

MEETING REPORT

Beyond the Genome: genomics research ten years after the human genome sequence

Amanda M Casto^{1*} and Clara Amid²

Abstract

A report on the meeting 'Beyond the Genome', Boston, USA, 11-13 October 2010.

Ten years ago, the first draft sequence of the human genome was released, ushering in the post-genomic era. Since then, the costs of sequencing have plummeted and many new sequencing technologies have been introduced. These conditions have allowed researchers to investigate a wide variety of genomes that may be important in human health and disease, including those of cancer cells and of the organisms that live on our skin and in our gut. Others have taken the opportunity to revisit lingering questions about the human genome itself, including those related to gene number, the importance of non-protein coding genes, and the identification of functional and selected variants. Investigators from both academia and industry interested in these questions and others gathered at the 'Beyond the Genome' conference in Boston recently to discuss the current state and future trajectory of genomics research. The meeting was organized by BioMed Central and *Genome Biology* to mark 10 years in science publishing.

In their keynote addresses, Steven Salzberg (University of Maryland, College Park, USA) reviewed the history of publications estimating the human gene count, while George Church (Harvard Medical School, Cambridge, USA) discussed how the rapidly shrinking costs of sequencing have facilitated the realization of the Personal Genome Project (PGP) [<http://www.personalgenomes.org>], an effort to obtain and interpret genome information by collecting medical information and full genome sequences for 100,000 individuals. This will include studies of interactions between genes and environmental

factors such as microbes and immune responses to allergens, viruses and toxins (Genome+Environment=Trait; GET). Currently all publicly available genomes from the PGP are integrated into GET-evidence [<http://evidence.personalgenomes.org>].

Sequencing cancer and complex disease genomes

The identification of genes with driver mutations critical in oncogenesis has been a central aim of cancer research, as these genes may represent new drug targets for cancer therapies. Alison Klein (Johns Hopkins School of Public Health, Baltimore, USA) is investigating susceptibility genes for pancreatic cancer, a particularly deadly neoplasm. The overall 5-year relative survival for this disease is less than 5% and most patients die within 6 to 8 months of diagnosis. A family history of pancreatic cancer can be found in about 10% of patients, while smoking, obesity and diabetes are some significant non-genetic risk factors. Traditional linkage methods and genome-wide association studies (GWAS) have revealed very few susceptibility loci, and the genetic basis of 90% of the familial clustering of pancreatic cancer remains unknown. However, recent cost reductions in sequencing enabled Klein and her group to identify *PALB2* (partner and localizer of BRCA2) as a candidate pancreatic cancer gene using exomic sequencing in patients with a family history of the disease.

The ability to routinely characterize individual tumor genomes may lead to the development of chemotherapeutic regimens tailored to individual cancer patients. Elaine Mardis (Washington University School of Medicine, St Louis, USA) described the Cancer Genome Initiative, which has produced data from 150 cancer genomes during the past year. She pointed out that whole-genome sequencing and the analysis of cancer genomes will help in understanding the evolution of cancer. However, as Steven Jones (British Columbia Cancer Agency, Vancouver, Canada) discussed, there are critical issues that need to be addressed if this technology is going to be routinely applied. These include development of software tools for a comprehensive analysis of the genome, the assessment of the right level of sampling to identify somatic changes

*Correspondence: morgan21@stanford.edu

¹Department of Genetics, Stanford University, Stanford, CA 94305, USA
Full list of author information is available at the end of the article

as well as the clarification over the number of false negatives. Jones and his group investigated the viability of high volume parallel sequencing to characterize a rare adenocarcinoma of the tongue before and after treatment with drugs that target growth-factor receptors. The study identified genes containing somatic protein-coding mutations and copy-number alterations in the tumor, and it also has the potential to identify genetic changes associated with drug resistance and to elucidate their connection to known cancer pathways.

How much of the genome is functional?

The true number of genes in the human genome is a question that has long engaged scientists. Researchers have estimated the human gene complement using many different approaches, including cross-species genome comparisons, extrapolations based on the number of CpG islands or expressed sequence tags in a particular genomic region and, ultimately, detailed analyses by automated and manual annotation. Nevertheless, 10 years after the first draft of the human genome sequence was released, this question still cannot be answered precisely, despite the advent of new technologies such as massively parallel sequencing of RNA (RNA-seq), which are aiding the discovery of new genes and alternative transcripts. Some of these new methods, including transcriptome reconstruction from RNA sequencing (RNA-seq) data, were described by Chad Nusbaum (Broad Institute, Cambridge, USA). Transcriptome reconstruction, however, is a difficult task currently. RNA-seq reads are short (75 bp); therefore, their assembly and alignment to the genome are difficult and can lead to ambiguous alignments. Also, since this technology allows direct sequencing of RNA fragments, it is tricky to know which reads came from which transcripts to distinguish the abundance of alternatively spliced variants.

Chris Ponting (University of Oxford, UK) addressed the importance of the human gene count. He and his group estimate that between 6.5% and 10% of the human genome is under functional constraint and that the human genome includes five to eight times more noncoding than coding bases. This ratio is only about two in *Drosophila melanogaster*, suggesting that the fraction of functional DNA in a genome may be a better predictor of organismal complexity than gene number. If this is the case, it calls for additional attention to be paid to the functional classification of genes. Work within the GENCODE (Encyclopedia of Genes and Gene Variants) consortium aimed at producing a human reference gene set has already moved in this direction by annotating alternative splice variants in rich detail and by estimating the functional potential of all transcripts. The current human gene count in the GENCODE set is 21,671 protein-coding genes

(described by Clara Amid, Wellcome Trust Sanger Institute, Hinxton, UK); a further 1,451 transcripts have been classified as large intergenic noncoding RNAs (lincRNAs), while another 8,529 have been annotated as processed transcripts, which will undergo additional sub-classification as the project proceeds.

Evolutionary insights from genomic analysis

There is not just one human genome. The nucleotide sequence varies from individual to individual, from population to population, and from continent to continent. Taking a quantitative view, Stephan Schuster (Pennsylvania State University, University Park, USA) reported that Europeans tend to differ from one another by about 3.3 million single-nucleotide polymorphisms (SNPs). However, he and his colleagues found that this number increases to 4 million SNPs when comparing two African Bushmen. In an effort to fully characterize the genetic diversity represented by African hunter-gatherer groups like the Bushmen, Schuster has begun a project to sequence several Bushmen individuals and compare these genomes to one from an African Bantu.

The genetic differences that distinguish individual Africans from each other and from non-Africans include functional variants. Sarah Tishkoff (University of Pennsylvania, Philadelphia, USA) has studied several African populations that demonstrate lactase persistence and found that the mutations that confer this phenotype in Africa are different from those found in Europe and the Middle East. She has also described a number of nonsynonymous mutations found only in Africans that interfere with the perception of bitter taste through their effects on *TAS2R38*, a taste-receptor gene. Polymorphic sites in the human genome that have phenotypic effects, like the SNPs in *TAS2R38* and those near the lactase gene, are of particular interest to researchers. One way to identify functional polymorphisms is to look for regions of the genome that appear to have been subjected to selection. This can be done by analyzing genomic data using various tests that detect unusual patterns of allele frequency variation at a given polymorphism or of haplotype structure surrounding the polymorphism. Shari Grossman (Broad Institute, Cambridge, USA) has developed a new technique called the Composite of Multiple Signals (CMS) method, which uses several such tests in concert to help distinguish selected variants from their linked neighbors. She is currently working on applying this approach to preliminary data from the 1,000 Genomes Project [<http://www.1000genomes.org>], which is expected to contain more information about rare variants in the human genome than previous SNP datasets.

Microbiomes in human and other environments

Humans play host to a huge number of microbes; whereas each of us is composed of about 10 trillion human cells, we also have about 100 trillion bacteria - the human microbiome - living inside us or on our skin. Human microbial communities change over time, and vary between individuals, body sites, and different pathologic and physiological states. The association of the human microbiome with various pathologies has recently become the subject of intense interest. Jun Wang (Beijing Genomics Institute, Shenzhen, China) described work showing that gut bacterial communities are strikingly different in individuals with ulcerative colitis or Crohn's disease relative to healthy controls. Julie Segre (National Human Genome Research Institute, National Institutes of Health, Bethesda, USA) reported similar findings for the skin microbiome in patients with eczema and other skin disorders. Segre's lab is also exploring whether these microbes play a causative role in disease traits, investigating the effects of skin bacteria on wound healing in diabetic mice. Recent work by Rob Knight (University of Colorado, Boulder, USA) has showed that microbes can influence the phenotype of their host. His group found that the gut flora of leptin-mutant mice is substantially different from that of normal mice and that the transfer of bacteria from mutant mice to control mice can result in weight gain. Working in conjunction with George Church, Knight is involved in an effort to characterize and document the microbiomes of the Personal Genome Project volunteers. Jennifer Wortman (University of Maryland School of Medicine, Baltimore, USA) spoke about an

initiative by the Human Microbiome Project to make human microbiome data like this publicly available through its web portal [<http://www.hmpdacc.org>].

Current genomics research is focused on both newly formulated inquires and longstanding questions about the composition, content and functionality of the human genome and the genomes of other organisms. The constant introduction of new technological innovations for collecting genomic information ensures that genomics research will continually be aided by new investigative techniques and driven by ever-increasing amounts of data. The future of this research is likely to include translational work leading to the further development of personalized and genetic medicine. At the same time, dramatic increases in the number of genomic sequences available for analysis are likely to add to our understanding of how genomes change over time and how they are used as the blueprints for organisms.

Acknowledgements

We thank Adam Frankish and If Barnes at the Wellcome Trust Sanger Institute for helpful feedback.

Author details

¹Department of Genetics, Stanford University, Stanford, CA 94305, USA.

²Wellcome Trust Sanger Institute, The Morgan Building, Genome Campus, Hinxton, Cambridge, CB10 1HH, UK.

Published: 30 November 2010

doi:10.1186/gb-2010-11-11-309

Cite this article as: Casto AM, Amid C: Beyond the Genome: genomics research ten years after the human genome sequence. *Genome Biology* 2010, 11:309.