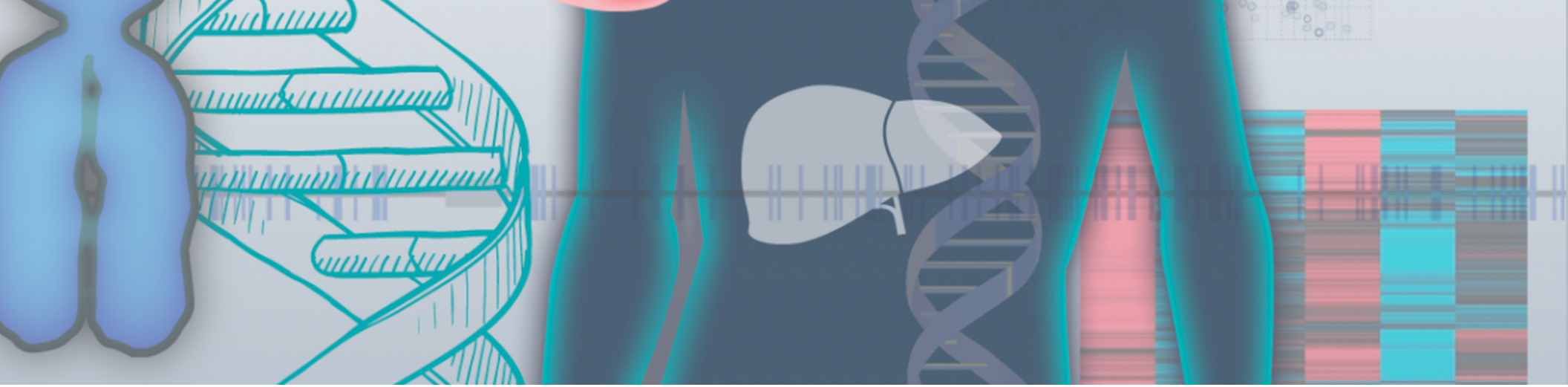


eQTL pipeline for V7 & V8

AWG call :: 08/14/2017

François Aguet

BROAD INSTITUTE

GTEx

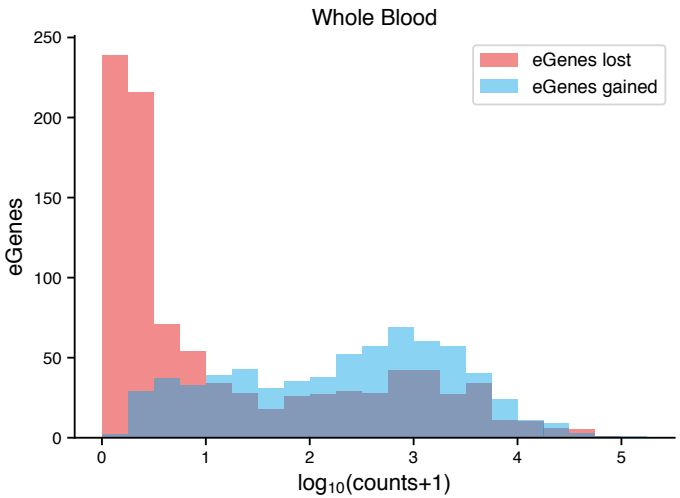
Summary

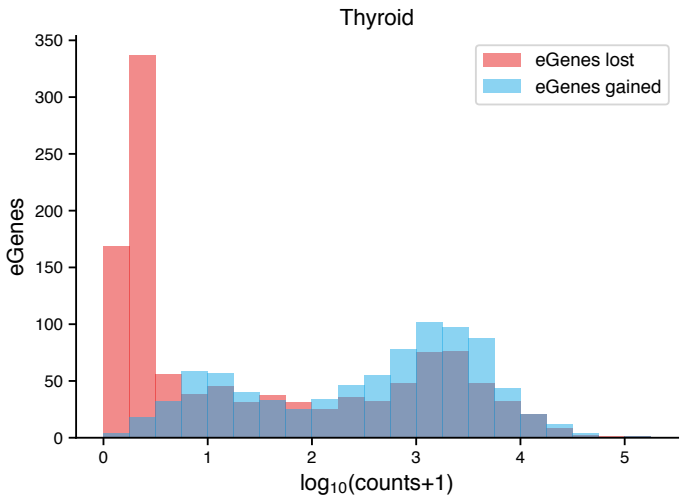
- **Proposed changes:**
  - Normalization: TMM instead of quantile normalization
    - Unchanged: inverse transform to standard normal
  - Expression/detection thresholds:
    - ≥6 counts in ≥20% of samples and
      >0.1 TPM in ≥20% of samples
    - Was:
      ≥6 counts in ≥10 samples and
      >0.1 FPKM in ≥10 samples
  - PEER factors: extension of prior approach

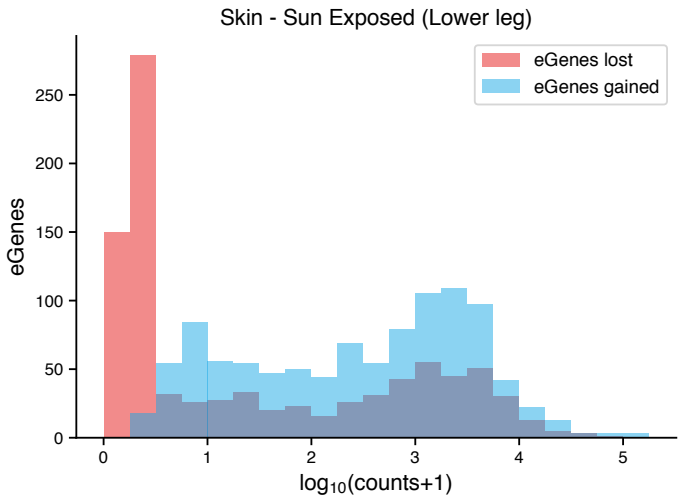
TMM normalization

- TMM: trimmed mean of M values (log fold-change)
  - Implemented in edgeR [Robinson & Oshlack, 2010]
  - Rescaling of count data; preserves zeros
    - Better for effect size calculations
- Consensus across benchmarks that TMM is generally the most suitable between-sample normalization method (together with DESeq)
  - References: Lin et al. BMC Genomics 2016; Rapaport et al. Genome Biology 2013; Dillies et al. Brief. Bioinformatics 2012

Comparison of normalization methods on V6p data



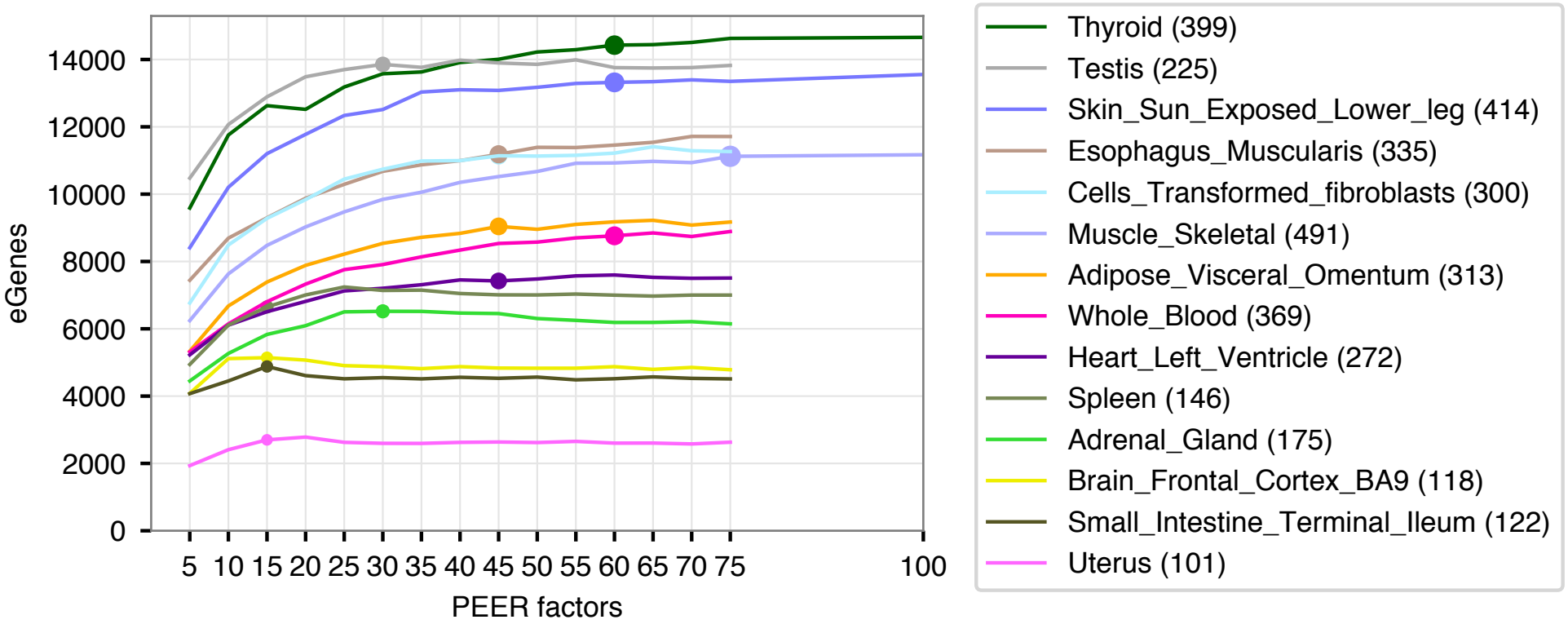




Tissue	v6p	v7 pipeline	shared	lost	gained
Thyroid	10610	10311	9461	1149	850
Skin - Sun Exposed (Lower leg)	9069	9167	8160	909	1007
Heart - Left Ventricle	4814	4649	4193	621	456
Whole Blood	7332	7055	6383	949	672

- Significant fraction of lost eGenes have median expression < 3 counts (!)

Selection of PEER factors



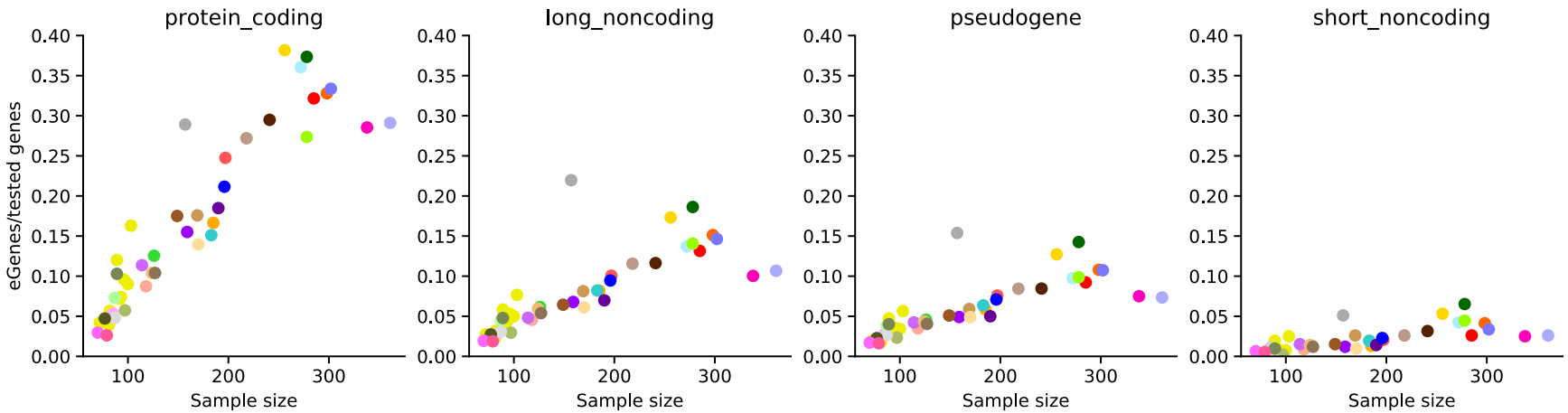
- Extension of approach from V6p paper
  - Selection of PEER factors based on eGenes detected, binned by sample size:

Sample size	PEER factors	Tissues
[0,150)	15	20
[150,250)	30	11
[250,350)	45	8
[350,450)	60	8
[450,550)	75	1

Discussion: FDR approach

- q-values or BH?
  - q-values: 'true' π_0 is 0 => use of q-values justified despite potential underestimation of FDR ?
  - BH: more conservative, but 'wrong' H_0

- Limited to protein coding genes and lincRNAs? Or all biotypes?



  - Based on stringent expression filters (unique mapping reads, edit distance), other biotypes unlikely artifacts
  - Consistent sets across papers and GTEx portal