

eQTL pipeline for V7 & V8

AWG call :: 08/20/2017

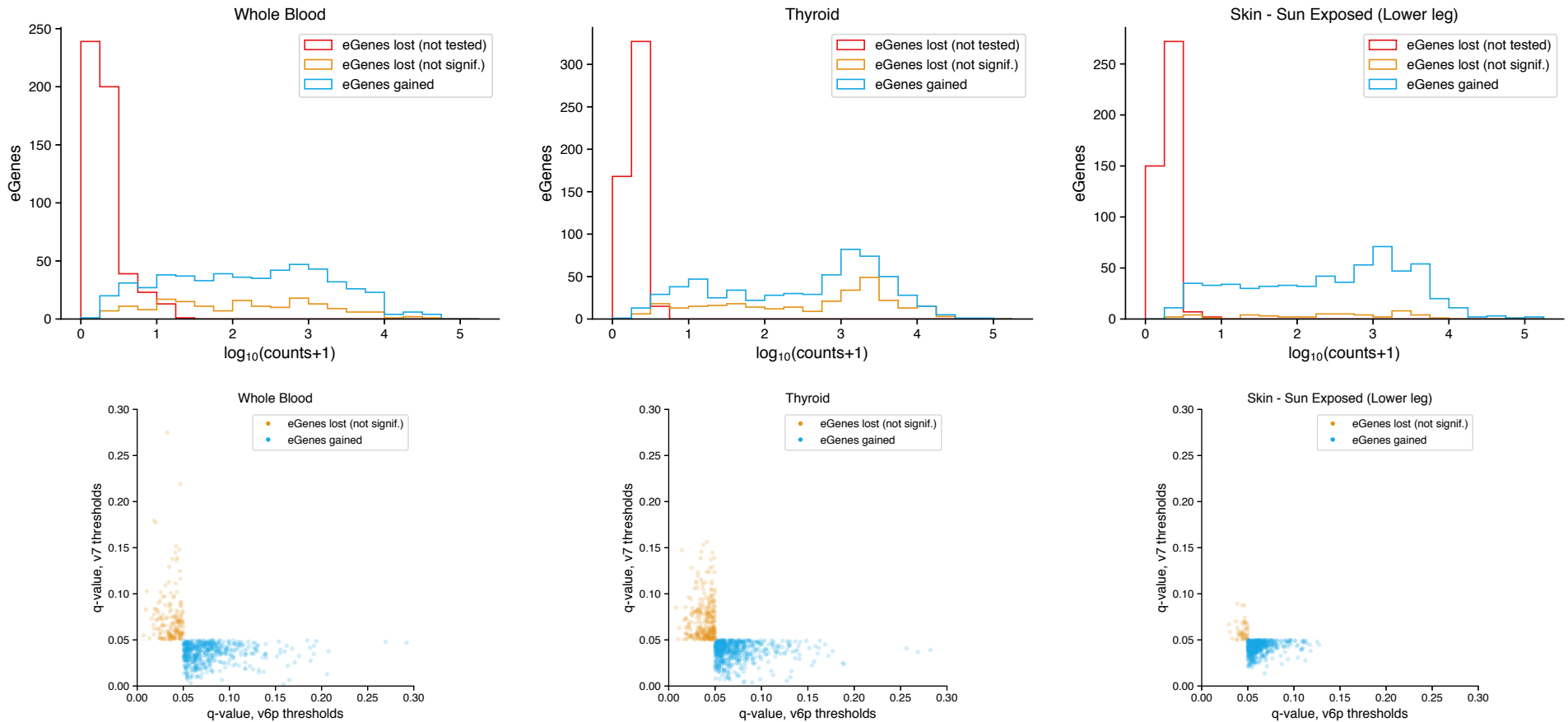
François Aguet

Summary

- **Proposed changes:**

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

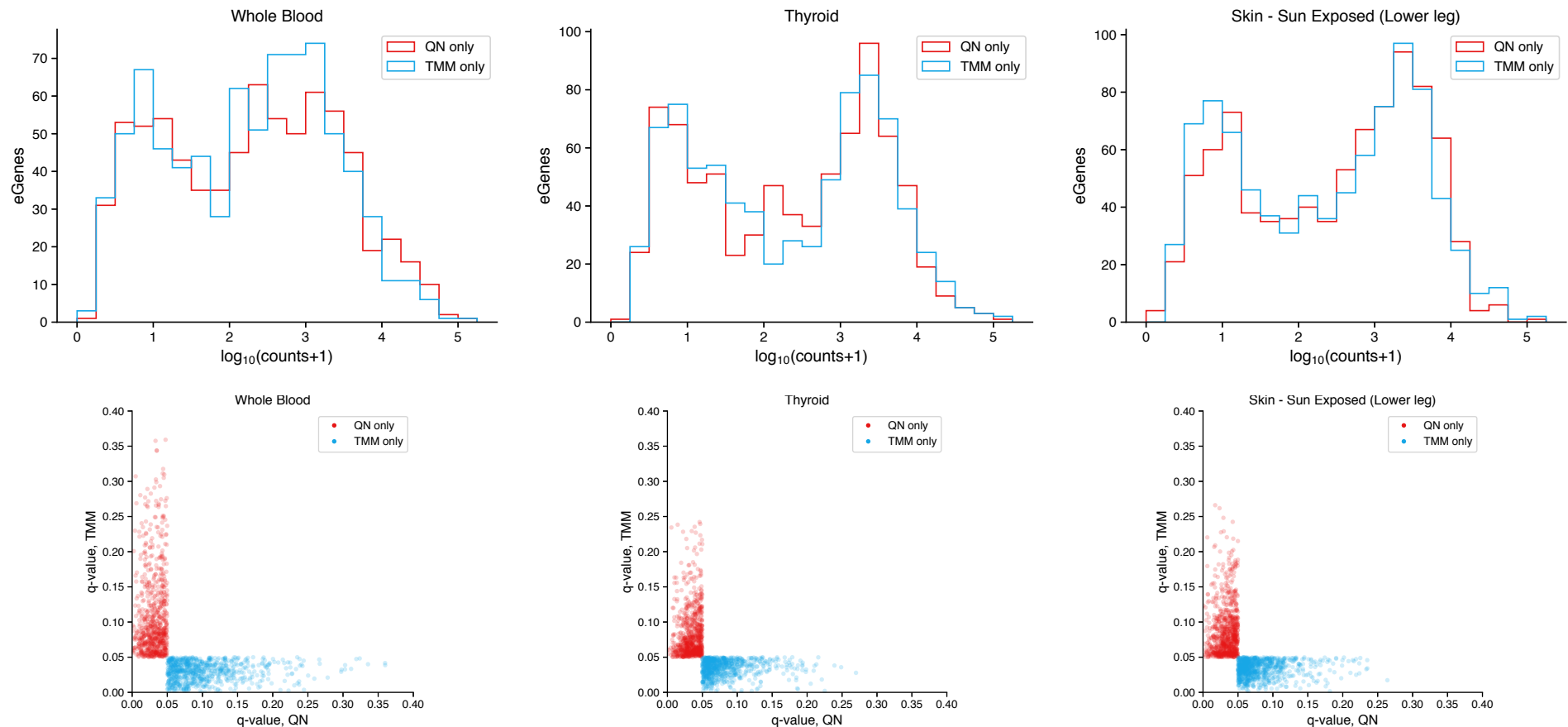
Effect of expression threshold on eGene discovery



Tissue	v6p	v7 thresholds	shared	lost	lost (n.s.)	gained
Thyroid	10610	10412	9808	510	292	604
Skin - Sun Exposed (Lower	9069	9172	8590	431	48	582
Heart - Left Ventricle	4814	4757	4522	165	127	235
Whole Blood	7332	7171	6647	515	170	524

- Comparison of v6p and v7 thresholds on v6p data
- Significant fraction of lost eGenes have median expression < 3 counts

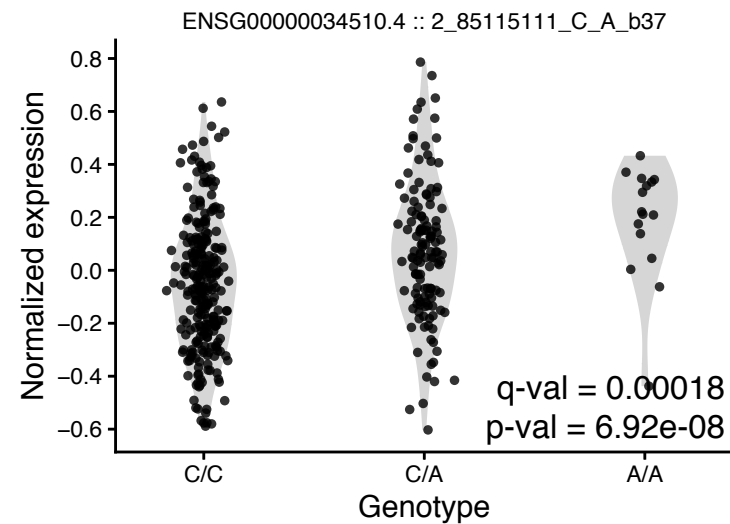
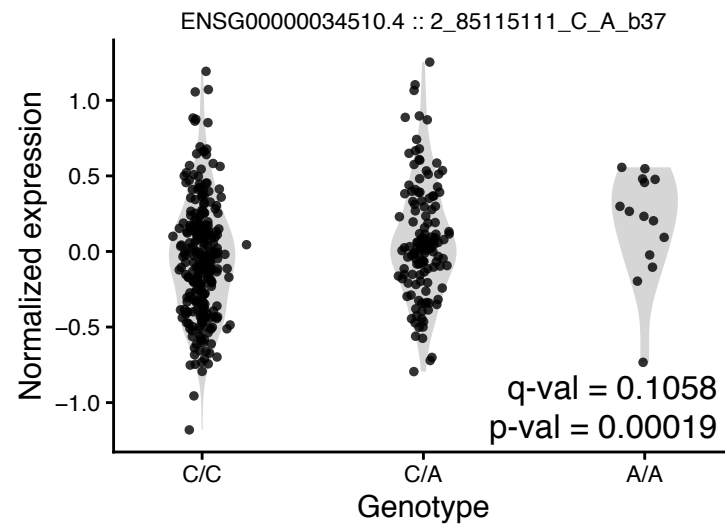
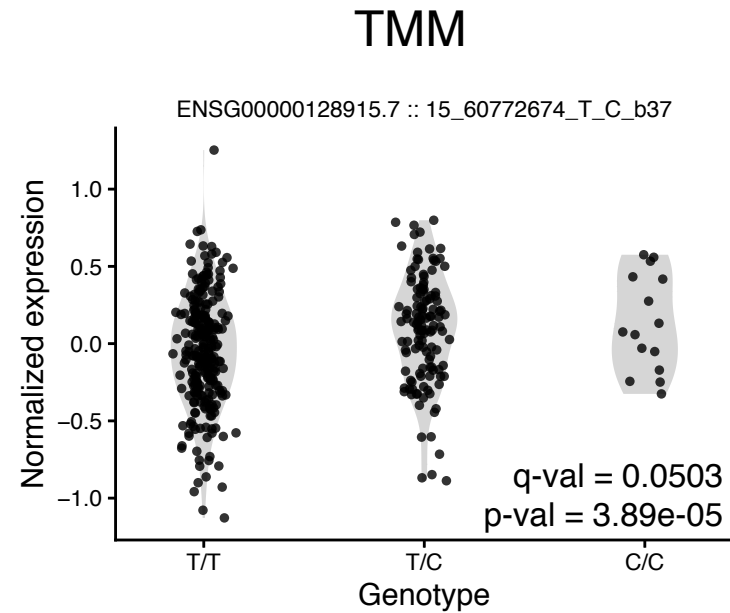
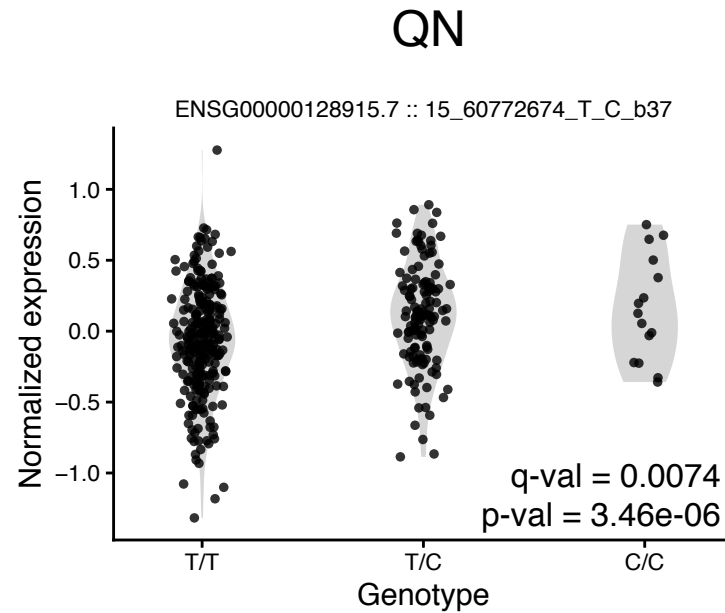
Effect of normalization method on eGene discovery



Tissue	QN	TMM	shared	QN only	TMM only
Thyroid	14340	14342	13544	796	798
Skin - Sun Exposed (Lower leg)	13090	13106	12222	868	884
Heart - Left Ventricle	7361	7370	6688	673	682
Whole Blood	8618	8659	7869	749	790

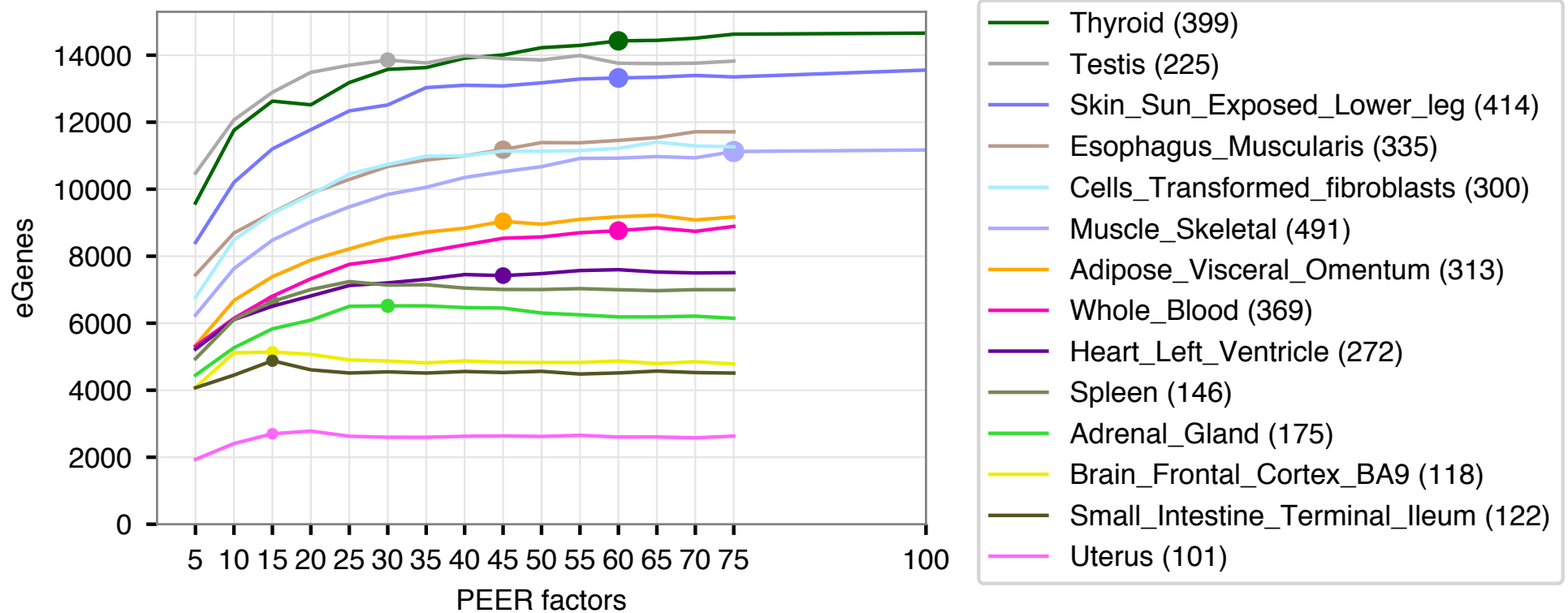
- Comparison of QN and TMM on v7 data

Examples of eGenes detected by one method only



- Changes in significance are generally small: s.d. of $\log_{10}(\text{p-value ratio}) \sim 1$

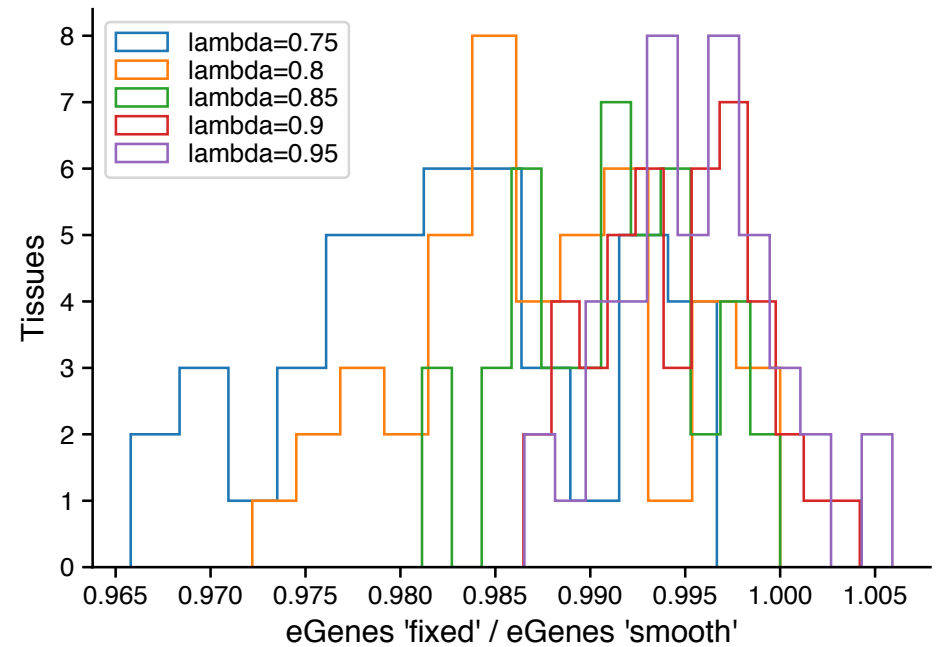
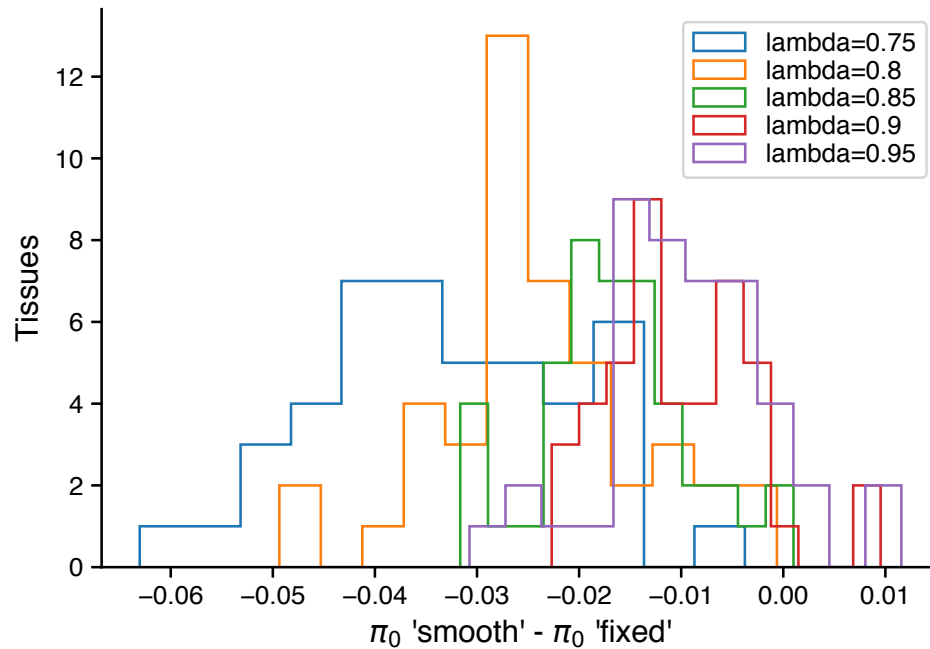
Selection of PEER factors



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

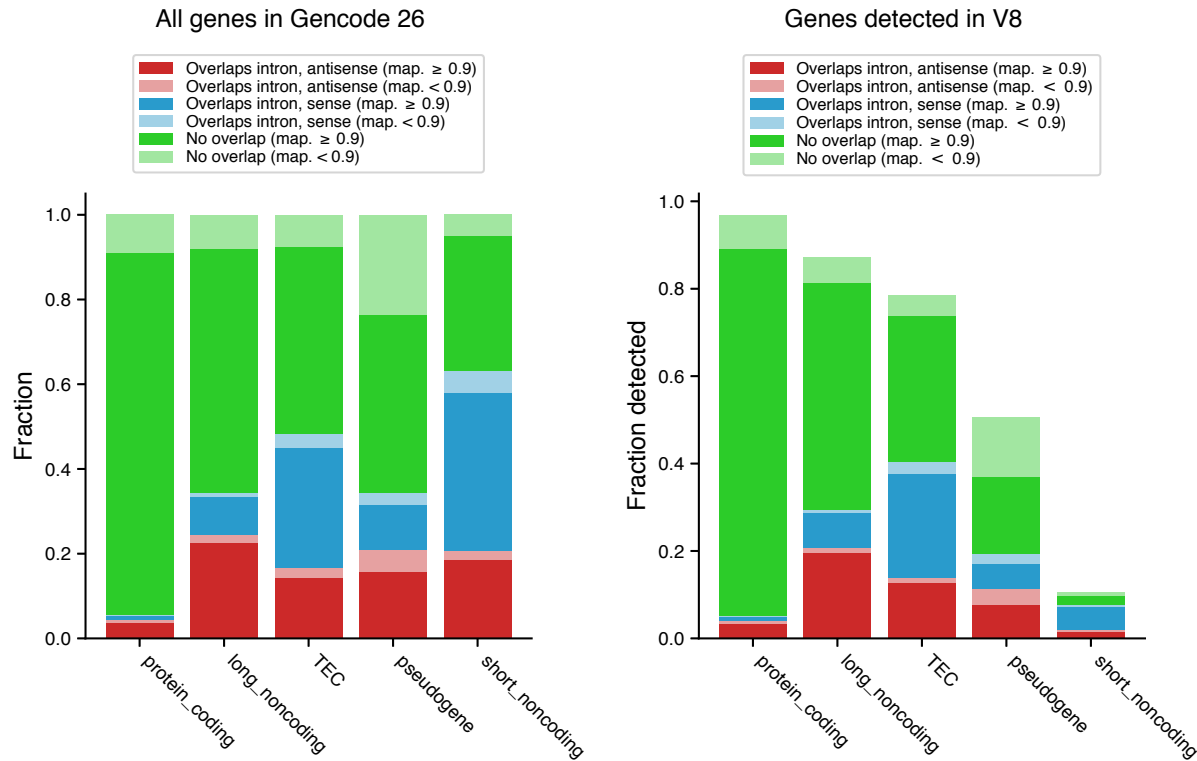
Sample size	PEER factors	Tissues
[0,150)	15	20
[150,250)	30	11
[250,350)	45	8
[350,...)	60	9

FDR: qvalue with fixed 'lambda'



- Comparison on v6p data
- Proposed value for V7/V8 pipeline: $\lambda = 0.85$

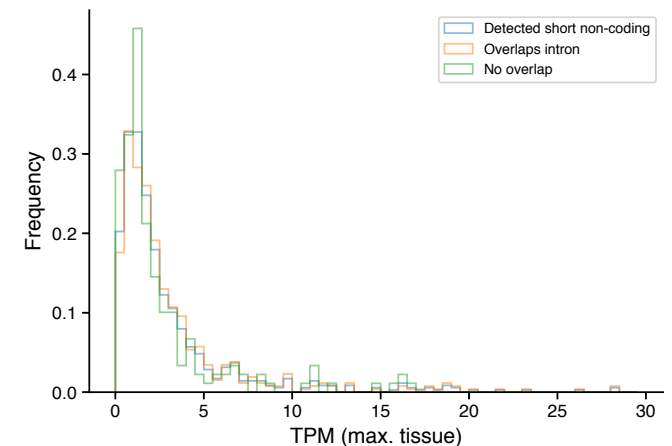
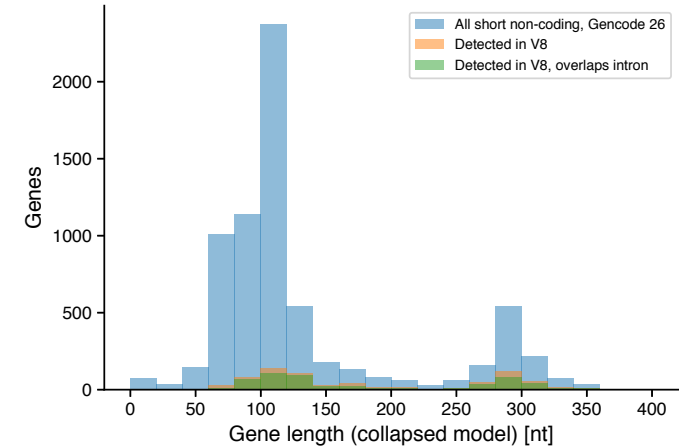
Biotypes for expression and eQTL analyses



TEC: protein-coding gene candidates

Sample size	Gencode 26*	Detected V8
protein_coding	19755	19147
long_noncoding	14100	12310
pseudogene	14407	7299
TEC	1033	812
short_noncoding	6905	724

*Collapsed model



- No unambiguous bias from introns in short non-coding RNAs; other biotypes also affected
- poly(A): degradation pathway for non-coding genes (exosome complex); miRNA precursors