

**CIFAR-GETX
CONFERENCE
BRAVE NEW WORLD:
UTILIZING GENETIC
NETWORKS TO
DECODE COMPLEX
DISEASE**

JUNE 4-5, 2018

IN PARTNERSHIP WITH

CIFAR



McLAUGHLIN
CENTRE



DAY ONE

KEYNOTE

GEORGE CHURCH, HARVARD UNIVERSITY - PERSONALIZING YOUR GENOME

- Human genome sequencing has provided us with an opportunity to understand which genetic variants affect disease risk and how genes affect an individual's response to drugs.
- Since 2005, the Personal Genome Project (PGP) has provided the world's only open-access information on human genomic, environmental and trait data, paving the way for precision medicine.
- In the past two decades, the cost to sequence a human genome has dropped from \$3 billion to roughly \$1000 and could drop to \$100 per genome in the next five years. This cost reduction is crucial to increasing participation in projects such as the PGP and in turn increasing data to enhance the predictive power of interpreting the human genome.
- There is still resistance by the public to get their genomes sequenced. Financially incentivizing the public to get their genome sequenced could be a mechanism to gain power from numbers.

SESSION 1: COMPLEX FACTORS THAT INFLUENCE GENOTYPE TO PHENOTYPE

MODERATOR

Hannah Carter, CIFAR Azrieli Global Scholar, University of San Diego

PRESENTERS

David Kelley, Calico – Interpreting Function from Sequence Using Deep Learning

Trey Ideker, University of California, San Diego – Using Networks to Translate Genotype to Phenotype

Jasper Rine, University of California, Berkeley – Space: The Final Frontier of Gene Silencing

- Many genetic variants found to contribute to complex disease are found in non-coding regulatory regions of the genome. The advent of artificial intelligence (AI) and machine learning has provided a powerful tool to analyze large data sets and predict the function of previously unknown regulatory elements.
- In dissecting regulatory sequences at a local level, better models and increased volumes of data are supporting rapid progress and informing a better understanding of how those sequences will behave in the next 10 years. In 10 years, we should also be able to map all of network biology to automatically construct a model of the cell.
- Translating a patient's genotype to their phenotype continues to be a challenge. In place of attempting to link the many variations in the genome directly to patient phenotype, using genetic networks as the architecture that informs a neural network can shed light on the genotype to phenotype connection. This approach Oxgrounds predictive modeling in real cell biology, greatly enhancing predictive power in healthcare and elsewhere.

- It is unclear whether machine learning algorithms are smart enough to recapitulate all cell biology, even given enough data. There is concern about what the models being developed are dependent on, such as the nature of data produced by experimentalists. An understanding of real biology is therefore still needed in the development and analysis of the outputs of these models.
- In addition to DNA sequence, we must also consider the local environment in the regulation of gene expression. The epigenetic changes can produce another dimension of complexity in gene expression and regulation and must be considered when designing therapeutics that leverage gene silencing.

SESSION 2: GENETIC INTERACTIONS, PENETRANCE AND HUMAN DISEASE

MODERATOR

Stephen Scherer, CIFAR Senior Fellow, McLaughlin Centre for Molecular Medicine, University of Toronto

PRESENTERS

Adam Shlien, Hospital for Sick Children – Finding Patterns of Mutation in Childhood Cancers

Jason Moffat, CIFAR Senior Fellow, University of Toronto – Identifying Essential Genes by Mapping Genetic Interactions

Shamil Sunyaev, Harvard Medical School – Genetic Variation and Penetrance

- Somatic mutations have the potential to drive cancer relapse. Novel tools and methods such as predictive modeling are being developed to trace somatic mutation evolution and trajectory throughout the life of a tumor, enabling doctors to get ahead of tumor progression.
- Studying genetic interactions has the power to drive understanding of genes with currently unknown function. Examining the effect of combinations of mutations in multiple genes will allow for an understanding of how genes combine and drive common and complex disease. CRISPR will enable a systematic effort to map genetic interactions at an unprecedented rate, allowing for new annotations of gene function and ordering genes in cellular pathways for the first time.
- Studies often use genetic variation to find SNPs that co-occur with disease symptoms and are statistically more likely to be associated with a disease. There is caution in using terms such as “co-occurrence” or “association” with respect to these potential disease genes, largely due to the factor of penetrance. Penetrance should be taken into consideration when designing methods to predict the functional effect of genetic variants and determining which variants are important in disease.
- Molecular studies of a mutation considered benign to the overall phenotype may show its interactions generate a condition that causes it to be deleterious. A mutation is considered pathogenic when it is causally relevant to the disease phenotype in question at relatively high penetrance.
- Every new disease gene discovery is an old disease gene discovery that is expressing in a different way phenotypically. If a variant is deemed to be of unknown significance, the right phenotype may not be under consideration or the phenotype may not yet have a clinical name. This is a daily challenge.

SESSION 3: DISRUPTIVE TECHNOLOGIES

MODERATOR

Hu Li, Mayo Clinic

PRESENTERS

Chad Myers, CIFAR Fellow, University of Minnesota – Computational Approaches to Understanding Complex Disease

Jennifer Listgarten, University of California, Berkeley – Machine Learning Meets Genome Editing

Prashant Mali, University of California San Diego – Pushing the Boundary of CRISPR Application

- Several variants are associated with any complex disease and it's thought that each variant contributes to a certain degree. Yet, modelling all the variants would not fully explain the disease. Complex diseases likely involve genetic interactions, as mutations can combine and become pathogenic. This is too vast to measure experimentally; computational approaches are needed to make inferences about genetic interaction networks and predictive models to further our understanding of gene function interactions.
- The unprecedented amount of data generated by high-throughput studies will revolutionize our understanding of biology. However, the combinatorial space is too great for experimental measurement; for example, pairwise mutations create 441 million experimental measurements. Sophisticated computational approaches are needed to interpret the data, enabling holistic cell models. New machine learning approaches could infer predictive models from genetic interaction networks and thereby further understanding of gene function and interaction.
- New machine learning tools could make CRISPR-Cas9 more efficient at staying on-target as well as targeting the optimal gene sequence for a desired knockout effect. To train the algorithms, computational scientists need to work alongside biologists to systematically measure CRISPR's efficacy for several genes. The algorithms can then learn from the data of real-world examples and later predict target sites for entirely different genes that have not been measured.
- The genome itself is now a "druggable" target. In the context of precision oncology, CRISPR is being used to discover novel synthetic lethal interactions. This can support discovery of new combination therapies where the oncogenic mutation can be exploited and targeted in combination with a drug targeting a parallel pathway. The synthetic lethal interaction produced can therefore induce cancer-cell-specific death while sparing normal cells.
- With regards to drug discovery, much of what applies to genetic interactions applies to chemical-genetic interactions. Many of the genetic interaction approaches developed in model organisms like yeast are now used as a template for human cells. Scientists can use chemical-genetic interactions to build profiles for compounds' effects on cells and in turn build increasingly precise maps of where in the cell the compounds target.
- Data integration is complex: the data can appear robust, but artifacts become apparent. Multiple approaches are needed to address this, such as assaying numerous different readouts, using cross-validation, and reproducing results across several research groups.
- Looking to the future, finding the right framework to do a disease screen will be important. For example, understanding whether a screen accounts for background heterogeneity and is representative of the tumour. Developing a framework that is truly representative of the disease system could allow for more meaningful outputs.

SYNTHESIS AND INSIGHTS: DAY 1

- The “Brave New World” is already here. We can read the genome sequence, write the genome sequence using synthetic biology, and edit the genome sequence using CRISPR.
- We need more comprehensive, genome-wide data now that technologies have advanced to be better and cheaper.
- Accessibility of data matters – namely, wider access to published data sets from model organisms.
- In leveraging genetic networks, greater emphasis is needed on how the lessons learned across different model organisms can be integrated – from yeast to multicellular organisms.
- Workable interdisciplinary teams need to be designed to support the identification of questions at the frontier of this brave new world. Interdisciplinary training is a huge challenge; biologists and computational experts have to communicate with each other, and systems must enable those interactions to happen more.

DAY TWO

SESSION 4: NATURAL VARIATION AND POPULATION GENETICS

MODERATOR

Philip Awadalla, Director, Computational Biology, Ontario Institute for Cancer Research

PRESENTERS

Nancy Cox, Vanderbilt University – Looking Across the Population’s Phenotypic Landscape

Frederick Roth, CIFAR Senior Fellow and Co-Director; University of Toronto – Interpreting Rare Genetic Variants

- Biobanks are discovery engines for gene-phenotype relationships. The strength of a biobank comes from its ability to look across an entire medical phenome at once. By collecting medical information and samples from large populations of millions of participants, they enable powerful studies to improve our understanding of complex disease.
- Interpreting the disease risk of variants in individual human genomes is a major challenge in medicine. With the effects of only a small portion of catalogued variants currently known, understanding which genetic variants matter remains a problem. Given the challenge in carrying out human studies at the scale and of the nature needed to interpret rare variants, a “humanized yeast” system offers the ability to pinpoint the molecular dysfunction of all clinical variants from patients with a given disorder. This will support a better understanding of the relationship between rare missense variants and protein function.
- As sequencing technologies improve, a shift from targeted to whole genome sequencing will be needed. Given issues of scalability, analyzing and interpreting whole genome sequencing poses a challenge in addition to cost and time.
- New computational tools to interpret genome sequencing can provide clinicians with functional data and computational predictors. However, when clinicians use research results in the clinic, they want one piece of evidence to support a final decision incorporating their expertise, knowledge of family history, and phenotype. They do not want terms such as “likely pathogenic” or “benign”. More useful will be likelihood ratios, indicating the increase in probability as a result of pathogenic evidence.

SESSION 5: BREAKING DOWN SILOS BETWEEN ACADEMIA, INDUSTRY AND THE CLINIC – MODEL SYSTEMS TO HUMAN POPULATIONS

MODERATOR

Andrew Hessel, Humane Genomics Inc

PRESENTERS

Calum MacRae, Harvard University – Human Genetics and the Rate-limiting Step: Phenotype

Marc Fiume, DNA Stack – The Internet is the Greatest Tool for Sharing Information

Amit Deshwar, DeepGenomics – Creating a New Universe of Genetic Medicines

- Patients don't care what their genome looks like; they care about feeling better and living longer. Phenotypes and genomics need to be incorporated into interventions that will support those outcomes.
- We're measuring the wrong things. We need to think systematically about where meaningful information might be (e.g. cellular mechanisms) and how we can build platforms to collect those readouts and data at the scale necessary to make meaningful conclusions.
- We need to harness recent advancements in technology as diverse as facial recognition, wearables, and short-term drug response to collect data and gain a system-wide view of a disease. Generalizable rules for interpreting this data are needed from researchers to fuel models of disease prediction. This can bring cell biology and physiology to the bedside.
- Pharmaceutical development will initially focus on rare diseases because the genotype-phenotype mapping is so clear and the medical need is so obvious but these pharmaceuticals could one day treat any sort of complex disease. Large scale sequencing projects will enable us to identify more outliers of very extreme phenotypes and, once the mutation is identified, a therapeutic target is created.
- The current pharmaceutical R&D system isn't working. The discovery pipeline is slow and costly to move a drug from discovery to testing to trials. Thousands of diseases have no known treatment and many more have substantial unmet medical need. Despite enormous advancements in genomics, sequencing, and systems biology, the ROI from pharmaceutical R&D expenditures has continuously dropped over the last 25 years to below the capital investment.
- The need for comprehensive automation in the drug discovery process is great; the predictive capacity of machine learning tools can help address this.
- Genomics lacks the availability of high quality, high volume data sets and the willingness and commitment to share the data. Data generation is the first limiting factor: fragmented infrastructure in the genomics landscape makes it difficult to access, use, and trust data, leading to a very low ROI on data generation.
- The second limited factor is data sharing capacity. Sharing data is critical as no single organization can generate enough data to cure all complex and rare diseases. The internet has shown that we can share data, and in nearly every industry we see the disruption.
- A new world for genomics would be powered by the internet to enable machine learning to power faster and more complex discoveries that increase ROI on data generation. A secure internet for genomics would pave the way for technical integrations that allow data to flow securely and efficiently between systems, provide on-demand access, break through traditional data silos, and help maximize the value of genomics and health data.

Jing Hou, PhD, contributed to the writing of this report.