**Regulatory variation quantified from allelic expression in GTEx: genetic architecture and disease applications**

Pejman Mohammadi, Tuuli Lappalainen

Estimating the amount of genetic variation affecting genes has become an essential metric for understanding the spectrum of genome variation, selective constraint on different genes, and prioritizing rare disease-causing mutations. While several existing approaches quantify the amount of genetic variation in coding regions, estimating the amount of regulatory variation has remained challenging.

We introduce Analysis of Expression Variation (ANEVA) to quantify the total amount of *cis* genetic variation affecting the expression of each gene in different tissues. This method uses allele-specific expression data from a population sample to derive genetic variation (Dg), and the total expression data to derive total variation that includes environmental components (Dt). Applying the method to the GTEx project v6 data, we obtained ANEVA estimates for 13,399 genes. We found that Dg/Dt is correlated with measures of coding constraint and haploinsufficiency but not with the size of regulatory regions per gene suggesting that it captures selective constraint on genes, rather than regulatory mutation rate.

Next, we developed a Dosage Outlier Test extension, ANEVA-DOT, that models total gene dosage from allelic expression to discover genes with allelic imbalance suggestive of a rare regulatory variant that can lead to a dosage outlier. Applying ANEVA-DOT to data from 26 patients with rare muscle disorders from Cummings et al. 2017, we found on average 24 such dosage outlier genes at 5% FDR. The genes where at least two patients were outliers were ~2.5 folds enriched for known neuromuscular disease genes (p=10-6). This demonstrates how careful modeling of genetic regulatory variation in the general population can be used to interpret potentially disease-causing variation in patient transcriptomes.