Genome **Biology**

MEETING REPORT

Clinical applications of sequencing take center stage

Gustavo Glusman*

Abstract

A report on the Advances in Genome Biology and Technology (AGBT) meeting, Marco Island, Florida, USA, February 20-23, 2013.

This year's Advances in Genome Biology and Technology (AGBT) meeting reflected the current state of 'next generation' sequencing (NGS) technologies: significantly reduced competition and innovation, and a strong focus on standardization and application. Announcements of technological breakthroughs - a hallmark of previous AGBT meetings - were markedly absent, but existing technologies continued to improve following the now expected exponential curve. Although applications ranged widely, there was a strong emphasis on clinical diagnosis.

Status of sequencing technologies

With the HiSeq and the MiSeq platforms, Illumina is now firmly established as the most widely used sequencing technology. Geoff Smith (Illumina, USA) announced the recent acquisition of Moleculo, a new protocol extending contiguous sequence length. Jonathan Rothberg (Life Technologies, USA) reported on the 100-fold increases in sequencing output in just two years on the Ion Torrent and Proton platforms, stating they are 'not seeing the end of capacity' for the semiconductor-based technology. In addition, Jonas Korlach (Pacific Biosciences, USA) presented impressive results on particularly difficult sequences and showed that pathogen genomes could be quickly finished using the SMRT platform's long reads. On the other hand, James Knight (Roche, USA) discussed the detailed sequencing of RP11 - the individual who contributed 70% of the reference genome - closing many reference gaps and leading to the identification of reference errors.

*Correspondence: Gustavo@SystemsBiology.org
Institute for Systems Biology, 401 Terry Ave N., Seattle, WA 98109, USA

The vast accumulation of genetic data, the use of combinations of sequencing technologies and a variety of algorithmic improvements have led to the identification of certain limitations and recurring error types of current technologies. Mark DePristo (Broad Institute, USA) described the observation of 'fake SNPs' that arise from misassembling reads around short gaps, and discussed how these were resolved using the HaplotypeCaller software. He also emphasized that PCR amplification in the library preparation step can be a major source of error. Meanwhile, Jay Shendure (University of Washington, USA) described an updated version of molecular inversion probes that is highly multiplexed and rebalanced for solving problems caused by extremes in GC content; he also reported finding many *de novo* mutations that were missed by exome sequencing. In addition, I reported observing thousands of genomic segments enriched in false-positive findings, based on analysis of hundreds of complete genomes (Gustavo Glusman, Institute for Systems Biology, USA) .

Exome sequencing is prevalent: more than 100,000 exomes have now been reported. Several presenters discussed limitations in exome sequencing, and contrasted results with whole genome sequencing (WGS). Stephen Scherer (Hospital for Sick Children, Canada) showed that approximately 13% of exons are absent from exome capture arrays, and that WGS data yield more uniform coverage of the exome. He further reported that 6 out of 56 clinically relevant variants observed in a WGS study of autism would not have been captured by exome sequencing. Interestingly, Malachi Griffith (Washington University School of Medicine, USA) showed examples of short intronic deletions, which are invisible to exome analysis, and demonstrated how these can affect gene expression levels. Michael Talkowski (Harvard Medical School, USA) stressed the difficulty in identifying copy-number variants and aneuploidy from exome data. Overall, it seems there is a clear trend for WGS becoming the ultimate diagnostic tool.

Application to the clinic

The central theme of this year's AGBT was the application of sequencing technologies to clinical needs. Recurring

topics of discussion included the need for speed in processing pipelines (from samples to interpretation), validation of results and clinical reporting, particularly for incidental findings.

How can results from NGS be validated?

Zivana Tezak (Food and Drug Administration, USA) gave an overview of regulatory considerations and questioned whether NGS technologies have the required quality for clinical use, if Sanger sequencing - the currently established standard - is still required for validation. Tezak discussed decoupling instrument validation from clinical test validation, and the need to define the minimal set of markers tested and percentage of genome covered for an instrument to be deemed reliable for clinical use. However, the question remains: to what extent can an instrument be deemed 'validated' given that failure modes are strongly sequence-specific?

How should NGS results be reported?

Elizabeth Worthey (Medical College of Wisconsin, USA) and several others cited the reporting recommendations of the American College of Medical Genetics. Worthey discussed the difficulty of establishing causality for each patient ('this variant/in this gene/causes this disease/in this case') and described WGS analysis yielding 100 to 120 variants flagged for in-depth review, only approximately six of which were considered clinically reportable. Worthey also highlighted the danger of 'wrong annotation creep' in databases and conference reports. Jonathan Berg (University of North Carolina at Chapel Hill, USA) described 'binning' the genome into three or four actionability levels and a reportability score combining disease severity, likelihood of causality, effectiveness and acceptability of interventions, and available knowledge. The exact cutoff used for reporting findings would then be left to patients' personal preferences.

An additional recurring topic was the need to end the 'diagnostic odyssey': the grueling, painful, expensive and sometimes decades-long journey from negative test to negative test, failing to diagnose a rare disease. Christine Eng (Baylor College of Medicine, USA) referred to the Undiagnosed Diseases Program at the National Institutes of Health (NIH) - soon to offer support for extramural research - and described an exome sequencing pipeline currently processing 140 samples a month, largely pediatric and neurologic in nature, with a conservative diagnosis rate of 25%. When asked how frequently diagnosis affected healthcare, Eng stressed that the focus was on reaching a diagnosis and ending the odyssey. Stephen Kingsmore (Children's Mercy Hospital, USA) described STAT-Seq, a WGS program to deliver (within 50 hours) a provisional report to the ordering neonatologist, including indications for pharmacogenomic dose adjustment. Kingsmore stressed the psychosocial benefits of rapid and definitive diagnosis, even in the absence of a cure.

Several advances were described using NGS to tackle cancer. Rebecca Leary (Johns Hopkins Kimmel Cancer Center, USA) described Personalized Analysis of Rearrangement Ends (PARE), a method for tracking cancer progression in a patient. Most promising, she described the ability to perform 'digital karyotyping' by sequencing circulating DNA in a patient's plasma - a sensitive method for early detection. Olivier Elemento (Weill Cornell Medical College, USA) performed deep sequencing of VDJ junctions and phylogenetic analysis to elucidate the personalized history of lymphoma - from primary tumor to relapse. Ira Hall (University of Virginia, USA) reanalyzed data from The Cancer Genome Atlas to identify chromosomal rearrangements, mapping thousands of somatic breakpoints to base resolution. This revealed non-random breakpoint clustering: Hall described three main modes of chromothripsis ('chromosome shattering'): focused, diffuse and multifocal, and reported significant enrichment of chromothripsis in glioblastoma.

Sequencing technologies are now being applied 'at all stages of life', not just for research and clinical diagnosis. Kevin Hrusovsky (Perkin Elmer, USA) described 'consumer genomics' uses of sequencing for dating, parental testing, neonatal diagnosis and 'no phenotype' personalized genomics. Dagan Wells (University of Oxford, UK) reported improved *in vitro* fertilization results by using multiplexed, low-coverage WGS on an Ion Torrent PGM to detect aneuploidies in single cells - a feat performed in just 12.5 hours, for a mere $70 per sample. Sunney Xie (Harvard University, USA) described a highly detailed preimplantation genetic diagnosis procedure that involves sequencing the first and second meiotic polar bodies, to deduce from them the genome of the female pronucleus.

Other applications

Beyond the clinic, a dizzying diversity of applications was presented. Kjersti Aagaard (Baylor College of Medicine, USA) described revelations of detailed metagenomic sequencing applied to obstetrics: her team found that the composition of the vaginal microbiome shifts during pregnancy, and that the placenta has a nonpathogenic commensal microbiome most similar in composition to that of subgingival plaques. Ross Hardison (Pennsylvania State University, USA) reported enrichment of gene-regulating variants outside coding regions, and modeling gene expression based on integration of epigenetic features, multi-species alignments and transcription factor motif enrichment. Eric Schadt (Pacific Biosciences, USA) highlighted massive copy number variation impact on gene expression levels. Leonid Moroz (University of Florida, USA) studied memory and epigenetic modifications

in giant snail neurons, and found massive DNA demethylation in quick response to neurotransmitters. He reported being able to perform sequencing on an expedition ship, simplifying oceanic exploration. Leonard Lipovich (Wayne State University, USA) reported thousands of novel long non-coding RNAs. Chia-Lin Wei (Joint Genome Institute, USA) presented a detailed and riveting connectivity map of chromatin interactions between promoters and enhancers, within and between chromosomes. Finally, Mark Yandell (University of Utah, USA) discussed applying 'pigeonomics' to the same pigeon species that Darwin studied.

Tools and data access

As attention shifts from sequencing to extracting meaning from the data, there is a proliferation of commercial analytical tools - frequently cloud-based solutions - from companies such as Ingenuity Systems, Agilent, Maverix Biomics, DNAnexus and Omicia.

There was a modest crop of novel algorithms for computational analysis of sequence data. Andrew Farrell (Boston College, USA) presented RUFUS, an ingenious algorithm for detecting differences between two samples that requires no reference sequence. RUFUS promises to have many applications, from cancer to RNA editing. Aaron Quinlan (University of Virginia, USA) reported on the development of LUMPY, a probabilistic framework integrating diverse signals for structural variant discovery.

Conclusions

Despite the wealth of data described at the meeting, one important issue remains unresolved: though vast numbers of exomes and whole genomes have been obtained, and some progress was reported in databasing and sharing of medical information and treatment outcomes, much work remains to be done to make these data available to the research community.