GTEx Studies Provide Comprehensive Map of Genetic Regulatory Variation Across Cell Types, Tissues

Sep 10, 2020 | Christie Rizk

NEW YORK – Concluding a 10-year multi-institutional research effort, the Genotype-Tissue Expression (GTEx) Consortium published its final set of studies this week, presenting a comprehensive atlas of genetic regulatory variation across cell types and tissues and an analysis of how these changes in regulation can contribute to risk for disease and the development of traits.

"GTEx was initially designed to partially address the dilemma of being able to identify what the functional mechanisms of genetic variants are that are associated with complex diseases and traits," said Tuuli Lappalainen, an assistant professor of systems biology at Columbia University and a core member at the New York Genome Center. Lappalainen was a member of the GTEx Consortium and co-led several studies published in *Science*.

"A little bit over 10 years ago when GTEx got started, the first GWAS started coming out and people realized that most of these associations were in noncoding regions," she said. "[The hypothesis was that] these variants must have some kind of a regulatory function, but what is that, and how do we find out what those are?"

GTEx published earlier-stage analyses in 2015 and 2017 — the 15 new papers coming out this week are based on the GTEx version 8 dataset, a deep survey of tens of thousands of regulatory variants, and made use of a variety of technologies. The researchers conducted RNA sequencing on 15,201 samples from 49 tissues of 838 post-mortem donors and analyzed whole-genome sequencing data for each donor. A key methodology was expression quantitative trait locus (eQTL) analysis to identify genetic variants that affected gene expression, as well as the analysis of splicing QTLs (sQTL). One of the papers also used CRISPR to analyze regulatory variants in rare diseases.

Five of the papers are published in *Science*, two are in *Science Advances*, one in *Cell*, five in *Genome Biology*, one in *Genetic Epidemiology*, and one in *Genome Medicine*.

In their main paper in *Science*, the researchers described their objectives and methods, emphasizing the ancestral and gender diversity of the donors in the dataset. Of the 838 donors, 715 (85.3 percent) were European American, 103 (12.3 percent) were African American, 12 (1.4 percent) were Asian American, and 16 (1.9 percent) had Hispanic or Latino ethnicity; 557 (66.4 percent) donors were male and 281 (33.5 percent) were female.

Overall, the authors wrote, this update to the GTEx dataset substantially expanded the catalogue of sQTLs, doubled the number of known genes with an eQTL per tissue, and "saturated the discovery of eQTLs with greater than twofold effect sizes in [about] 40 tissues."

They conducted fine-mapping analysis of *cis*-eQTLs, resulting in a set of thousands of likely causal functional variants. They also analyzed cell type interaction *cis*-eQTLs and *cis*-sQTLs and mapped them with computational estimates of cell type enrichment, amplifying their understanding of the effects of cell types within tissues.

The highly similar patterns shared across tissues cross these data types suggested a shared biology from cell type composition to transcriptome variation and genetic regulatory effects, the researchers noted. These results

indicated that shared cell types between tissues may be a key factor behind tissue sharing of genetic regulatory effects.

Finally, GWAS co-localization with *cis*-eQTLs and *cis*-sQTLs provided them with data for further functional follow-up studies and characterization of regulatory mechanisms of GWAS associations.

In a second *Science* paper, the researchers explored the impact of sex on gene expression across human tissues. By integrating sex-aware analyses of GTEx data with gene function and transcription factor binding annotations, they found multiple sex-differentiated genetic effects on gene expression that co-localized with complex trait genetic associations.

For example, they identified 58 gene-trait associations driven by genetic regulation of gene expression in a single sex, including loci where sex-differentiated cell type abundances mediated genotype-phenotype associations, as well as loci where sex may play a more direct role in the underlying molecular mechanism of the association.

In another *Science* paper, the team explored cell type-specific genetic regulation of gene expression across tissues, concluding that the large majority of cell type-specific QTLs remains to be discovered. The investigators' co-localization results indicated that comprehensive mapping of cell type-specific QTLs will be highly valuable for gaining a mechanistic understanding of complex trait associations.

In yet another study in *Science*, researchers analyzed determinants of telomere length across human tissues, and found that variation in relative telomere length was attributable to tissue type, donor, and age, as well as, to a lesser extent, race or ethnicity, smoking, and inherited variants known to affect leukocyte telomere length. For example, African ancestry was associated with longer relative telomere length across all tissues and within specific tissue types, suggesting that ancestry-based differences in telomere length exist in germ cells and are transmitted to the zygote. The researchers also found evidence that relative telomere length may mediate the effect of age on gene expression in human tissues.

An additional study in *Science* explored functional rare genetic variation found in transcriptomic signatures across human tissues. Here, the researchers assessed how rare genetic variants contributed to extreme patterns in gene expression (eOutliers), allelic expression (aseOutliers), and alternative splicing (sOutliers). After identifying multi-tissue eOutliers, aseOutliers, and sOutliers, they found that outliers of each type were significantly more likely to carry a rare variant near the corresponding gene, and developed a probabilistic model called Watershed for personal genome interpretation that improved over standard genomic annotation-based methods to score rare variants by integrating these three transcriptomic signals from the same individual and replicating them in an independent cohort. They found that transcriptome-assisted prioritization identified rare variants with larger trait effect sizes and were better predictors of effect size than genomic annotation alone.

In a *Science Advances* study, researchers looked at how tissue-specific genetic features could inform the prediction of drug side effects in clinical trials. They determined that drug target genes with five genetic features — tissue specificity of gene expression, Mendelian associations, phenotype- and tissue-level effects of genome-wide association loci driven by eQTL, and genetic constraint — conferred a 2.6-fold greater risk of side effects, compared to genes without such features.

For a second *Science Advances* study, researchers developed a resource called PhenomeXcan in order to map the genome to the phenome through the transcriptome. Using PhenomeXcan, the researchers synthesizes 8.87 million variants from GWAS summary statistics on 4,091 traits with transcriptomic data from 49 tissues in the GTEx v8 dataset into a gene-based, queryable platform that included 22,515 genes. They were able to provide examples of novel and underreported genome-to-phenome associations, complex gene-trait clusters, shared causal genes between common and rare diseases via further integration of PhenomeXcan with ClinVar, and potential therapeutic targets.

In the *Cell* study, a team led by Stanford's Micheal Snyder reported a quantitative proteome map of the human body. They quantified the relative protein levels of more than 12,000 genes across 32 normal human tissues,

and identified tissue-specific or tissue-enriched proteins and compared them to transcriptome data. The researchers found that many ubiquitous transcripts encoded tissue-specific proteins, and that discordance of RNA and protein enrichment pointed to potential sites of synthesis and action of secreted proteins. Importantly, they also noted that protein tissue-enrichment information can explain phenotypes of genetic diseases, which cannot be obtained by transcript information alone.

Looking back at the 10 years of the project and what the consortium was able to accomplish, Lappalainen said the researchers have learned a lot about biology and were able to answer many questions about the genome's function.

"I think that GTEx has fulfilled its promise in many ways. It has provided very comprehensive data," she said. "And also, the technology development during these years has introduced new approaches and new ways of answering these same questions. There is not a silver bullet, there is no thing or approach or study design that will resolve the function of the genome. So, I think that some of the things that we also allude to and partially addressed in these papers of thinking about cell type composition, thinking about gene regulation and genetic variant effects at the tissue level and at the cell type level is a very important angle that we can study here."

As for the data itself, it's widely accessible, as it has been through the past 10 years, through the GTEx portal. The consortium has been doing work to empower the larger research community, to power functional genomics research, provide backup for GWAS studies, and provide understanding for research on the potential regulatory effects of disease-associated variants, Lappalainen said. The cancer genomics community is also using GTEx widely to power its own studies, she noted, as well as researchers looking for answers on splicing patterns or variability for a given gene.

Despite the massive volume of data and the number of papers the consortium has now published, there's still more work to do, Lappalainen added. She plans to look at ways to combine genomic and RNA sequencing data and phenotypic data to better understand disease mechanisms, and is also thinking of looking at how the combination of environmental factors and genetic factors influence risk of disease.

Filed Under   **+ Genetic Research**   **+ Gene Expression Research**   **+ Sequencing**

+ Proteomics & Protein Research   **+ Cell Biology Research**   **+ Asia/Oceania**   **+ Europe**

+ North America   **+ rare disease**   **+ Broad Institute**   **+ Johns Hopkins**   **+ NYGC**

+ Stanford   **+ RNA-seq**