans outlined by Mark Caulfield (Genomics England, UK) are going to constitute a landmark transformation in our approach to clinical diagnostics and treatment. Genomics England is a company that was recently set up by the Department of Health to sequence 100,000 genomes of rare diseases, common cancers and pathogens in NHS patients. Among common cancers, expert panels have carefully chosen to focus on childhood cancers and lung cancer; similarly, HIV, hepatitis C and tuberculosis will be the primarily targeted pathogens. Caulfield described the plans of the company for patient sample collection, sequencing and data management, a well thought-through process that has established solid ground from which the project will develop. The first pilot study, planned for 2014, will aim to sequence 2,000 genomes from patients with rare disease at 30× coverage. Beyond the primary objectives of encouraging new drug discovery and providing faster diagnostics, Genomics England hopes to drive down the cost of sequencing, and intends to train the wider healthcare community to use the technology, enabling the NHS to embrace genomic medicine at an unparalleled scale. Caulfield is under no illusion that this

monumental effort may be seen as trying to walk before
you can crawl in the minds of many, but remains
confident that such leaps in human endeavor have proven
in the long term to be world-changing.

Competing interests
The author declares that he has no competing interests.

Author details
[1]Centre for Trophoblast Research, University of Cambridge, Cambridge, UK.
[2]Epigenetics Programme, Babraham Institute, Cambridge, UK. [3]Current
address: The Blizard Institute, Barts and The London School of Medicine and
Dentistry, Queen Mary University of London, London, UK.