

Estimating Genotype-Tissue Specific Gene Expression Using Hybrid Deep Learning

Jiahong Dong¹, Stephen Brown¹, Kevin Truong^{1,2,*}

¹Edward S. Rogers, Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Rd, Toronto, Ontario M5S 3G4.

²Institute of Biomedical Engineering, University of Toronto, 170 College St, Toronto, Ontario M5S 3E3.

*Corresponding author. E-mail: kevin.truong@utoronto.ca

Supplementary Materials

The receptive field is calculated using:

$$RF_{new} = RF_{old} + (k - 1) \times stride$$

where $k = 8,6,4$ is the kernel size and the stride refers to the step size of the operation. For our CNN, the receptive field expands as follows:

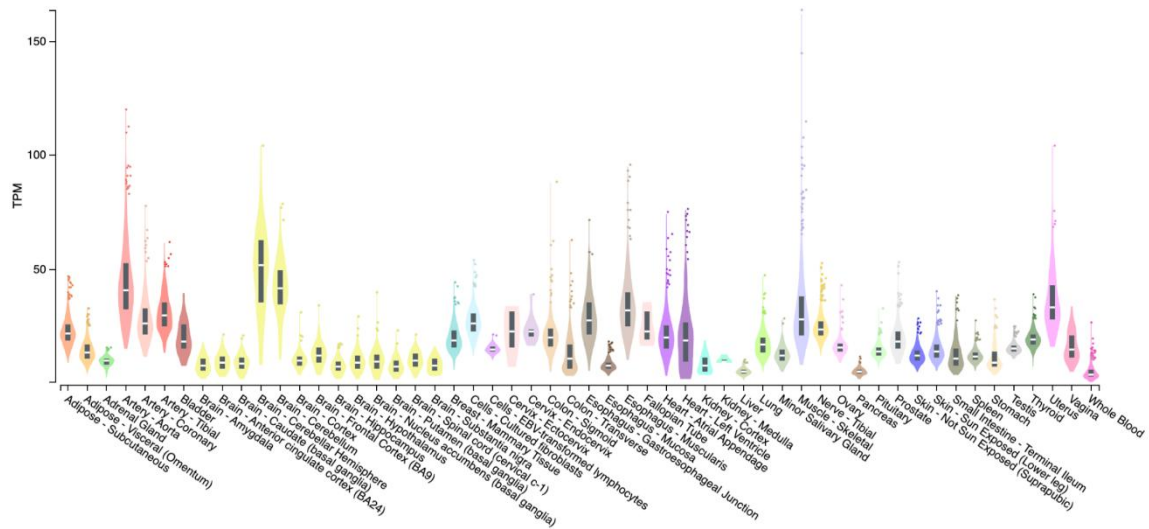
$$RF_{conv1} = 1 + (8 - 1) \times 1 = 8$$

$$RF_{max_pool1} = 8 + (2 - 1) \times 1 = 9$$

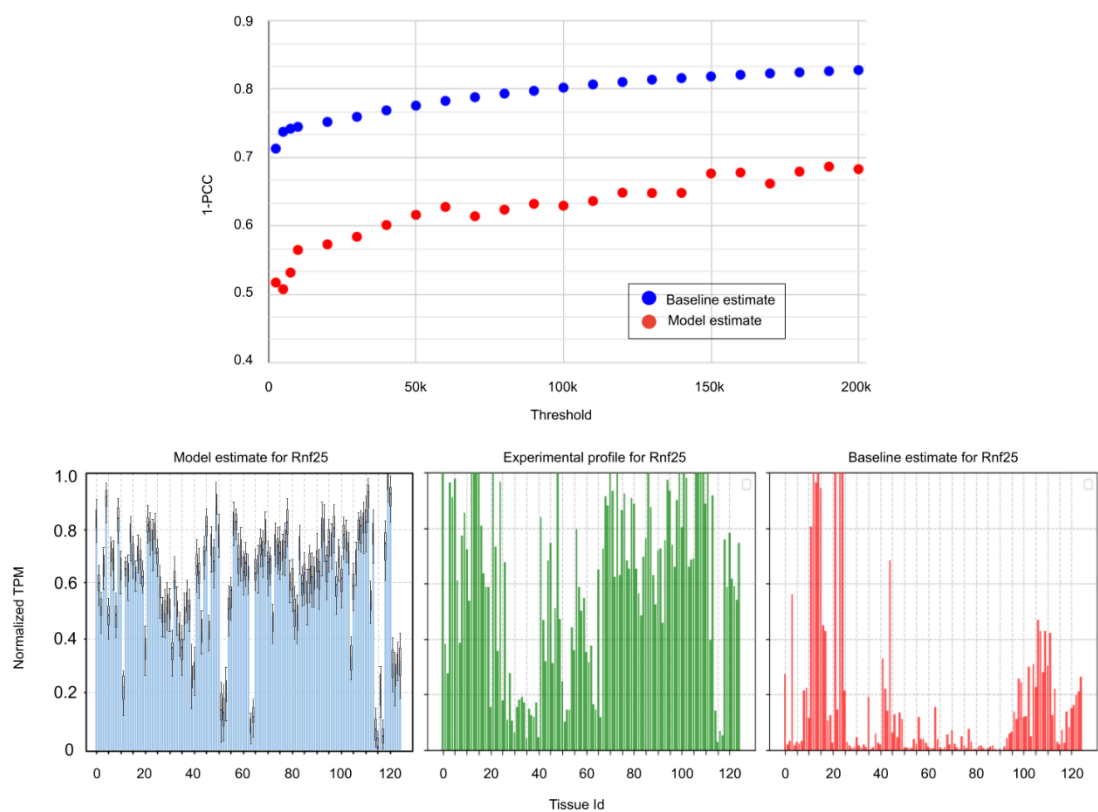
$$RF_{conv2} = 9 + (6 - 1) \times 2 = 19$$

$$RF_{max_pool2} = 19 + (2 - 1) \times 2 = 21$$

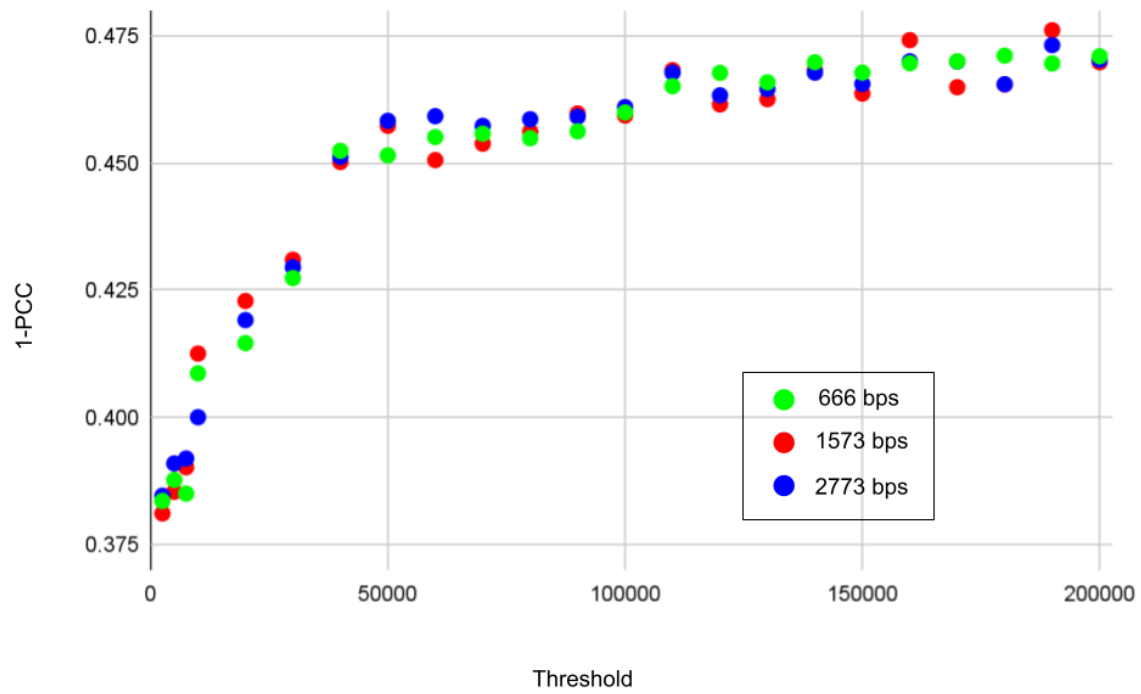
$$RF_{conv3} = 21 + (4 - 1) \times 4 = 33$$



Supplementary Figure S1. An example of what a GTEx genotype-tissue specific expression profile looks like.



Supplementary Figure S2. Extending the hybrid model to estimate tissue-specific gene expression profiles in the mouse genome.



Supplementary Figure S3. Scatter plot showing mean estimation error (1–PCC) for promoter lengths of 666 bp, 1573 bp, and 2773 bp.

Supplementary Table S1

Fold	Chromosomes	# of genes	1-PCC	MSE
1	chr1, chr5, chr8, chr18	3557	0.397412	0.218823
2	chr19, chr7, chr10, chr20, chr21	3517	0.419198	0.218892
3	chr11, chr6, chr4, chr14	3357	0.380512	0.211320
4	chr2, chr12, chr9, chr15	3339	0.406590	0.218101
5	chr17, chr3, chr16, chr22, chr13	3536	0.390476	0.220122

Mean 1-PCC $\pm$ std = 0.3988376 $\pm$ 0.013284

Mean MSE $\pm$ std = 0.217451 $\pm$ 0.003133