

MEETING REPORT

Looking beyond the post-genomic era

Ester Falconer

Abstract

A report on BioMed Central's fourth annual Beyond the Genome conference held at the University of California, San Francisco Mission Bay Conference Center, USA, 1–3 October 2013.

Keywords: Single-cell genomics, Structural variation, Genome assembly, Bioinformatics, Information technology, Precision medicine

Meeting Report

I arrived at this year's Beyond the Genome (BTG) meeting in San Francisco with a particular interest in the single-cell genomic analysis session, as it is the basis of my own research. I was honored to have the opportunity to present a novel template-strand sequencing technique (Strand-seq) for finding genomic instability events such as sister chromatid exchanges, translocations and aneuploidy, adding to the genomic profiling possibilities in single cells. I was fortunate that this meeting also presented the opportunity to discover how plant genomics is at the forefront of epigenomics, and is pushing sequencing technologies and bioinformatics to catch up with the complex genomes of crops. Equally interesting were the innovative bioinformatics and IT tools that help translate the A, T, C and G into a narrative of where we came from as a species, and to where we are headed in clinical research.

We are all meta-genomes (single-cell analysis)

It is very revealing that many of the speakers in the single-cell genomics session prefaced their talks by remarking that they no longer have to convince an audience of the advantages of single-cell genomic analysis. The intersection of lower sequencing costs and validated techniques has shifted single-cell genomics from concept to mainstream. However, there is plenty of room for improving these techniques. Keynote speaker Stephen

Quake (Stanford University, USA) showed that integrating a microfluidics platform allows precise control over the fluid dynamics, to maximize precision, reproducibility and efficiency of reactions. Session chair Nicholas Navin (MD Anderson Cancer Center, USA) reminded us that we are still at the beginnings of single-cell genomics, and predicted that there will be further complexity as epigenomic and spatial organization maps will fulfill complete single-cell genomic profiles in the future.

As pointed out by Sunney Xie (Harvard University, USA), each cell has a unique genome. Thus, one major focus of this session was to profile these unique genomes in individual primary tumor cells, metastases and circulating tumor cells to identify potentially targetable features responsible for the evolution of a cancer to acquire mobility, adaptability, resistance and response to therapy. A recurring idea was that copy number variation (CNV) may be a more accurate signature to follow cancer evolution than the accumulation of single nucleotide variants (SNVs). Xie reported that in individual lung cancer patients, while SNVs vary between individual circulating tumor cells (CTCs), CNV signatures are more consistent and match those from metastatic but not primary tumor cells. In addition, CNV signatures of CTCs between different patients can accurately cluster patients into lung cancer subtypes, matching their disease progression state. These data argue for CNVs as an important diagnostic tool, a hypothesis seconded by James Hicks (Cold Spring Harbor Laboratory, USA), who also argued that CNV breakpoints reveal a cancer cell's history and lineage. However, he reported CNV heterogeneity between CTCs of different prostate cancer patients as well as within individual patients, arguing for a more dynamic evolution and adaptability of cancer cells. Devon Lawson (University of California, San Francisco, USA) also demonstrated heterogeneity between individual cells isolated from metastatic breast cancer xenografts, using gene expression profiling.

The idea of having many genomes within one individual was demonstrated by a high proportion of *de novo* translocations and aneuploidy during early zygotic cell divisions in pre-implantation human embryos (Thierry

Correspondence: efalconer@bccrc.ca
Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, BC V5Z 1 L3, Canada

Voet, Sanger Institute, UK). Finding structural variations as well as SNV signatures for pre-implantation genetic screening of *in vitro* fertilized (IVF) embryos was also shown by Fei Gao (BGI-Shenzhen, China) and Fuchou Tang (Peking University, China). Tang showed that a complete genomic profile of polar bodies (which are byproducts of meiosis and do not contribute to embryonic or extra-embryonic tissues) can be used to infer the CNV profile, chromosomal abnormalities, and potentially SNVs and haplotypes of oocytes and pre-implantation embryo cells, allowing the healthiest oocytes and embryos to be selected for IVF and implantation.

Good crop, bad crop (plant genomics)

The favored plant model organism *Arabidopsis thaliana* is diploid, has a relatively small genome with little repetitive DNA, and several decades of genetic studies to link DNA sequence to epigenetic inheritance and phenotype. Not surprisingly, *Arabidopsis* was well represented in the plant genomics session. Rob Martienssen (Cold Spring Harbor Laboratory, USA) and Ryan Lister (University of Western Australia, Australia) both detailed the extraordinary small RNA-dependent pathways that lead to methylation and silencing of repetitive transposable elements (TEs). Lister presented evidence that TE silencing via these small RNA silencing pathways is upregulated in stem cell niches such as the meristem and columella, to protect the development and differentiation programs in stem and progenitor cells from being disrupted by TE de-repression. The role of companion cells that upregulate and transport small RNAs to stem cells in the root tip or to germ cells in pollen illustrated the complex pathways that can be decoded in plants.

Another major focus was the contribution of the epigenome and epigenetic inheritance to the genotype-phenotype equation. Robert Schmitz (Salk Institute, USA) uncovered spontaneous epialleles via methylome sequencing of multiple generations, which are not due to structural variations. However, Magnus Nordborg (Gregor Mendel Institute of Molecular Plant Biology, Austria) argued in his keynote address that 'genetics rules!', meaning that the effects of genetics far outweigh the effects of environmental and epigenetic variation. His study of *Arabidopsis* from natural populations found that associations between the flowering time phenotype and differential methylation patterns could also be explained by SNVs. He proposed that many of the epigenetic variations in these populations are the outcome of sequence variation.

But plant genomics is often not afforded the luxury of having an ideal diploid model organism such as *Arabidopsis*. The extremely complex polyploid, repetitive and heterozygous genomes of many crop plants led wheat to be lightheartedly labeled as the 'bad guy' of plant

genomics by session chair Mario Caccamo (The Genome Analysis Centre, UK). However, Jorge Dubcovsky (University of California, Davis, USA) argued that the potential to knock out every gene in wheat makes it ideal for functional genomics, and therefore, a 'good guy.' He outlined efforts to sequence genomes of 1,000 wheat mutants and make sequence and seeds publicly available to researchers, with no intellectual property restrictions. It was very clear, however, that the large 'beast'-like genomes of crop plants are pushing the limits of sequencing and bioinformatics technology. The current approach of building a high quality assembly from short sequencing reads is confounded by the reality of these complex genomes. As Caccamo stated, there must be an integrated effort to combine whole-genome shotgun (WGS) sequencing, RNA sequencing, long mate pairs and single molecule reads with genetic maps to achieve a high-quality genome assembly for wheat and other complex crops. The continual advancement of sequencing and bioinformatics technologies will undoubtedly boost this effort.

Some assembly required (bioinformatics)

Session chair Michael Schatz (Cold Spring Harbor Laboratory, USA) proposed that the formula for a good genome assembly is long reads, good coverage and high quality. Achieving this from the degraded DNA from fossils is a major challenge, as Janet Kelso (Max Planck Institute, Germany) discussed. She presented a novel technique to obtain high-quality, high-coverage assemblies from ancient genomes, and showed that the genomic recipe for what makes modern humans differ from Neanderthals is surprisingly small. Assembling a quality transcriptome from short RNA-seq reads is equally challenging, but Alicia Oshlack (Murdoch Children's Research Institute, Australia) presented a clever algorithm called CORSET that clusters RNA-seq read fragments based on overlapping sequence and expression level, helping distinguish paralogs and alternatively spliced genes.

There is also urgency for clever bioinformatics tools to wade through the mountains of data that both the average lab and large sequencing centers can generate from genomics studies. Jeremy Goecks (Emory University, USA) outlined the features of the intuitive and user-friendly Galaxy platform for high-throughput analysis. This web-based approach to computing nicely demonstrated the power of bringing the computing to the data, and not *vice versa*. But bioinformatic challenges of genomic studies go far beyond needing clever software. As Geoffrey Noer (Panasas, USA) and Chris Dagdigan (BioTeam, USA) both pointed out, the science of sequencing is generating data faster than most underlying IT infrastructures can handle. Data storage and file retrieval,

even for the average lab, will continue to be major issues as datasets become larger and need to be more integrated. There will need to be more forethought put into IT policy, infrastructure setup and administration, integrating physical and cloud-based resources prior to undertaking projects.

The final keynote of the meeting, delivered by David Haussler, was a powerful reminder that the current reality of 'silos' of genomic data isolated by outdated confidentiality policies is preventing the true potential of these rich genomic datasets, and is threatening to become a lost opportunity to learn from available data. He argued that we should act now to establish an open and shared genomic database resource that can be easily accessed by all researchers to increase the power of clinical diagnosis and treatment. The goal of precision medicine for each individual will only be possible if we learn from the collective data of a large population.

Conclusions

It was evident from this meeting that there is no such thing as a single genome. There is seemingly infinite genomic diversity between individual cells, and we are only now starting to connect this diversity to cell function and disease. Technological advancements will add more complexity, integrating epigenomics, spatial organization and genomic evolution into the complete profiles of cells. It is up to researchers and bioinformaticians to realize that these data will only be useful if they are interpretable, accessible and open. The right tools to analyze, store and access data will need to evolve in parallel with forward-looking policies and data management systems to realize the full potential of the post-genomic era.

Abbreviations

CNV: Copy number variation; CTC: Circulating tumor cell; IVF: *in vitro* fertilized; SNV: Single nucleotide variant; TE: Transposable element; WGS: Whole-genome shotgun.

Competing interests

The author declares that she has no competing interests.

Acknowledgements

I would like to thank Ashley D Sanders and Carolina Novoa for input and comments on the manuscript. I also extend my apologies to the researchers who also presented excellent work that I did not discuss here due to space considerations.

Published: 31 October 2013

doi:10.1186/gb4138

Cite this article as: Falconer: Looking beyond the post-genomic era. *Genome Biology* 2013 **14**:313.