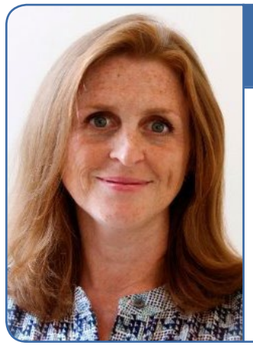
SEIZING THE OPPORTUNITY OF THE DECODED HUMAN GENOME



JANE WILKINSON, Senior Director, **Broad Institute**

Jane is a Senior Director at the Broad Institute where she leads the Broad Genomics Alliance Management team. Jane has over 20 years of high-throughput genomics experience from the Wellcome Trust Sanger Genome Center, UK where she was a key leader on the Human Genome Project and at Monsanto Company, USA where she led a new directive in plant genomics. Jane has been at the Broad Institute for fifteen years and has worked on various initiatives including Cancer, Mendelian, Infectious and Common Diseases.

Could you tell me a little about the Broad Genomics team and their mission?

The Broad Institute was founded to seize the opportunity that arose from the Human Genome Project -- the international effort that successfully deciphered the entire human genetic code. Despite that accomplishment, scientists knew they still lacked a clear understanding of the genetic basis of disease, and how to translate that understanding into more effective prevention, diagnosis, and treatment.

Since the Human Genome Project, the Broad Genomics Platform has played a leadership role in the design, data generation, and methods development in support of major genomic resource projects including: the HapMap, the 1000 Genomes Project, The Cancer Genome Atlas (TCGA), Comparative Reference Genomes, ENCODE, Genotype-tissue Expression Project (GTEx), Human Microbiome Project, Exome Sequencing Program Center for Mendelian Genomics, Human Cell Atlas (HCA) Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomAD) projects and All Of Us.

Over the past ten years, we have been one of the largest producers of human genomic information in the world. Currently, the group produces approximately 500 terabases of genomic data per month - a rate equivalent to a 30x human whole genome every 10 minutes. The group has processed more than 3 million samples from more than 1,400 groups in more than 50 countries.

Thinking about the challenges of genomics - what have been the key developments of the Broad Genomics within the past few years?

Over the past 20 years, the wave of genomic innovation and development has been incredible. The introduction of the first 'next-generation' sequencing machines (454 and Solexa) in 2005 and 2006 led the way to whole exome sequencing and the <$1000 whole genome. Applying lean manufacturing work design has allowed us to scale-up to handling over 10,000 samples per week - we completed our 100,000th whole human genome on the 15th anniversary of the completion of the Human Genome in 2018. And recently we have launched half a dozen new services, including Clinical Whole Genome Sequencing, Liquid Biopsy, and Single Cell Sequencing. We have also redesigned our Whole Exome which has resulted in an increase in capacity and more importantly - a significant drop in price.

You mentioned the new Whole Exome - can you tell me more how this was developed and the biggest impact it has made?

As sequencing costs continue to drop, the utility of whole genome sequencing for common disease research is beginning to be realized. However, even with cheaper whole genomes, complex common disease studies still often require larger case-control cohorts in the 10s of thousands and greater. Even with a $1000 genome, achieving the necessary power to find meaningful association is expensive. This realization has driven us to redevelop our germline whole exome product and workflow to reduce cost and simultaneously increase quality. The new Broad Custom Exome developed with Twist BioSciences leverages our custom hyb capture workflow, extremely even coverage, and a high on target percentage (~90%) to allow us to offer a high quality exome at ~USD$200. In addition to a high quality exome and lower price, we have re-engineered our automated workflow to enable a 3-fold increase in capacity in the past year, and are now capable of processing >300,000 exomes per year to better support the internal research needs of the Broad Community as well as external customers.

There's a lot of interest in Liquid Biopsy in the community - tell me more about how Broad Genomics has implemented this and how it's being used.

We have developed a custom unified target enrichment workflow to support a broad suite of somatic studies that require high sensitivity and specificity for low allele fraction variants. This method incorporates extremely stringent error correction using duplex UMIs and ultra-deep targeted sequencing. Using array-synthesized custom panels from Twist BioScience, we can now rapidly and in a cost effective manner target disease and study-specific genes to high depth. With Twist custom panels and our workflow, panel sizes can range from 10's of kilobases to multiple megabases in size and routinely achieve greater than 90% of sequenced bases on or near target, maximally utilizing gigabases of sequencing purchased. With cheaper cost of entry into a custom panel synthesis, investigators can now sequence only their regions of interest to extreme depth (20,000-50,000X raw coverage) to maximize the likelihood of detecting variants of importance at very low allele fractions. This allows the investigator to decide whether a larger fixed pan-cancer panel or a more targeted small panel is best suited to their study regardless of sample / patient number. This method and workflow is particularly well suited for blood biopsy cfDNA as well as bone marrow aspirate clonal hematopoiesis studies.

Just recently The Multiple Myeloma Research Foundation announced that it is launching the MMRF CureCloud™, a data hub that generates, aggregates and visualizes data to accelerate the delivery of precision medicine to multiple myeloma patients. Broad Gennomics will be providing our liquid biopsy capabilities as part of this initiative.

Connecting this large scale genomic data generation to translational analysis can be a pinch point for many researchers - can you tell me more about how Broad Genomics has tackled this?

We have tackled this issue in two specific ways. 1. Through the creation of a cloud-platform, Terra, we have enabled researchers to retrieve, analyse, and share their data in a scalable, accessible, and secure environment. We have further created repositories of open-source, best-practice methods in that platform so that researchers can use the same tools as we do to analyze these datasets. 2. For groups who do not have a fleet of bioinformaticians sitting around, we have created the Translational Analysis Group or TAG. TAG is comprised of 3 PhD-level computational scientists, an associate computational biologist, and a variant curation scientist. This group uses best-practice pipelines developed at Broad and beyond, to analyse our users data in a fee-for-service model. The group is fully cloud-native and all data and analytical outputs are made available through our cloud-platform. The team has run >15,000 analyses on samples ranging from whole genome sequencing for germline variant calling to somatic analysis of tumor variants from liquid biopsy. The team can validate analytical pipelines for use in clinical trials, clinical diagnostics, or therapeutic development.