



## U01: Haussler and Paten - Unbiased analysis of genomic correlates of gene expression in health and disease

---

There is increasing recognition that rare, non-coding, and structural genomic variations are all important contributors to disease risk. We propose to create a graphical map of all the variations found by the TOPMed consortium - ultimately incorporating information from 100,000 genomes - that will allow unbiased analysis of all forms of variation in concert. The proof of the utility of this map will be a partially phased diploid reconstruction of each of the TOPMed genomes, compactly represented as a pair of paths for each chromosome in a global genome graph reference.

To demonstrate how such a genome graph can be transformative for integrative analysis, we will build the first *population gene annotation* - defined as the set of all the splice isoforms of genes being expressed, including their underlying haplotypes, within a comprehensive sampling of a population. A population gene annotation can be used to study the association between genetic variation and isoform expression in a statistically meaningful way. For example, given a population annotation, one could ask: "How does a given genetic variant - which may be in a non-coding region - affect the expression of a given isoform?", or "Which variants are associated with high expression of a given isoform in a particular disease state?". Typically such integrative analysis questions are hard-to-impossible to answer with current representations.

Hemoglobin disorders caused by genetic variants affecting the production or structure of the alpha and beta globin proteins are the most common inherited blood disorders, affecting millions of individuals worldwide. To demonstrate the power of the graphical approach, we will create the most comprehensive map of genetic variants in the alpha and beta globin loci and associated regulatory genes to date. We will show how combining this map with the phenotype data available from projects such as the Jackson Heart and the Women's Health Initiative, and RNA-Seq data from TOPMed and other projects such as the Genotype-Tissue Expression (GTEx) project ([gtexportal.org](https://gtexportal.org) [1]) allows us to identify novel candidates for causal variants and provide evidence that other rare variants are benign. This demonstration project will drive the research and improve both speed and precision in the diagnosis of hemoglobin disorders. It will provide a convincing demonstration of the value of this type of integrated approach to the analysis of all forms of genetic variation. As these hemoglobin disorders disproportionately affect certain genetic subpopulations, this study will show how the use of a more comprehensive reference structure, tunable to specific ethnic subpopulations, can reduce the potential biases that occur when relying on a single reference genome.

---

NHLBI Program officer: Maarten Leerkes

|                      |                                                                                                    |
|----------------------|----------------------------------------------------------------------------------------------------|
| <b>PI:</b>           | David Haussler                                                                                     |
| <b>Co-PI:</b>        | Benedict Paten                                                                                     |
| <b>Award Type:</b>   | U01 NHLBI TOPMed Program: Integrative Omics Approaches for Analysis of TOPMed Data (RFA-HL-17-011) |
| <b>Award number:</b> | U01 HL137183-01                                                                                    |
| <b>Start Year:</b>   | 2017                                                                                               |

**Source URL (modified on 06/10/2019 - 4:31pm):**<https://topmed.nhlbi.nih.gov/awards/2425>

#### Links

[1] <http://gtexportal.org/>