Uncovering explanatory power of large-scale expression data

The ability to measure molecular traits in biological systems has dramatically expanded in recent years, enabling the measurement of tens of thousands molecular phenotypes across thousands of individuals and multiple tissue types and associate those with millions of variants in the genome. Deriving actionable, biological interpretations from these datasets remains a challenging problem, and this Data Analytics challenge focuses on generating useful interpretations from large, populationwide, multi-tissue, molecular phenotyping assays. Specifically, the genes that predict tissue state will be identified, and genetic perturbations will be leveraged to characterize the genes that act in concert to orchestrate overt phenotypes.

Challenge

The GTEx datasets [1] can be found at their Download page. GTEx Portal uses Google Sign-In, so access will require a Google account.

Description	Name	Size
Gene RPKM	GTEx_Analysis_v6p_RNA-seq_RNA-SeQCv1.1.8_gene_rpkm.gct.gz	1.9G

Table 1: RNA-seq Data: RPKMs for genes x tissues

Description	Name	Size
eGene and significant snpgene associations based on permutations	GTEx Analysis v6p eQTL.tar	782M

Table 2: Single Tissue cis-eQTL Data: association statistics for markers (snpgene) x tissues

Given the GTEx data sets, find solutions to the following questions.

1. How accurately can tissue identity be predicted from expression state, and what are the key genes driving that classification?
2. What are the modules of genes whose expression is being genetically perturbed?
3. For any given gene (or sub-network of genes), what effect will increasing or decreasing its expression have on the system and what is the confidence of that prediction?
4. What is the extent with which directionality of gene effects can be inferred (i.e. causal inference)?

References

[1] The GTEx Consortium, Kristin G. Ardlie, David S. Deluca, Ayellet V. Segr`e, Timothy J. Sullivan, Taylor R. Young, Ellen T. Gelfand, Casandra A. Trowbridge, Julian B. Maller, Taru Tukiainen, Monkol Lek, Lucas D. Ward, Pouya Kheradpour, Benjamin Iriarte, Yan Meng, Cameron D. Palmer, T˜onu Esko, Wendy Winckler, Joel N. Hirschhorn, Manolis Kellis, Daniel G. MacArthur, Gad Getz, Andrey A. Shabalin, Gen Li, Yi-Hui Zhou, Andrew B. Nobel, Ivan Rusyn, Fred A. Wright, Tuuli Lappalainen, Pedro G. Ferreira, Halit Ongen, Manuel A. Rivas, Alexis Battle, Sara Mostafavi, Jean Monlong, Michael Sammeth, Marta Mele, Ferran Reverter, Jakob M. Goldmann, Daphne Koller, Roderic Guig´o, Mark I. McCarthy, Emmanouil T. Dermitzakis, Eric R. Gamazon, Hae Kyung Im, Anuar Konkashbaev, Dan L. Nicolae, Nancy J. Cox, Timoth´ee Flutre, XiaoquanWen, Matthew Stephens, Jonathan K. Pritchard, Zhidong Tu, Bin Zhang, Tao Huang, Quan Long, Luan Lin, Jialiang Yang, Jun Zhu, Jun Liu, Amanda Brown, Bernadette Mestichelli, Denee Tidwell, Edmund Lo, Mike Salvatore, Saboor Shad, Jeffrey A. Thomas, John T. Lonsdale, Michael T. Moser, Bryan M. Gillard, Ellen Karasik, Kimberly Ramsey, Christopher Choi, Barbara A. Foster, John Syron, Johnell Fleming, Harold Magazine, Rick Hasz, Gary D. Walters, Jason P. Bridge, Mark Miklos, Susan Sullivan, Laura K. Barker, Heather M. Traino, Maghboeba Mosavel, Laura A. Simino↵, Dana R. Valley, Daniel C. Rohrer, Scott D. Jewell, Philip A. Branton, Leslie H. Sobin, Mary Barcus, Liqun Qi, Je↵rey McLean, Pushpa Hariharan, Ki Sung Um, Shenpei Wu, David Tabor, Charles Shive, Anna M. Smith, Stephen A. Buia, Anita H. Undale, Karna L. Robinson, Nancy Roche, Kimberly M. Valentino, Angela Britton, Robin Burges, Debra Bradbury, Kenneth W. Hambright, John Seleski, Greg E. Korzeniewski, Kenyon Erickson, Yvonne Marcus, Jorge Tejada, Mehran Taherian, Chunrong Lu, Margaret Basile, Deborah C. Mash, Simona Volpi, Je↵ery P. Struewing, Gary F. Temple, Joy Boyer, Deborah Colantuoni, Roger Little, Susan Koester, Latarsha J. Carithers, Helen M. Moore, Ping Guan, Carolyn Compton, Sherilyn J. Sawyer, Joanne P. Demchok, Jimmie B. Vaught, Chana A. Rabiner, Nicole C. Lockhart, Kristin G. Ardlie, Gad Getz, Fred A. Wright, Manolis Kellis, Simona Volpi, and Emmanouil T. Dermitzakis. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science, 348(6235):648–660, 2015.