Attention on Genes

Unveiling Key Genes For Cell-state Predictions of the Geneformer Model by Inspecting the Attention Weights

M. A. Trützschler von Falkenstein

1. Introduction into Geneformer

Machine learning models within the medical sector are faced with the ***need for diverse datasets for training***. ***Transfer learning*** emerges as a powerful tool ***within limited-data conditions***. ***Geneformer*** |1| is a ***transformer***, a model that uses attention, pretrained on *Genecorpus-30M*. During pretraining, ***Geneformer learns which genes are correlated***. This knowledge is retained within the fine-tuned model. By adding a fine-tuning layer, the model may be used for ***cell-state predictions*** and potentially also to ***discover target genes through perturbation experiments***. Through the attention mechanism ***the model learns to pay attention to the most important parts of the input***. This research aims to ***quantify the attention shift*** that Geneformer undergoes during pretraining, and ***discover key genes*** for cell-state predictions by monitoring which genes receive the most attention.

2. Background

The model consists of ***6 layers*** of transformer encoder units, which each have ***four attention heads*** that process the data in parallel. A single-cell transcriptome is encoded as a ***rank-value encoding***. Through the ***attention mechanism***, the model captures the ***correlation between genes***. After normalisation, the output of the attention layer passes through a ***FFN*** which adds non-linearity |2|.

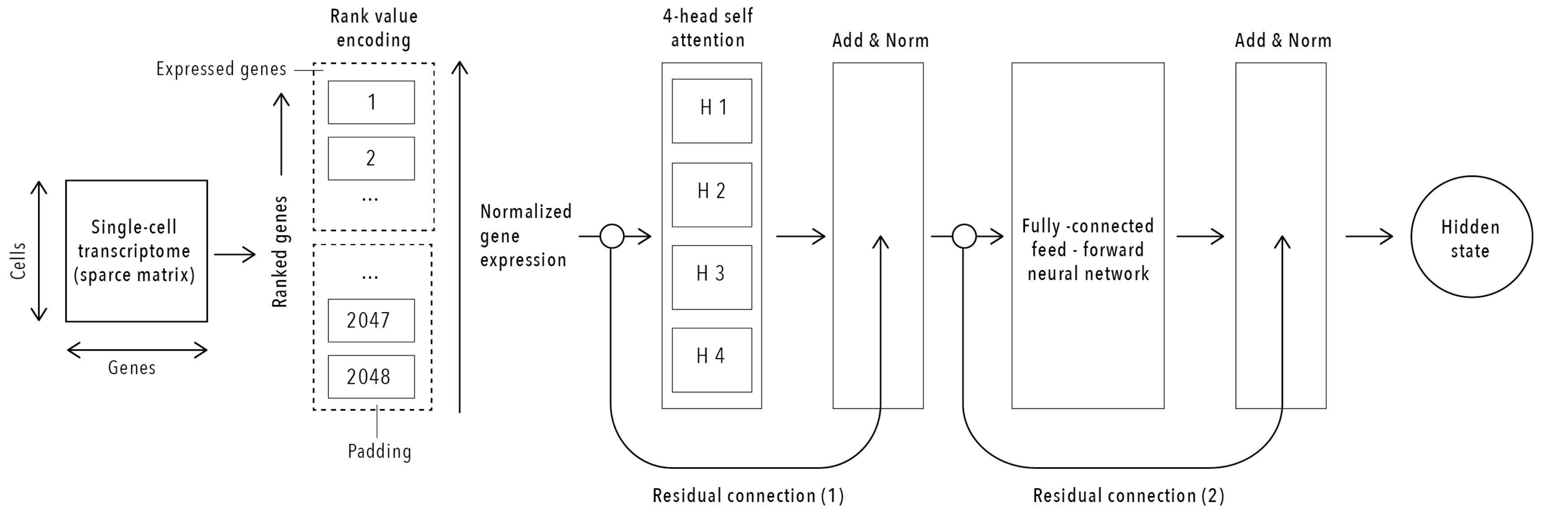


Figure 1. The transformer encoder unit.

3. Preprocessing and Fine-tuning

The model was ***fine-tuned*** with the sciplex-2 |3| dataset which contains ***single-cell transcriptomic data of cancer cells***. We used data of cells which were subjected to the nutlin-3a drug.

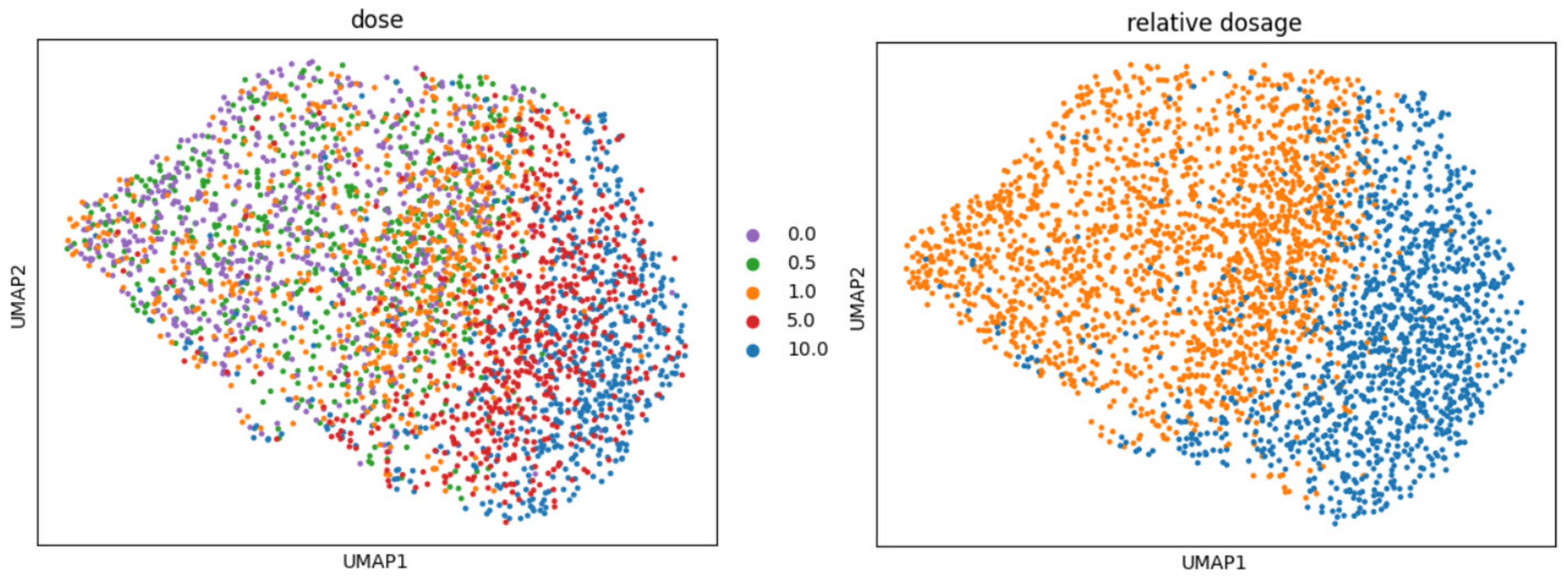
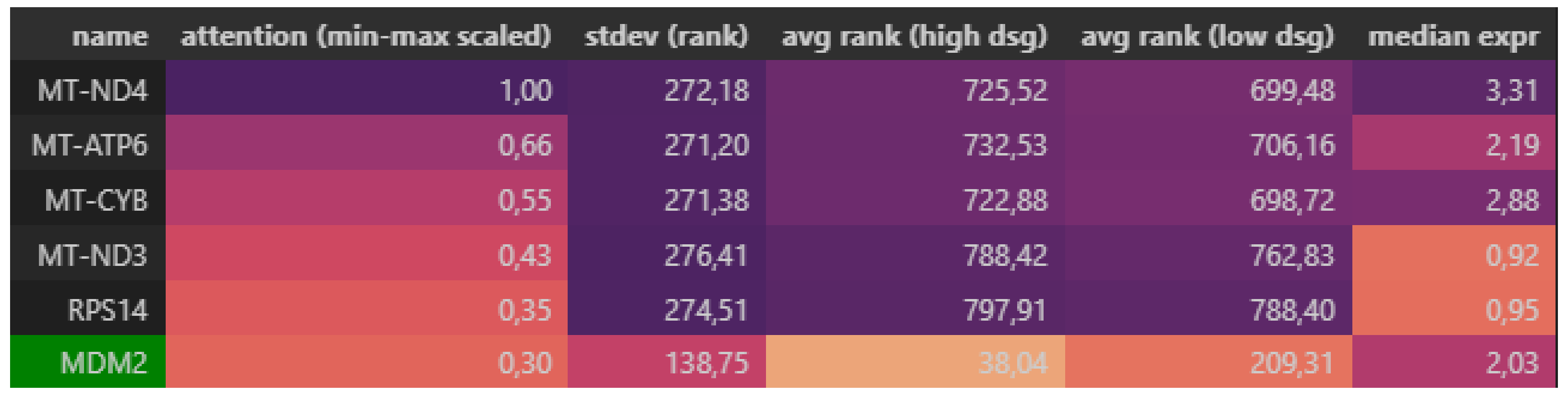


Figure 2. Umap into two-dimensional space of the data after preprocessing.

The ***Wilcoxon test*** was applied to identify ***differntially expressed genes*** across the classes. The training objective was to identify the dosage which was applied, in order to assess wether geneformer would focus on ***similar genes***. Geneformer reached ***low predictive accuracy***, but it was able to ***discern between high and low dosage***, with ***high predictive accuracy***.

4. Attention-weights Analysis

The attention weights of the model were analysed as a ***percentage change*** (fig. 4) and ***mapped back to the genes*** (fig 3), which showed that while Geneformer is able to find key genes within the dataset, there still seems to be noise, as the top genes are not differential in gene expression across the classes.



name	attention (min-max scaled)	stdev (rank)	avg rank (high dsg)	avg rank (low dsg)	median expr
MT-ND4	1,00	272,18	725,52	699,48	3,31
MT-ATP6	0,66	271,20	732,53	706,16	2,19
MT-CYB	0,55	271,38	722,88	698,72	2,88
MT-ND3	0,43	276,41	788,42	762,83	0,92
RPS14	0,35	274,51	797,91	788,40	0,95
MDM2	0,30	138,75	38,04	209,31	2,03

Figure 3. Top 6 genes which receive the most attention by geneformer. MDM2, the target gene of Nutlin-3a was ranked sixth.

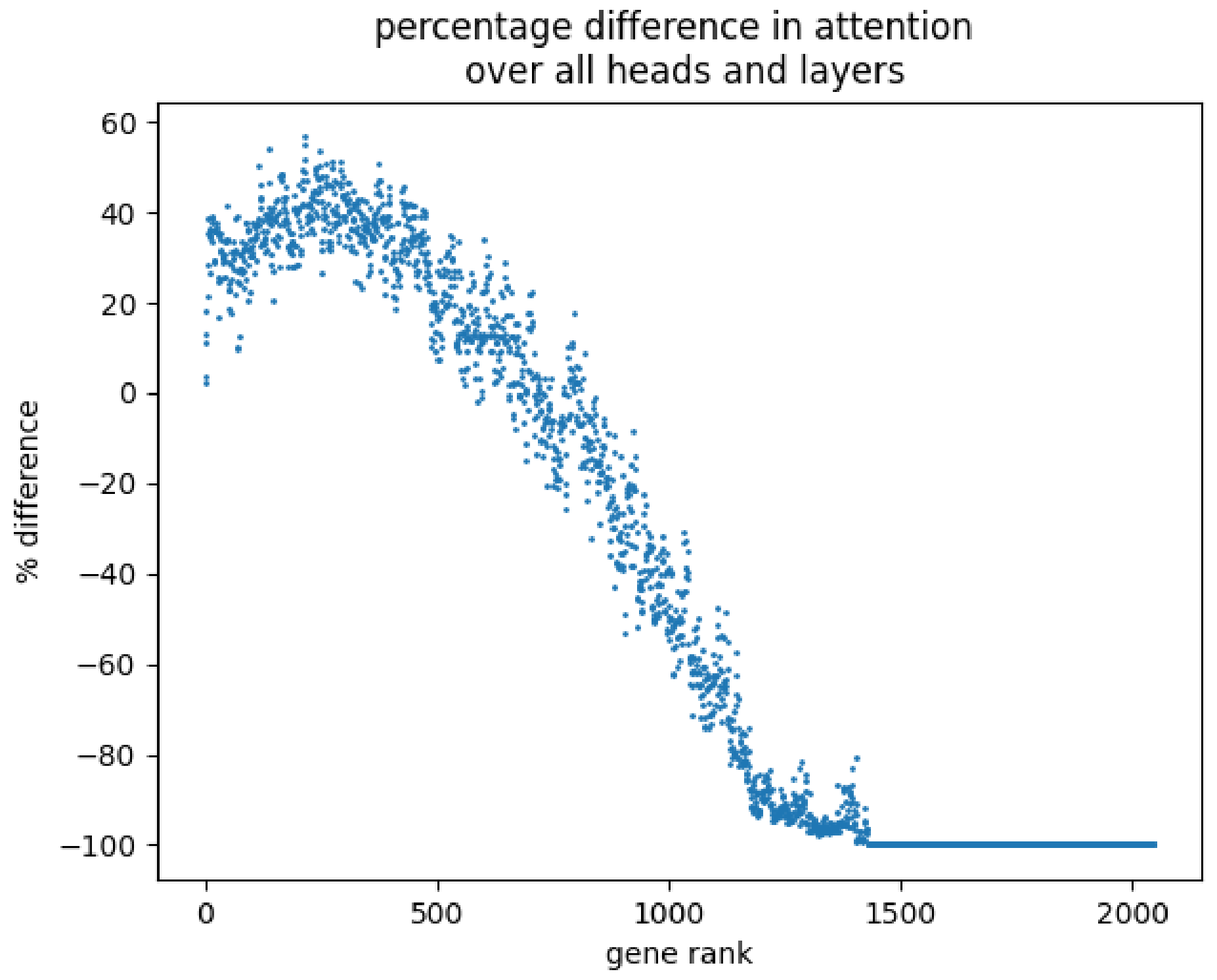


Figure 4. Plot showing the percentage difference of the trained and pretrained attention heads. The largest change can be found around rank 200, which is were the embedding of MDM2 resides for low dosages of nutlin-3a. This shows Geneformers aptitude to find a key gene within the dataset.

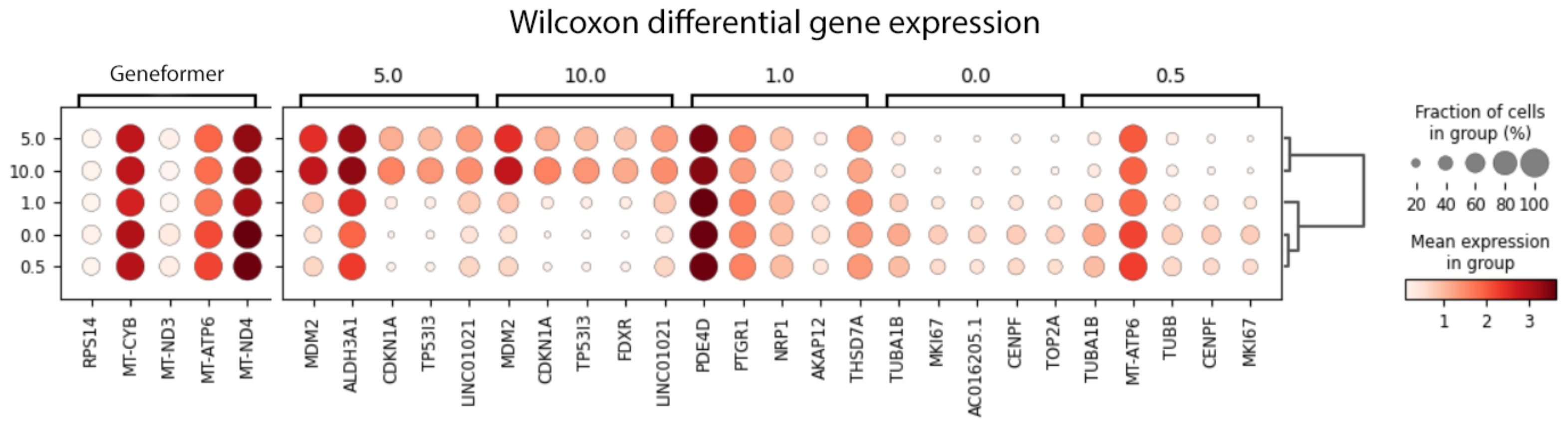
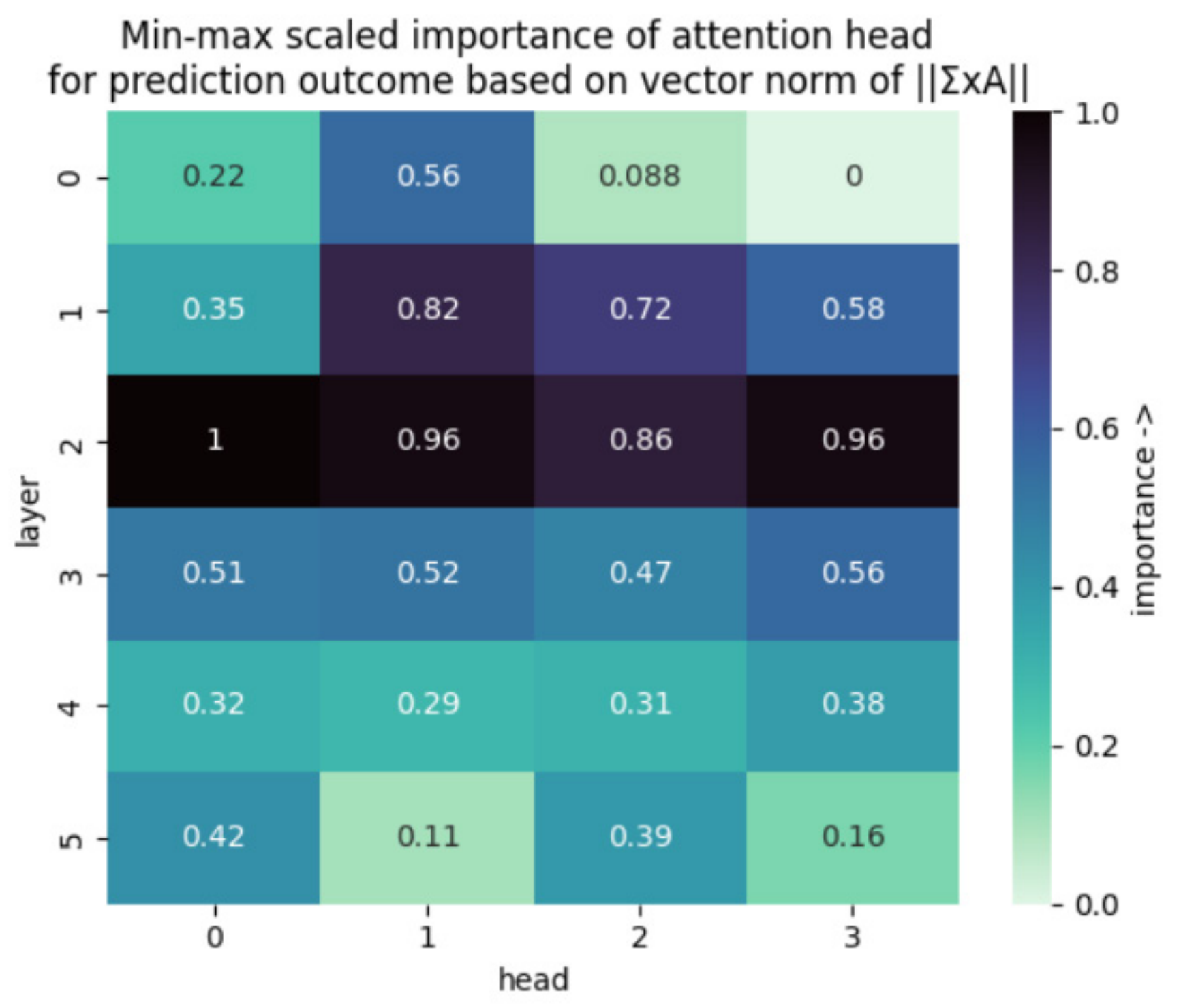


Figure 5. Differential gene expression of the four classes as estimated by the Wilcoxon test. The most attended genes and their gene expression are depicted on the left.

Figure 6. The significance of the changes within each individual head were quantified by measuring || ΣxA||, where x was the input vector. This showed that the changes could have a large impact on the prediction outcome. |4|



5. Discussion

While ***significant changes in the attention weights*** were observed, this ***did not improve the overall accuracy*** of the model, possibly this means that the classes within the high or low dosage cluster are too similar.

There also seems to be noise in the attention: the t***op genes are not differentially expressed across the classes***. Possibly Geneformer is still subject to some ***batch-effect: the amount of genes expressed***.

6. Conclusion and Future Work

Geneformer was able to find key genes within the data to seperate some of the classes. However, there still seems to be ***noise which could be destructive when performing perturbation experiments***. Future work should focus on ***mitigating the effect of the amount of genes expressed***.

References

|1| C. V. Theodoris, L. Xiao, A. Chopra, M. D. Chaffin, Z. R. Al Sayed, M. C. Hill, H. Man- tineo, E. M. Brydon, Z. Zeng, X. S. Liu, and P. T. Ellinor, “Transfer learning enables predictions in network biology,” Nature, vol. 618, pp. 616–624, Jun 2023. |2| J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” 2019. |3| V. Alexander, “scrnaseq exposed to multiple compounds.” https://www.kaggle.com/datasets/alexandervc/scrnaseq-exposed-to-multiple-compounds, May 2021. Accessed: 08-05-2024. |4| G. Kobayashi, T. Kuribayashi, S. Yokoi, and K. Inui, “Attention is not only a weight: Analyzing transformers with vector norms,” 2020.