CENTER FOR THE MULTIPLEX ASSESSMENT OF PHENOTYPE

NEWSLETTER

Summer | August 26, 2020

https://www.cmap.gs.washington.edu

WELCOME

The Center for the Multiplexed Assessment of Phenotype, (CMAP) is a Center of Excellence in Genome Sciences, supported by the National Human Genome Research Institute. Our goal is to develop technologies to assess the functional impact of variants in human genes. Linking phenotype to genotype is one of the most pressing problems in biology and our goal is to facilitate variant interpretation to enable genome- guided precision medicine in clinical decision making. We are based at the University of Washington and at the University of Toronto.

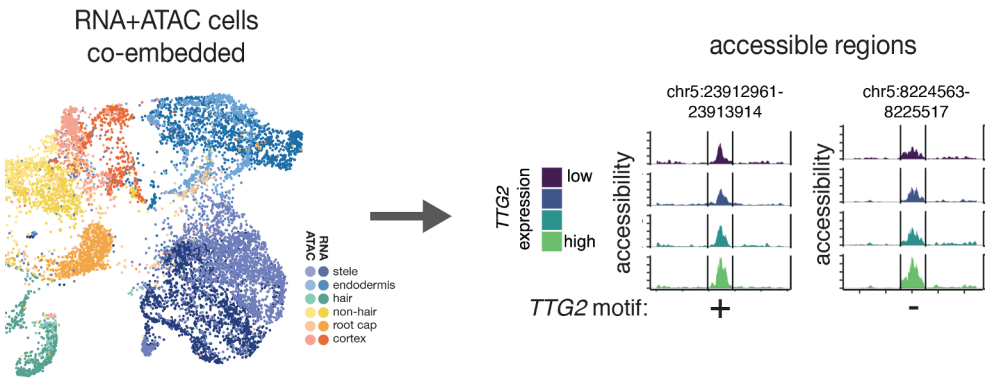
ANNUAL RETREAT

Thanks to everyone for attending CMAP's very first Annual Retreat! We met virtually for two back to back morning sessions on August 3rd and 4th. We certainly packed it in! The first day started with a State of the Center address by Doug followed by presentations each lab group. We also heard about cutting edge research and work in progress from the team in 12 lightning talks, which were followed by randomized breakout discussion sessions. The second day was dedicated to brainstorming. Working groups tackled questions such as " How do we make variant effect mapping more time and cost efficient?" and "What new information-rich molecular and cellular phenotypes should we consider?" These breakout sessions were very energizing and spawned new ideas. And hey, CMAP SWAG is the best. Who doesn't love a mug and a mask combo?

Capturing complexity in gene regulation

by Michael Dorrity

How integration of diverse molecular phenotyping data can allow us to better understand genetic perturbations



The advent of single-cell genomics has rapidly realized a dream shared by many who have analyzed microarray or RNA-sequencing data: to view the state of each cell individually, without confounding factors of cell type, cell cycle progression, or developmental stage. With single cell measurements, each of these sources of heterogeneity in a typical biological sample can be deconvolved, so long as we have the right tools to do so. As we profile single cells for molecular phenotypes beyond transcription, the potential to resolve individual cell states is greater still, but depends on our ability to leverage computational tools to integrate different types of data. In our recent work, “The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution,” we use profiling of open chromatin by single cell ATAC-seq (scATAC) alongside previously generated transcriptional profiling data by single cell RNA-sequencing (scRNA) to discern cell states of the root, the continuously-developing tissue responsible for acquisition of nutrients and water required for plant growth.

Open chromatin profiling is a valuable source of data because: (1) it contains information about cell identity –a root epidermis cell, for example, has many accessible chromatin regions that are not “open” in root vasculature cells; (2) regions of accessible chromatin represent sites of transcription factor binding, which allow for the dynamic gene regulatory events that are required for cell differentiation during development, or for responding to environmental cues.

Beyond more precise definition of cell states of the root, we combined these types of data to infer the individual regulatory events that specify cell type identity. Can we link the expression level of transcription factors, the regulatory proteins responsible for switching genes on and off, to the accessible sites that “open” up in root epidermis cells? We find, using a combination of dataset integration tools (Stuart et al. 2019) and modeling with Monocle3 (Cao et al. 2019), that transcription factors with cell type specific expression patterns can be linked to individual accessible regions of the genome, providing a first step toward building stepwise models of gene regulatory programs that are deployed during development.

Because gene regulation is complex, involving thousands of individual genes, and thousands of distinct accessible regulatory sites, each differing depending on cell type, spatial tissue patterning, and genetic background, the models described in our work are a simple starting point. As more molecular phenotyping data accumulates to describe how cell states change during development and in response to genetic perturbations, we will need to push our tools even further to capture the full complexity of eukaryotic gene regulation.

References

• The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. Dorrity et al. 2020 *bioRxiv.*

• The single-cell transcriptional landscape of mammalian organogenesis. Cao et al. 2019 *Nature*.

• Comprehensive integration of single-cell data. Stuart et al. 2019 *Cell.*

to learn more

Please visit our website - https://www.cmap.gs.washington.edu/publications

RECENT NEWS

Our team has adapted as best we can to the pandemic. Classes and meeting are held over Zoom and we are taking time to tend to our mental health by spending time outdoors.

Words of wisdom from Stanley Fields who is going on Sabbatical in 2021.

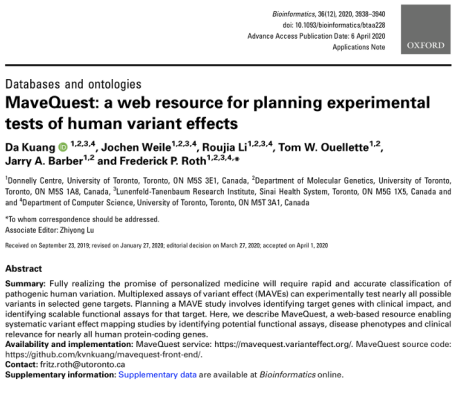
"I tell graduate students to try wild things—try ideas that might be off-the-wall. This is the time in your life to have fun, be creative, tackle tough projects, and chase big ideas. If you become cautious right at the beginning, I think science becomes too much like a real job as opposed to something we do because we enjoy it and we're passionate about it." ~ Stanley Fields

IN THE SPOTLIGHT

Recent publication! "**Kuang, et al. 'MaveQuest: a web resource for planning experimental tests of human variant effects.**' Bioinformatics (2020)"

Fully realizing the promise of personalized medicine will require accurate and rapid classification of pathogenic human variation. Multiplexed assays of variant effect (MAVEs) can experimentally test nearly all possible variants in selected gene targets. MaveQuest, developed by CMAP researchers Kevin Kuang, Jochen Weile, Roujia Li, and others in the Roth Lab, is an online resource for querying literature-curated functional assays, phenotypes and evidence of clinical interests of human genes for MAVE studies. Visit **https://mavequest.varianteffect.org** to start exploring potential assays and phenotypes for human gene variants!

Roth Lab / University of Toronto



Bioinformatics, 36(12), 2020, 3938–3940
doi: 10.1093/bioinformatics/btaa228
Advance Access Publication Date: 6 April 2020
Applications Note

OXFORD

Databases and ontologies

MaveQuest: a web resource for planning experimental tests of human variant effects

Da Kuang [1,2,3,4], Jochen Weile [1,2,3,4], Roujia Li [1,2,3,4], Tom W. Ouellette [1,2], Jarry A. Barber [1,2] and Frederick P. Roth [1,2,3,4,*]

[1]Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada, [2]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada, [3]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON M5G 1X5, Canada and and [4]Department of Computer Science, University of Toronto, Toronto, ON M5T 3A1, Canada

*To whom correspondence should be addressed.
Associate Editor: Zhiyong Lu

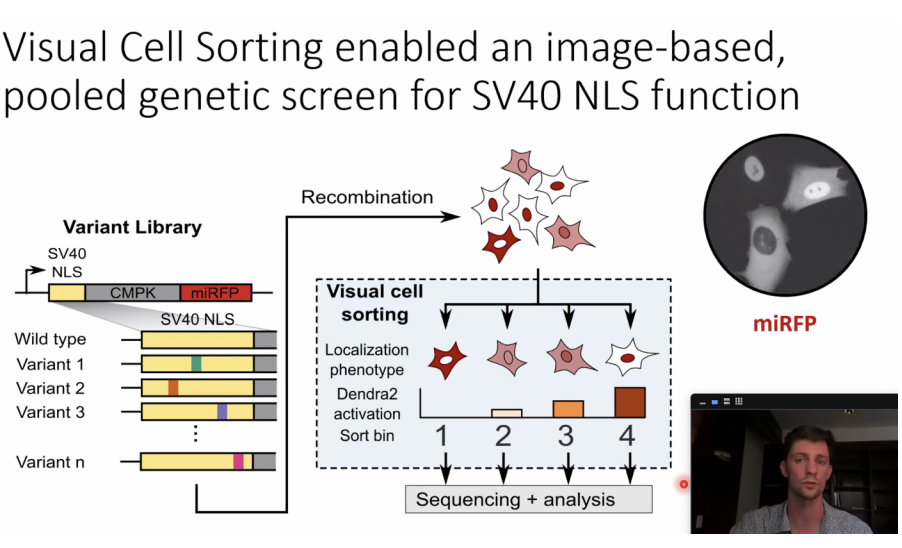
Received on September 23, 2019; revised on January 27, 2020; editorial decision on March 27, 2020; accepted on April 1, 2020

Abstract

Summary: Fully realizing the promise of personalized medicine will require rapid and accurate classification of pathogenic human variation. Multiplexed assays of variant effect (MAVEs) can experimentally test nearly all possible variants in selected gene targets. Planning a MAVE study involves identifying target genes with clinical impact, and identifying scalable functional assays for that target. Here, we describe MaveQuest, a web-based resource enabling systematic variant effect mapping studies by identifying potential functional assays, disease phenotypes and clinical relevance for nearly all human protein-coding genes.
Availability and implementation: MaveQuest service: https://mavequest.varianteffect.org/. MaveQuest source code: https://github.com/kvnkuang/mavequest-front-end/.
Contact: fritz.roth@utoronto.ca
Supplementary information: Supplementary data are available at *Bioinformatics* online.

Kuang et al 2000

THESIS DEFENSE - NICK HASLE



CMAP researcher **Nick Hasle** defended his thesis "Using visual phenotypes to dissect sequence-function relationships and complex drug responses" on May 19th via Zoom. Congratulations Dr. Hasle!