Improving tissue-specific expression prediction of sequence-to-function models

Kanav Mittal[1], Ruchir Rastogi[1], Nilah M. Ioannidis[1]

[1]University of California, Berkeley, CA, USA.

1 Introduction

Sequence-to-expression deep learning models, such as Enformer [1], Borzoi [2], and Decima [3], predict gene expression from DNA, typically via multitask training across numerous biological contexts (e.g. cell lines, tissues, disease states). Performance is conventionally assessed by computing cross-gene correlations within each context, a metric on which existing models do quite well. However, these models often struggle to predict context-specific regulatory features [4] or expression, a critical limitation as disease heritability is enriched in cell-type-specific peaks and near genes with cell-type-specific expression [5].

Numerous approaches have been developed to improve context-specific chromatin accessibility prediction, including increasing model capacity [4] and including an additional cross-context loss [6]. However, less effort has been devoted to improving context-specific gene expression prediction. Recently, the Decima model [3] aimed to improve context specificity not only by fine-tuning on pseudobulked single-cell RNA-seq data from additional contexts, but also by introducing an additional cross-context loss term. This cross-context loss, implemented in Decima as a multinomial loss, focuses on predicting a gene's relative expression profile across contexts, rather than its absolute expression in each. However, the Decima study lacked ablations to fully characterize the impact of using this additional loss term.

In this work, we fine-tune Enformer on bulk RNA-seq data from 218 tissues, incorporating an additional cross-tissue Pearson loss. We demonstrate that this strategy improves tissue-specific gene expression predictions. We further show that this improvement can be obtained without modifying the attention maps or even training the vast majority of the network. These findings highlight the utility of cross-context losses and help better characterize their impact on the model.

2 Methods

Data. We fine-tuned Enformer to predict log RPKM values from bulk RNA-seq (as preprocessed by [7]) for 218 tissue samples from GTEx [8], Roadmap Epigenomics [9], and ENCODE [10]. We excluded rRNA genes and genes on sex chromosomes. For each gene, we extracted a 50,048 bp sequence centered at its CAGE-representative TSS. Because Enformer was not trained on chromosome 14, we split our dataset so that all genes in our test set come from that chromosome.

Loss Function. Enformer is trained with a Poisson negative log-likelihood (NLL) loss, which helps it to output values reasonably close to the ground truth. However, the expression of a gene across tissues is much more similar than the expression of different genes in the same tissue. (For example, in our data, the average cross-tissue coefficient of variation (CV) is 3.33, while the average cross-gene CV is 7.79.) This means that a model trained with Poisson NLL loss has a harder time learning the ordering of the expression of a gene across tissues than the ordering of the expression of different genes in the same tissue.

We introduce a hybrid loss function intended to explicitly penalize poor cross-tissue predictions. The hybrid loss is a weighted combination of MSE loss (chosen over Poisson NLL because our targets are continuous log RPKM values) and cross-tissue Pearson loss (chosen for its speed over differentiable Spearman correlation [11]). Pearson loss is defined as 1 minus Pearson correlation. We test and show results for various weightings, i.e. combinations of α and β.

$$L(y, \hat{y}) = \alpha \cdot (1 - \text{Pearson}(y, \hat{y})) + \beta \cdot \text{MSE}(y, \hat{y}) \text{ where } y, \hat{y} \in \mathbb{R}^{218}$$

Model. We fine-tune Enformer using the hybrid loss without freezing any weights. We pass each 50 kb sequence into Enformer, we use attention pooling to aggregate information across the central 10 bins of the model's embeddings before the output heads, and we pass this into a linear layer to predict our log RPKM RNA-seq values across all 218 tissue samples.

Due to limited compute, we did not perform a hyperparameter search, and we instead used hyperparameters largely inspired by [12]. We trained using a learning rate of 1×10^{-4} and a weight decay of 1×10^{-3}. We utilized a learning rate scheduler that linearly increased the learning rate over 1000 steps and then performed cosine annealing to reduce the learning rate. We trained for 50 epochs, with early stopping on the validation loss using a patience of 5. We used a batch size of 2.

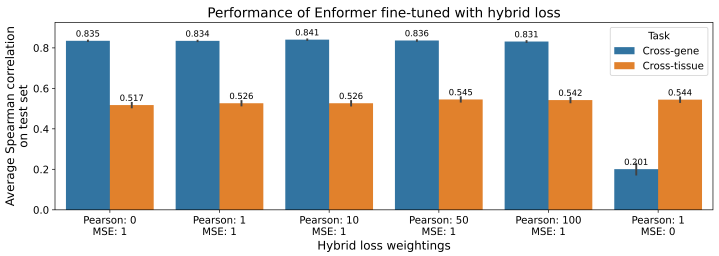

Fig. 1: **Performance of Enformer when fine-tuning with the hybrid loss on cross-gene and cross-tissue tasks**. We fine-tune Enformer with 6 hybrid losses with different Pearson and MSE weightings. For each tissue, we compute the cross-gene Spearman correlation across the 420 genes in the test set and plot the average across 218 tissue samples. For each gene, we compute the cross-tissue Spearman correlation across 218 tissue samples and plot the average across the 420 genes in the test set. Error bars reflect ±1 standard error of the mean.

3 Results

Performance with hybrid loss. We evaluated the cross-gene and cross-tissue performance on the test set for Enformer models fine-tuned with 6 different hybrid loss functions, as seen in Figure 1. As the weighting of the Pearson component (α) increases, the fine-tuned model performs better on the cross-tissue task with minimal change in performance on the cross-gene task. For example, cross-tissue performance increases by 5.4% from $\alpha = 0$ to $\alpha = 50$. However, if MSE loss is weighted at 0, cross-gene performance drops more than over 0.8 to 0.201 Spearman correlation. Cross-tissue Pearson loss alone is not enough for the model to understand which genes are more highly expressed than others in a single tissue.

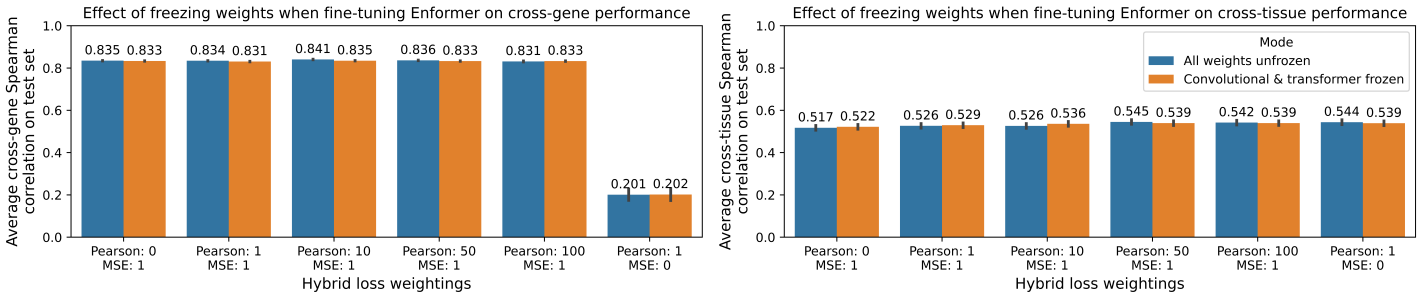

Fig. 2: **Effect of freezing convolutional and transformer modules on cross-gene and cross-tissue performance of fine-tuned Enformer.** We fine-tune Enformer with the same 6 hybrid loss functions as before. Blue bars represent when all weights are unfrozen; orange bars represent when convolutional and transformer modules are frozen. We compute cross-gene and cross-tissue Spearman correlations in a similar way as in Figure 1. Error bars reflect ±1 standard error of the mean.

Freezing. Enformer struggles to capture signal from distal enhancers [13]. Because distal enhancers help modulate tissue-specific expression [14], we hypothesized that incorporating a cross-tissue component into our loss function would help Enformer better capture this distal signal—in particular, via updates to the multi-head attention (MHA) layers [1]. Thus, we hypothesized that if we freeze the weights of the MHA layers when finetuning, the model will not learn to better integrate distal information, and the model's performance should drop.

We compare Enformer fine-tuned without freezing any weights to Enformer fine-tuned while freezing the weights in the convolutional and transformer modules (which includes the MHA layers) and surprisingly find that freezing these modules does not significantly impact the model's cross-gene or cross-tissue performance, as seen in Figure 2. Freezing these modules can reduce the computational cost of fine-tuning Enformer for our tasks and our data, and perhaps other tasks or datasets.

Effect of cross-tissue loss on attention maps. Next, we directly analyzed how our hybrid loss affects the weights in the MHA layers. We compared base Enformer (i.e., not fine-tuned), Enformer fine-tuned with MSE loss, and Enformer fine-tuned with a hybrid loss (we chose a hybrid loss with a Pearson weighting of 50 and MSE weighting of 1 as it performed best on the cross-tissue task in Figure 1). We did not freeze any parameters while finetuning. For each model, we passed in the 50 kb sequence centered at the TSS of each gene in our test set and computed the attention weight between the TSS bin and all other bins, averaging across all layers and all heads (Figure 3a). In addition, we calculated the number of distal bins (defined as more than 50 bins from the TSS) that are highly attended by the TSS for different attention weight thresholds; our results are reported in Figure 3b.

While the base Enformer was trained on both expression data and regulatory chromatin data, we fine-tuned Enformer using only expression data. Since distal enhancers play a critical role in regulating expression [14], our fine-tuning process should make the model pay more attention to distal bins where enhancers may lie. Indeed, we see that our fine-tuned models have higher attention weights than the base Enformer model at the edges of the 50 kb sequence we test (as seen in Figure 3a). Likewise, there are more highly attended distal bins for our fine-tuned models than the base model (as seen in Figure 3b).

However, finetuning Enformer with MSE loss and finetuning with a hybrid loss produce highly similar attention patterns, especially in the distal regions. The cross-tissue loss component does not help the model pay more attention to distal bins.

4 Conclusion

In this paper, we introduce a hybrid loss that incorporates a cross-tissue Pearson loss component. We find that fine-tuning Enformer with a hybrid loss improves cross-tissue performance—even if we freeze the majority of the network—and does not significantly alter the attention maps compared to fine-tuning with an MSE loss. We hope that future research will explore the impact of cross-tissue losses throughout model training (not just fine-tuning) and on tracks besides gene expression and will design loss functions that better integrate signal from distal elements.

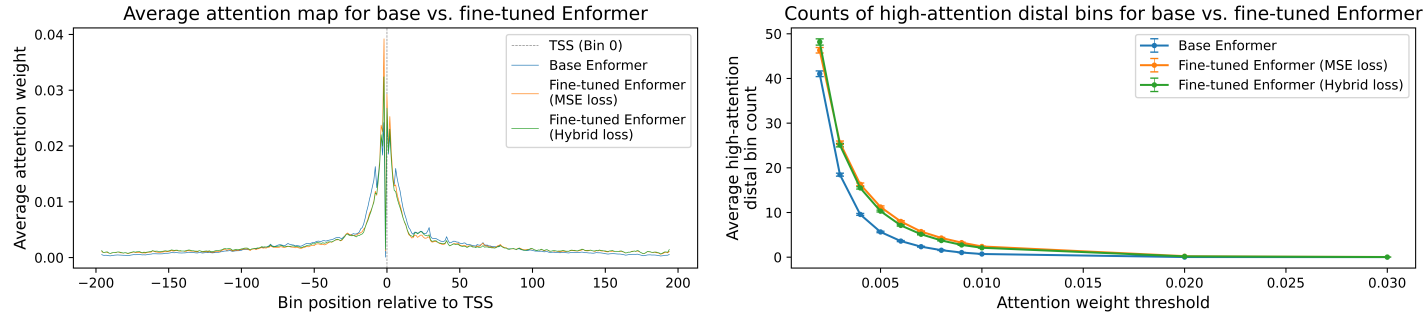

Fig. 3: **Fine-tuning on RNA-seq data alters the attention weights of Enformer, irrespective of the specific loss used.** a) Left. We calculate the attention map using the attention weights where the query is the TSS (i.e., bin 0), and we average the attention maps across all genes in the test set. We plot attention maps for the base Enformer, Enformer fine-tuned with MSE loss, and Enformer fine-tuned with a hybrid loss. Vertical gray dashed line indicates TSS. b) Right. We calculate the number of high-attention distal bins, using different thresholds labeled on the x-axis, for the base Enformer, Enformer fine-tuned with MSE loss, and Enformer fine-tuned with a hybrid loss. We plot the mean number of high-attention distal bins across the genes in our test set with error bars reflecting ±1 standard error of the mean.

References

[1] Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J.R., Grabska-Barwinska, A., Taylor, K.R., Assael, Y., Jumper, J., Kohli, P., Kelley, D.R.: Effective gene expression prediction from sequence by integrating long-range interactions. Nature methods **18**(10), 1196–1203 (2021)

[2] Linder, J., Srivastava, D., Yuan, H., Agarwal, V., Kelley, D.R.: Predicting rna-seq coverage from dna sequence as a unifying model of gene regulation. Nature Genetics, 1–13 (2025)

[3] Lal, A., Karollus, A., Gunsalus, L., Garfield, D., Nair, S., Tseng, A.M., Gordon, M.G., Blischak, J., Geijn, B., Bhangale, T., et al.: Decoding sequence determinants of gene expression in diverse cellular and disease states. bioRxiv, 2024–10 (2024)

[4] Kathail, P., Shuai, R.W., Chung, R., Ye, C.J., Loeb, G.B., Ioannidis, N.M.: Current genomic deep learning models display decreased performance in cell type-specific accessible regions. Genome Biology **25**(1), 202 (2024)

[5] Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., *et al.*: Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature genetics **47**(11), 1228–1235 (2015)

[6] Maslova, A., Ramirez, R.N., Ma, K., Schmutz, H., Wang, C., Fox, C., Ng, B., Benoist, C., Mostafavi, S., Project, I.G.: Deep learning of immune cell differentiation. Proceedings of the National Academy of Sciences **117**(41), 25655–25666 (2020)

[7] Zhou, J., Theesfeld, C.L., Yao, K., Chen, K.M., Wong, A.K., Troyanskaya, O.G.: Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. Nature genetics **50**(8), 1171–1179 (2018)

[8] Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., *et al.*: The genotype-tissue expression (gtex) project. Nature genetics **45**(6), 580–585 (2013)

[9] Kundaje Anshul 1 2 3 Meuleman Wouter 1 2 Ernst Jason 1 2 4 Bilenky Misha 5, R.E.C.I., Chadwick Lisa H. 53, S., Bernstein Bradley E. 2 26 42 Costello Joseph F. 14 Ecker Joseph R. 9 Hirst Martin 5 18 Meissner Alexander 2 6 Milosavljevic Aleksandar 7 Ren Bing 8 13 Stamatoyannopoulos John A. 10 Wang Ting 21 Kellis Manolis 1 2, P.: Integrative analysis of 111 reference human epigenomes. Nature **518**(7539), 317–330 (2015)

[10] Consortium, E.P., *et al.*: An integrated encyclopedia of dna elements in the human genome. Nature **489**(7414), 57 (2012)

[11] Blondel, M., Teboul, O., Berthet, Q., Djolonga, J.: Fast differentiable sorting and ranking. In: International Conference on Machine Learning, pp. 950–959 (2020). PMLR

[12] Rastogi, R., Reddy, A.J., Chung, R., Ioannidis, N.M.: Fine-tuning sequence-to-expression models on personal genome and transcriptome data. bioRxiv (2024)

[13] Karollus, A., Mauermeier, T., Gagneur, J.: Current sequence-based models capture gene expression determinants in promoters but mostly ignore distal enhancers. Genome Biology **24**(1), 56 (2023)

[14] Heinz, S., Romanoski, C.E., Benner, C., Glass, C.K.: The selection and function of cell type-specific enhancers. Nature reviews Molecular cell biology **16**(3), 144–154 (2015)