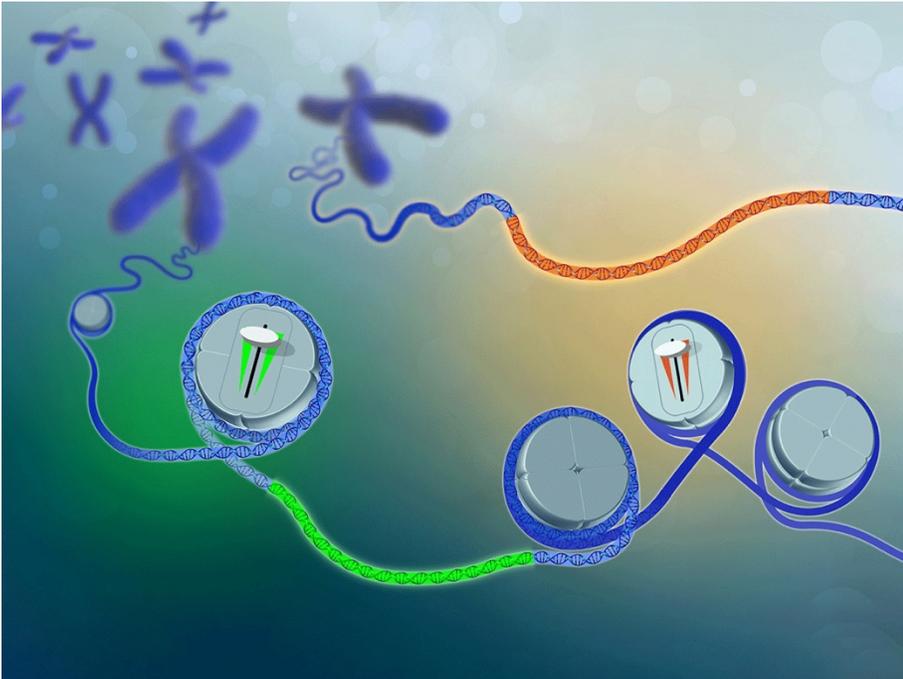# BioWorld™

# GTEx explores variation in genome bureaucracy



Sections of the genome, known as expression Quantitative Trait Loci (eQTL), work to control
how genes are turned off and on. Variants in eQTLs affect how much of a protein a cell will produce.
Credit: Darryl Leja, National Human Genome Research Institute, NIH

September 10, 2020
By Anette Breindl

The Genotype-Tissue Expression (GTEx) project, a multiyear, multi-institutional attempt to catalog how
expression quantitative trait loci (eQTL) and splicing quantitative trait loci (sQTL) affect protein
levels, reported data from its final phase in 15 papers in the Sept, 10, 2020, online issues of
the *Science* and *Cell* family of journals, as well as in *Genome Biology*.

Collectively, the findings report multiple new insights into how protein expression levels can vary by genotype,
sex, genetic ancestry, tissue type and other factors.

The Encyclopedia of DNA Elements (ENCODE) consortium, which reported data from its own third phase in
July, is also focused on noncoding elements of the genome. But it doesn't directly measure variation, which is
at the heart of GTEx's approach.

In that sense, the GTEx approach is "ultimately similar to GWAS" – genome-wide association studies – Tuuli
Lappalainen told *BioWorld*.

Lappalainen is a core faculty member at the New York Genome Center, an assistant professor of systems
biology at Columbia University, and the co-leader of the GTEx Consortium.

The consortium had previously published preliminary data on selected tissue types.

The current publications are much more comprehensive. To gain insight into variation in gene expression, the team looked at more than 15,000 RNA samples from 49 tissues of 838 postmortem donors as well as the whole genome sequencing data of each donor.

Individual papers used those data to analyze multiple different sources of variation.

The dataset skews to males, which make up two-thirds of donors, and Caucasians, at 85%.

Nevertheless, "now that we have a larger sample size, it's been possible to look at sex differences and different ancestries and… capture some of that diversity," Lappalainen said.

Although the dataset does include some samples from individuals, such as accident victims, that were healthy right up to the time of their death, the samples of necessity skew to elderly individuals that were ill.

"We correct away a lot of the sort of variance that has to do with causes of death, or ischemic time – how long it takes from the time of death to sample collection," she said. Nevertheless, the database "captures the complexity of human biological tissues and systems, and that's both a good thing and a bad thing."

The dozens of tissue types analyzed in the studies enable unprecedented insights into how variants translate into variation at a practical level. The researchers looked at samples from nearly a dozen different brain regions, as well as both sun-exposed and non-sun exposed skin.

They also support the overarching conclusion that an understanding of the effects of noncoding variants on health and disease will hinge on understanding differences not just at the level of tissue types, but at the level of cell types within those tissues.

The samples the consortium analyzed consisted of bulk tissue, although the researchers did use computation methods to estimate how much each of 64 major cell types contributed to each sample.

"Human tissues are not a homogenous mass of different cell types," Lappalainen said, and the team's analysis "demonstrates the importance of [cell types] in very concrete ways."
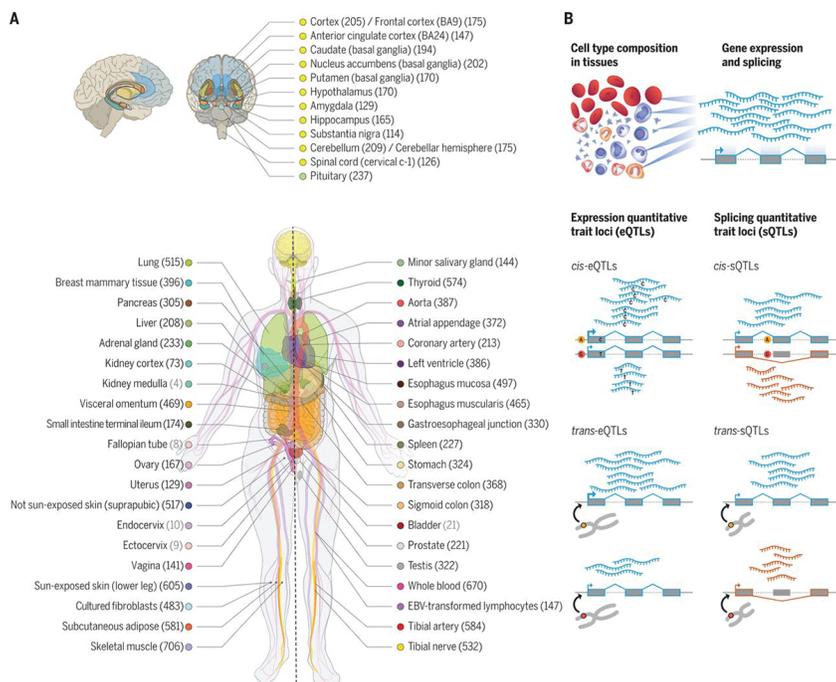


A. Illustration of the 54 tissue types examined. Tissues with 70 or more samples were included in QTL analyses.
B. Illustration of the core data types used throughout the study. Credit: Aguet et al., Science (2020)