

Transformers Enhance the Predictive Power of Network Medicine

Jonah Spector^{1,2,*}, Andrés Aldana^{2,*}, Michael Sebek², Joseph Ehlert²,
Christian DeFrondeville³, Susan Dina Ghiassian⁶, and Albert-László
Barabási^{1,2,4,5}

¹Department of Physics, Northeastern University

²Network Science Institute, Northeastern University

³College of Engineering, Northeastern University

⁴Department of Medicine, Brigham and Women's Hospital, Harvard Medical
School

⁵Department of Network and Data Science, Central European University

⁶Scipher Medicine Corporation

* Authors contributed equally to this work

January 27, 2025

Abstract

Background

Self-attention mechanisms and token embeddings behind transformers allow the extraction of complex patterns from large datasets, and enhance the predictive power over traditional machine learning models. Yet, being trained to make predictions about individual cells or genes, it is not clear if transformers can learn the inherent interaction patterns between genes, ultimately responsible for their mechanism of action. We use Geneformer, pretrained on single-cell transcriptomes, to ask if transformers can implicitly capture molecular dependencies, including protein-protein interactions (PPIs), allowing us to explore the use of transformers to improve network medicine tasks such as disease gene identification and drug repurposing.

Methods

We extracted the cosine similarity of gene embeddings and the attention weights contained in Geneformer, allowing us to test if these weights capture experimentally validated protein interactions. Using dilated cardiomyopathy as a case study, we evaluated the effectiveness of the resulting weighted networks in disease module detection and drug repurposing.

Results

We found that Geneformer displays awareness of experimentally documented Protein-Protein Interactions, exhibiting higher cosine similarity and attention weights for gene pairs with physical interactions. Weighting PPI networks with the cosine similarity and attention weights improved the detection of disease-associated genes and the accuracy of drug repurposing predictions for dilated cardiomyopathy, surpassing the accuracy of unweighted networks. Finally, we find that combining attention weights and cosine similarities with ranking methods enhances drug candidate prioritization for drug repurposing.

Conclusions

We find that transformers, by implicitly learning the interactions between genes, offer a promising pathway for advancing medicine and drug discovery when integrated with the graph theoretic algorithms used in network medicine.

Introduction and Background

Transformers [1], a class of deep learning models designed for natural language processing (NLP), are revolutionizing biological research by capturing complex patterns in data through self-attention mechanisms. The transfer learning approach behind transformers has been applied to improve protein structure prediction [2], genetic sequence analysis [3, 4], and molecular interaction prediction [5], and to extract knowledge from large datasets [3, 6, 7].

In NLP, models are given inputs in the form of tokens, which are vector representations of the words in the sentence. In Geneformer, a transcriptome-based transformer pretrained with 30 million single-cell transcriptomes, genes are the individual tokens. While in language the order of the tokens is predefined by the grammar, Geneformer uses Rank-Value Encoding (RVE) to order genes by normalized expression defining the ‘grammar’ of the cell [7]. The pretraining gives Geneformer a versatile context awareness that can be harnessed through posterior task-specific training phases (fine-tuning).

Machine learning models are typically trained to predict individual tokens or classify groups of tokens, ignoring the inherent interactions between them. In biology, however, the interactions between genes are responsible for the mechanism of action of the individual genes and proteins. In this study, we ask to what degree transformers capture the known molecular dependencies between proteins, allowing us to explore the utility of transformers to enhance the tool set of network medicine. Indeed, the activity pattern of a gene is the result of multiple network-based dependencies, that are mapped and studied by network medicine. By conceptualizing biological processes as networks of genes, proteins, and RNAs [8, 9], network medicine offers a series of graph theoretic tools to predict drug repurposing opportunities or drug responses [10, 11, 12, 13]. Here we show that Geneformer, a six-layer transformer trained with single-cell transcriptomes, develops an awareness of experimentally documented Protein-Protein Interactions (PPI). This awareness allows us to adapt the transformer to perform standard network medicine tasks, finding that it can improve the performance of disease gene discovery and drug repurposing. Overall, our research reflects a convergence of network medicine and deep learning, offering a path towards a better un-

derstanding of complex diseases by providing predictive models of biological interactions and disease mechanisms that can help develop more effective treatments.

Results

The Human Interactome

Network medicine unveils disease mechanisms by mapping out the physical interactions between cellular components (genes, proteins, RNAs). Geneformer, trained on expression data, lacks direct information on these physical interactions. However, the experimentally observed expression pattern of each gene is determined by regulatory processes driven by the underlying physical interactions. Therefore, we hypothesize that information about physical interactions is implicitly encoded in Geneformer. To test this hypothesis, we examined whether an experimentally validated network of protein interactions can be extracted by exploiting the relationships between their corresponding internal representations within Geneformer. This ground truth interactome is represented as a network of 18,260 protein-coding genes (nodes) and 520,009 experimentally validated physical interactions (edges) [14]. By filtering for the 18,062 genes common to Geneformer’s vocabulary and the interactome, we derive a PPI network with 514,674 interactions. This network serves as the foundation for the analyses presented in this paper.

Embeddings Predict Human Interactome

Geneformer encodes genes in 256-dimensional vectors known as embeddings [7]. These vectors are initially oriented randomly and change direction during pretraining, updated as

they progress through each layer of the model (see Fig. 1a-c). After pretraining, the model develops a vocabulary of 25,424 genes, each corresponding to a 256 dimensional embedding that represents a single protein coding gene or miRNA. Cosine similarity of these vectors was previously used for *in silico* network analysis of congenital heart disease [7], prompting us to hypothesize that cosine similarity may reflect the underlying physical interactions between gene pairs (Figure 1c).

We run 10,000 healthy single-cell transcriptomes from the 30M corpus through the pre-trained model. For each sample, we extract the embeddings from the second to last layer and calculate the cosine similarity matrix C that records the maximum cosine similarity between each gene pair across the 10,000 samples (Methods). If the embeddings are informed by the physical interactions between the genes, we would expect that gene pairs with experimentally validated binding interactions would have larger cosine similarity. In line with this hypothesis, we find that the mean value 0.219 ± 0.001 of the cosine similarities (CSs) corresponding to experimentally validated PPIs is significantly different from the background (nonexistent interactions) with the mean value 0.179 ± 0.001 , a difference whose statistical significance is verified by the Kolmogorov-Smirnov test (KS-test, $p = 0$) (Figure 2a).

Attention Weights Also Predict Interactions

“Attention is All You Need” stated the publication that introduced the transformer architecture [1]. Indeed, the attention mechanism helps the transformer dynamically focus on different parts of the input data by weighting the significance of each input element differently, determining how much “attention” each part needs.

Each of Geneformer’s six layers contains four attention heads (Figure 1e-f) [7]. Each head is a square matrix whose dimensions match the length of the input transcriptome, capturing the “attention” each gene assigns to every other gene within that transcriptome. We created an aggregated attention matrix A from the same 10,000 single-cell samples used to generate the cosine similarity matrix C (Methods). We find that the PPI edges have a higher mean attention weight (AW) ($2.25 \pm 0.04 \times 10^{-3}$) than the background ($1.62 \pm 0.02 \times 10^{-3}$), whose significance is again verified by a KS-test ($p = 0$).

Degree Dependence of Cosine Similarity and Attention Weights

Given the central role the node degree k (number of physical interactions with other nodes) has in network medicine, we explored the relationship between the total attention and cosine similarity associated with a gene and the degree k of that gene in the network. By summing along the rows of the cosine similarity matrix C , we calculated the total cosine similarity C_{tot} associated to each gene. Similarly, we calculated the gene’s total attention A_{tot} . By fitting the logistic regression,

$$C_{tot} = \beta_C \ln(k), A_{tot} = \beta_A \ln(k), \quad (1)$$

we found that the total cosine similarity and attention of each gene have a strong positive correlation with the gene’s degree k . C_{tot} shows a dependence on $\ln(k)$ with an R^2 of 0.27, while A_{tot} has an R^2 of 0.09 with $\ln(k)$, indicating that the natural log of the degree explains 27% of the variance in C_{tot} and 9% in A_{tot} (Figure 2d).

Taken together, we find that both cosine similarities and the attention weights show an

awareness of the underlying physical interactions within the cell. While the resulting predictive power for protein interactions is relatively modest (Figure S4), the higher cosine similarity and attention weights associated with these interactions prompt us to ask if the transformer technique may be able to contribute to the traditional network medicine tasks, like disease module identification and drug repurposing, questions that we explore next.

Pretrained Model Pays Attention to Disease Modules

A foundational result in network medicine is that the disease genes associated with a specific disease or phenotype are not randomly distributed in the interactome, but tend to be located in a specific network neighborhood, forming a connected subgraph known as the disease module [15, 16]. We hypothesize that attention weights and embeddings, by revealing relationships between disease genes, highlight the specific network neighborhood within the interactome, helping us identify the disease module.

To test this hypothesis, we start from a gene-disease association (GDA) list consisting of 110 disease genes associated with dilated cardiomyopathy [10]. Of the 110 proteins, 74 form a connected subgraph in the PPI, corresponding to a disease module connected by 174 physical interactions. We take expression data from 10,000 dilated cardiomyopathy (DCM) cardiomyocytes [17] and run them through the model to create a new aggregated cosine similarity matrix C' . We categorize the obtained CSs between gene pairs corresponding to the 174 edges within the disease module and those corresponding to all physical interactions in the PPI. We find that the CSs within the disease module have a mean weight of 0.313 ± 0.002 , slightly larger than 0.253 ± 0.001 , observed for all PPI interactions (Figure 2e), a difference

significant by KS-test ($p = 6 \times 10^{-5}$). Similarly, we find that the attention weights within the disease module have a mean weight of $9.29 \pm 0.19 \times 10^{-3}$, which is 3 times larger than the mean weight of all PPI interactions ($2.78 \pm 0.003 \times 10^{-3}$, Figure 2f, statistically significant by the KS-test, $p = 2 \times 10^{-14}$).

Improving Disease Module Detection

Next, we ask if we can use Geneformer to improve disease gene predictions [18], a central task for network medicine, particularly pertinent for under-studied and rare diseases with a limited number of associated genes. Current network medicine algorithms for identifying the disease module aim to construct a connected component of disease genes in the network, relying on the unweighted protein-protein interactome (PPI) [19]. While a weighted interactome offers myriad potential improvements in network medicine [8, 10, 19], weighting techniques can introduce batch effects [20, 21], methodological differences [22], or low confidence edges [23].

Here, we hypothesize that an interactome weighted with either Geneformer AWs or CSs will allow us to improve network medicine tasks such as disease module discovery. The leading algorithm for disease module detection, random walk with restart (RWR) [24, 25, 26], begins with a set of known disease genes acting as seeds and then ranks new candidate genes in the interactome by calculating the steady-state probability that a walker, starting from any seed gene, will reach each candidate gene. The probability that the walker reaches a candidate gene defines the probability that the gene should be a part of the disease module.

We applied the aggregated Geneformer AWs from the pretrained model as weights to the

PPI edges. To be specific, for each link (i, j) in the PPI, the interaction strength is defined as $W_{ij} = \max(A_{ij}, A_{ji})$, where A_{ij} represents the aggregated attention weight between genes i and j , derived from the transcriptomes of 10,000 cardiomyocytes from dilated cardiomyopathy (DCM) samples. We repeated this process using AWs from the Geneformer model fine-tuned on cardiomyopathy. Finally, we repeat the entire process using cosine similarities (CSs), for a total of 5 networks (unweighted, pretrained embedding weighted, pretrained attention weighted, fine-tuned embedding weighted, and fine-tuned attention weighted.) We divided the set of 110 known cardiomyopathy genes into seed (80%) and recovery (20%) sets. The seed set served as the starting point for RWR, while the recovery set is used to measure the predicted rank of the known disease genes.

We find that all 4 weighted networks perform at least as well as the unweighted network, and the attention weighted networks offer a notable improvement in disease module detection. Both the pretrained and fine-tuned attention weighted networks generate an AUROC of 0.76 ± 0.02 , which is within the standard error in the mean of the unweighted network (AUROC = 0.74 ± 0.02 , Figure 3a). For the purposes of disease module detection though we are mainly interested in the top candidates. Therefore, we evaluated the precision and cumulative true positives in the top 100 candidates (Figure 3c,d), since these metrics can represent the efficacy of the algorithm more accurately than the AUROC. We find that in the top 100 candidates, the attention weighted networks record nearly twice the cumulative true positives compared to the unweighted network(Figure 3d). The fine-tuned AWs in particular generate an initial precision of nearly 60%, a notable improvement on the unweighted network, which records an initial precision of 0 (Figure 3b,c).

Attention Weights Improve Drug Repurposing

Drug repurposing, the process of identifying new therapeutic uses for FDA-approved drugs, has become increasingly important due to its potential to reduce the time and cost associated with drug development. While some approaches leverage computational methods and extensive biological databases to predict drug-disease associations [27, 28, 29], network medicine based repurposing, using either graph theory tools [11, 13], graph neural networks [30], or a combination of these two methods [10], was able to identify promising experimentally tested drug candidates, along with the specific disease mechanism responsible for the pathogenesis [31].

We evaluated whether the AWs and CSs between Geneformer embeddings could improve drug repurposing accuracy. As a baseline, we used the dilated cardiomyopathy (DCM) fine-tuned Geneformer model [7]. We extracted the AWs and CSs from the model and used them to assign weights to the PPI network. Subsequently, we calculated the network-based proximity scores [32] between the DCM disease genes and the targets of 618 drugs, including 171 DCM-related (Positive) and 447 unrelated (Negative) drugs extracted from DrugBank (see Methods) [33]. We ranked the drug list according to each drug's proximity score and tracked the number of positive and negative drugs predicted by the DCM-Geneformer weighted networks.

We repeated this procedure by extracting the AWs and CSs values from all six transformer layers of Geneformer. As a control, we compared the results from these weighted networks with those from the unweighted PPI network. We find that the PPI network weighted with AWs from layer 5 and CSs from layer 0 enhanced the overall prediction capabilities for identifying

DCM-related prescribed drugs, achieving an AUROC of 0.72 for AWs and 0.71 for CSs. This performance considerably surpassed the predictive power of the unweighted PPI network (AUROC = 0.60), as well the predictive power extracted from the other layers (Figures 4a,d and g).

The aim of drug repurposing is to prioritize the available drugs, so that experimental efforts can be focused on the highest-ranked compounds. For this reason, we calculated the number of true positive drugs among the top 100 predicted drugs (precision), and the fraction of all possible positive drugs among the top 100 predicted drugs (recall). For precision, the PPI network weighted with AWs from layer 5 (AUPC = 0.43) and CSs from layer 0 (AUPC = 0.45) outperformed both the unweighted control network (AUPC = 0.40) and other layers. Similarly, for recall, the PPI network weighted with AWs from layer 5 (AURC = 0.43) and CSs from layer 0 (AURC = 0.50) also surpassed the control network (AURC = 0.39) and other layers (Figure 4b,e,h and 4c,f,i).

Since AWs from layer 5 and CSs from layer 0 demonstrated the best overall performance from each set, we further explored the relationship between their drug proximity scores. As shown in Figure 5a, the scores derived from AWs and CSs exhibit a strong positive correlation (Pearson correlation coefficient = 0.77, $R^2 = 0.59$), indicating that both methods capture similar network-based relationships between drug targets and DCM genes. However, disagreement between the two scores suggests that each method captures different aspects of the relationships, indicating their potential complementarity for improving prediction accuracy when combined. To assess this, we applied three ranking methods: Borda Count [34], Dowdall Count [35], and CRank [30], to integrate the scores provided by AWs and CSs. These

combined ranking methods yielded a marginally higher predictive performance than AW or CSs scores alone, achieving AUROC values of 0.74 for Dowdall Count and CRank (Figure 5b). Additionally, precision and recall analyses of the top 100 ranked drugs revealed that CRank is the most effective combination method, outperforming Borda and Dowdall counts, as well as the individual AW and CSs scores and the unweighted PPI network. This suggests that integrating multiple layers of information from Geneformer enhances the accuracy of drug repurposing predictions (Figure 5c,d).

Discussion

In this study, we demonstrated that transformer-based models, specifically Geneformer, have the potential to enhance network medicine tasks, such as disease gene identification and drug repurposing. Our results illustrate how transformer embeddings and attention weights can be leveraged to capture and reinforce biological interactions within protein-protein interaction (PPI) networks. While network medicine traditionally relies on unweighted or experimentally weighted networks [36, 37, 38], we show that attention-weighted networks outperform traditional methods.

A key finding of this study is the superior performance of attention weights compared to cosine similarity in disease gene identification and marginally in drug repurposing. While both mechanisms capture relationships between gene pairs, we find that attention weights outperform cosine similarity in accuracy and applicability.

One possible explanation for this difference is that attention weights are dynamic and context-

dependent, in contrast with the more static, global relationships between gene embeddings offered by cosine similarity [1]. This dynamic nature of attention weights makes them particularly well-suited for tasks like disease module detection and drug repurposing, where functional relationships between genes are context-specific.

We also find that fine-tuning Geneformer on disease-specific data, such as cardiomyopathy transcriptomes, further enhances its ability to identify disease-relevant network regions. In our analysis of the DCM disease module, the fine-tuned model's cosine similarity values and attention weights showed greater separation between PPI and disease module edges than the pretrained model (Figure S5). This increased focus on the disease module, alongside with a down-prioritization of unrelated interactions, highlights the value of fine-tuning transformer models for specific diseases.

One of the most impactful applications of Geneformer's attention weights and cosine similarity is in drug repurposing. Our results show that AWs from layer 5 and CSs from layer 0 provided the highest overall accuracy, outperforming both the unweighted PPI network and other layers. Moreover, integrating AWs and CSs from the Geneformer model slightly improves drug repurposing predictions for dilated cardiomyopathy (DCM). While AWs and CSs scores are strongly correlated, their differences suggest they capture complementary information, with combined ranking methods like CRank [30] further enhancing predictive power. These findings highlight the value of integrating transformer models with network medicine approaches, demonstrating that attention weights and cosine similarity together can improve the identification of therapeutic candidates for complex diseases.

While Geneformer's performance in network medicine tasks is promising, limitations remain.

Though improved, the AUROC values indicate room for further optimization, particularly in integrating additional biological data or refining the transformer model architecture. Additionally, exploring other deep learning architectures or expanding the training datasets could help improve the accuracy of disease gene identification and drug repurposing predictions. Further research is needed to determine how these transformer-based models can be applied to more diverse biological contexts, including signaling pathways or metabolic networks.

This convergence of AI and network medicine offers unprecedented opportunities for understanding complex diseases. Using transformers in network biology could significantly accelerate drug discovery efforts, reduce the cost and time of drug development, and improve patient outcomes through more targeted therapeutic interventions.

Methods

To evaluate Geneformer's ability to capture biological information, we compared embeddings and attention weights from a pretrained and fine-tuned model with a protein-protein interaction (PPI) network and curated disease-gene association lists (SI 1.1). We sampled 10,000 single-cell transcriptomes and calculated the aggregated cosine similarity and attention weights to assess the distribution of known PPI and disease module edges (SI 1.2). We further used the cosine similarity and attention weight values to construct weighted PPI networks (see SI 1.3). We used the weighted networks to perform disease module discovery (SI 1.3) and drug repurposing studies (see SI 1.4) in dilated cardiomyopathy.

Acknowledgments

We acknowledge Katarina Petrovich, Giulia Menichetti, and Abel Elekes for valuable discussions, and Christina Theodoris for advice related to the internal workings of Geneformer.

Competing interests

A.L.B. is a scientific founder of Scipher Medicine, Inc., which applies network medicine strategies to personalized drug selection. All other authors have no competing interests.

Materials & correspondence

Correspondence and requests for materials should be addressed to A.L.B.

Data and code availability

All raw datasets used in this manuscript are directly cited. The source code is available at <https://github.com/Barabasi-Lab/Geneformer-NetworkMedicine/tree/main>, along with instructions for recreating and replicating the analyses in the paper. The aggregated weight matrices used to create the figures are available at https://huggingface.co/datasets/jspector792/Geneformer_NetworkMedicine.

References

1. Vaswani A, Shazeer NM, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. *Neural Information Processing Systems*. 2017. Available from: <https://api.semanticscholar.org/CorpusID:13756489>
2. Hwang Y, Cornman AL, Kellogg EH, Ovchinnikov S, and Girguis PR. Genomic language model predicts protein co-regulation and function. *Nature Communications* 2024; 15(1):2880. DOI: [10.1038/s41467-024-46947-9](https://doi.org/10.1038/s41467-024-46947-9)
3. Ji Y, Zhou Z, Liu H, and Davuluri RV. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. *Bioinformatics* 2021; 37(15):2112–20. DOI: [10.1093/bioinformatics/btab083](https://doi.org/10.1093/bioinformatics/btab083)
4. Outeiral C and Deane CM. Codon language embeddings provide strong signals for use in protein engineering. *Nature Machine Intelligence* 2024; 6(2):170–9. DOI: [10.1038/s42256-024-00791-0](https://doi.org/10.1038/s42256-024-00791-0)
5. Chakraborty C, Bhattacharya M, and Lee SS. Artificial intelligence enabled ChatGPT and large language models in drug target discovery, drug discovery, and development. *Molecular Therapy - Nucleic Acids* 2023; 33:866–8. DOI: [10.1016/j.omtn.2023.08.009](https://doi.org/10.1016/j.omtn.2023.08.009)
6. Devlin J, Chang MW, Lee K, and Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* 2019. Available from: <https://arxiv.org/abs/1810.04805>

7. Theodoris CV, Xiao L, Chopra A, Chaffin MD, Sayed ZRA, Hill MC, et al. Transfer learning enables predictions in network biology. *Nature* 2023; 618(7965):616–24. DOI: [10.1038/s41586-023-06139-9](https://doi.org/10.1038/s41586-023-06139-9)
8. Barabási AL, Gulbahce N, and Loscalzo J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics* 2011; 12(1):56–68. DOI: [10.1038/nrg2918](https://doi.org/10.1038/nrg2918)
9. Loscalzo J, Barabási AL, and Silverman EK, eds. *Network Medicine: Complex systems in human disease and therapeutics*. Harvard University Press, 2017. DOI: [10.4159/9780674545533](https://doi.org/10.4159/9780674545533)
10. Gysi DM, Valle Í do, Zitnik M, Ameli A, Gan X, Varol O, et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proceedings of the National Academy of Sciences* 2021; 118(19):e2025581118. DOI: [10.1073/pnas.2025581118](https://doi.org/10.1073/pnas.2025581118)
11. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nature Communications* 2018; 9(1):2691. DOI: [10.1038/s41467-018-05116-5](https://doi.org/10.1038/s41467-018-05116-5)
12. Fiscon G, Conte F, Farina L, and Paci P. SAveRUNNER: A network-based algorithm for drug repurposing and its application to COVID-19. *PLOS Computational Biology* 2021; 17(2):e1008686. DOI: [10.1371/journal.pcbi.1008686](https://doi.org/10.1371/journal.pcbi.1008686)
13. Cheng F, Lu W, Liu C, Fang J, Hou Y, Handy DE, et al. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nature Communications* 2019; 10(1):3476. DOI: [10.1038/s41467-019-10744-6](https://doi.org/10.1038/s41467-019-10744-6)
14. Gysi DM and Barabási AL. Noncoding RNAs improve the predictive power of network medicine. *Proceedings of the National Academy of Sciences* 2023; 120(45):e2301342120. DOI: [10.1073/pnas.2301342120](https://doi.org/10.1073/pnas.2301342120)

15. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, et al. Uncovering disease-disease relationships through the incomplete interactome. *Science* 2015; 347(6224):1257601. DOI: [10.1126/science.1257601](https://doi.org/10.1126/science.1257601)
16. Choobdar S, Ahsen ME, Crawford J, Tomasoni M, Fang T, Lamparter D, et al. Assessment of network module identification across complex diseases. *Nature Methods* 2019; 16(9):843–52. DOI: [10.1038/s41592-019-0509-5](https://doi.org/10.1038/s41592-019-0509-5)
17. Chaffin M, Papangelis I, Simonson B, Akkad AD, Hill MC, Arduini A, et al. Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. *Nature* 2022; 608(7921):174–80. DOI: [10.1038/s41586-022-04817-8](https://doi.org/10.1038/s41586-022-04817-8)
18. Wang X, Gulbahce N, and Yu H. Network-based methods for human disease gene prediction. *Briefings in Functional Genomics* 2011; 10(5):280–93. DOI: [10.1093/bfpg/blr024](https://doi.org/10.1093/bfpg/blr024)
19. Ghiassian SD, Menche J, and Barabási AL. A Disease Module Detection (DIAMOND) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome. *PLOS Computational Biology* 2015; 11(4):e1004120. DOI: [10.1371/journal.pcbi.1004120](https://doi.org/10.1371/journal.pcbi.1004120)
20. Micheletti S, Schlauch D, Quackenbush J, and Guebila MB. Higher-order correction of persistent batch effects in correlation networks. *Bioinformatics* 2024; 40(9). DOI: [10.1093/bioinformatics/btae531](https://doi.org/10.1093/bioinformatics/btae531)
21. Vandebon A. Evaluation of critical data processing steps for reliable prediction of gene co-expression from large collections of RNA-seq data. *PLOS ONE* 2022; 17(1):e0263344. DOI: [10.1371/journal.pone.0263344](https://doi.org/10.1371/journal.pone.0263344)

22. Corchete LA, Rojas EA, Alonso-López D, Rivas JDL, Gutiérrez NC, and Burguillo FJ. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Scientific Reports* 2020; 10(1):19737. DOI: [10.1038/s41598-020-76881-x](https://doi.org/10.1038/s41598-020-76881-x)
23. Bozhilova LV, Whitmore AV, Wray J, Reinert G, and Deane CM. Measuring rank robustness in scored protein interaction networks. *BMC Bioinformatics* 2019; 20(1):446. DOI: [10.1186/s12859-019-3036-6](https://doi.org/10.1186/s12859-019-3036-6)
24. Le DH. Random walk with restart: A powerful network propagation algorithm in Bioinformatics field. *2017 4th NAFOSTED Conference on Information and Computer Science*. IEEE, 2017 :242–7. DOI: [10.1109/NAFOSTED.2017.8108071](https://doi.org/10.1109/NAFOSTED.2017.8108071)
25. Andersen R, Chung F, and Lang K. Local Graph Partitioning using PageRank Vectors. *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*. IEEE, 2006 :475–86. DOI: [10.1109/FOCS.2006.44](https://doi.org/10.1109/FOCS.2006.44)
26. Saelens W, Cannoodt R, and Saeys Y. A comprehensive evaluation of module detection methods for gene expression data. *Nature Communications* 2018; 9(1):1090. DOI: [10.1038/s41467-018-03424-4](https://doi.org/10.1038/s41467-018-03424-4)
27. Jarada TN, Rokne JG, and Alhajj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *Journal of Cheminformatics* 2020 Dec; 12(1):46. DOI: [10.1186/s13321-020-00450-7](https://doi.org/10.1186/s13321-020-00450-7)
28. Kim Y, Jung YS, Park JH, Kim SJ, and Cho YR. Drug-Disease Association Prediction Using Heterogeneous Networks for Computational Drug Repositioning. *Biomolecules* 2022; 12(10):1497. DOI: [10.3390/biom12101497](https://doi.org/10.3390/biom12101497)

29. Luo H, Wang J, Li M, Luo J, Peng X, Wu FX, et al. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* 2016; 32(17):2664–71. DOI: [10.1093/bioinformatics/btw228](https://doi.org/10.1093/bioinformatics/btw228)
30. Zitnik M, Sosič R, and Leskovec J. Prioritizing network communities. *Nature Communications* 2018; 9(1):2544. DOI: [10.1038/s41467-018-04948-5](https://doi.org/10.1038/s41467-018-04948-5)
31. Casas AI, Hassan AA, Larsen SJ, Gomez-Rangel V, Elbatreek M, Kleikers PWM, et al. From single drug targets to synergistic network pharmacology in ischemic stroke. *Proceedings of the National Academy of Sciences* 2019; 116(14):7129–36. DOI: [10.1073/pnas.1820799116](https://doi.org/10.1073/pnas.1820799116)
32. Guney E, Menche J, Vidal M, and Barábasi AL. Network-based in silico drug efficacy screening. *Nature Communications* 2016; 7(1):10331. DOI: [10.1038/ncomms10331](https://doi.org/10.1038/ncomms10331)
33. Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Research* 2024; 52(D1):D1265–D1275. DOI: [10.1093/nar/gkad976](https://doi.org/10.1093/nar/gkad976)
34. Borda JC de. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences* 1781; 12:657–64
35. Reilly B. Social Choice in the South Seas: Electoral Innovation and the Borda Count in the Pacific Island Countries. *International Political Science Review* 2002; 23(4):355–72. DOI: [10.1177/0192512102023004002](https://doi.org/10.1177/0192512102023004002)
36. Ferrari R, Lovering RC, Hardy J, Lewis PA, and Manzoni C. Weighted Protein Interaction Network Analysis of Frontotemporal Dementia. *Journal of Proteome Research* 2017; 16(2):999–1013. DOI: [10.1021/acs.jproteome.6b00934](https://doi.org/10.1021/acs.jproteome.6b00934)

37. Sharma A, Menche J, Huang CC, Ort T, Zhou X, Kitsak M, et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Human Molecular Genetics* 2015; 24(11):3005–20. DOI: [10.1093/hmg/ddv001](https://doi.org/10.1093/hmg/ddv001)
38. Silverman EK, Schmidt HHHW, Anastasiadou E, Altucci L, Angelini M, Badimon L, et al. Molecular networks in Network Medicine: Development and applications. *WIREs Systems Biology and Medicine* 2020; 12(6). DOI: [10.1002/wsbm.1489](https://doi.org/10.1002/wsbm.1489)

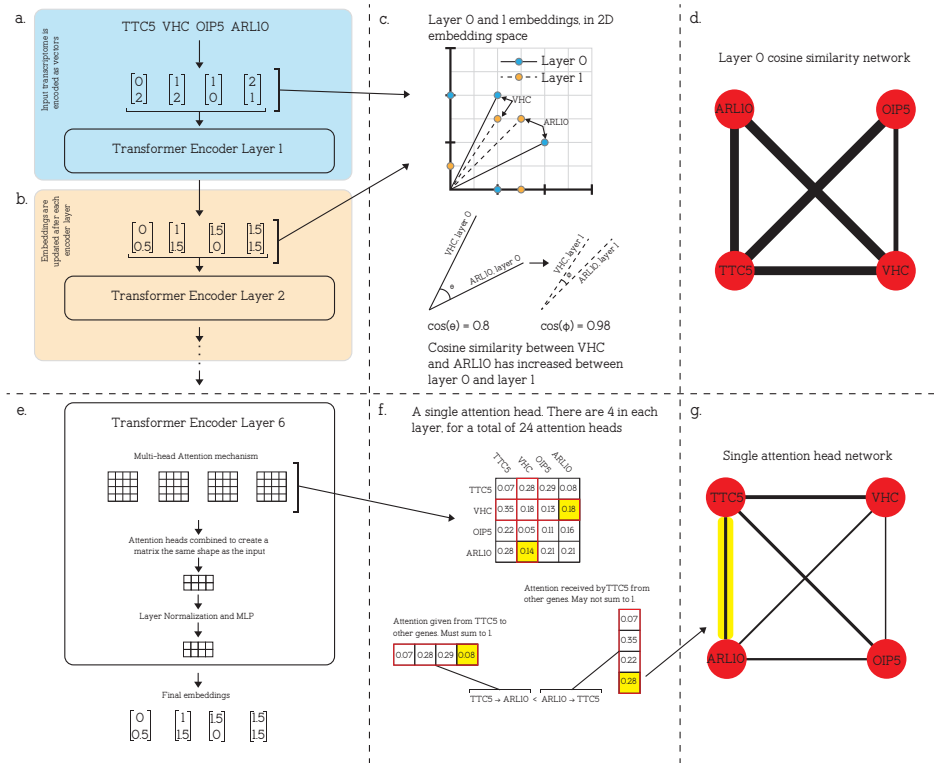


Figure 1: Similarity Measurements in Geneformer. (a) Starting with a transcriptome, genes are tokenized into vector embeddings and fed through the first transformer encoder layer. (b) Each layer outputs an updated set of embeddings before passing them into the next layer. (c) The initial embeddings (layer 0) and the layer 1 embeddings are plotted to show how the genes can move in embedding space between layers. In this case, VHC and ARL10 get closer in embedding space after layer 1. (d) The layer 0 embeddings are turned into a cosine similarity network, where each gene is connected to every other gene with an edge weighted by its cosine similarity. Here, ARL10 and OIP5 are orthogonal in embedding space, so no edge exists. (e) The final layer of the model is expanded to show the multi-head attention mechanism. All six layers have an identical mechanism, giving the model 24 total attention heads. (f) An attention head from the 6th layer is magnified. These attention weights are another way to compare gene similarities, as each gene attends to every other gene in the sequence. Since the matrix is asymmetric, the attention to TTC5 from ARL10 is different from the attention to ARL10 from TTC5. (g) The layer 6 attention weights are turned into a weighted directed network. This is an example of maximum aggregation, where we take the larger of a_{ij} , a_{ji} as the edge weight between i and j .

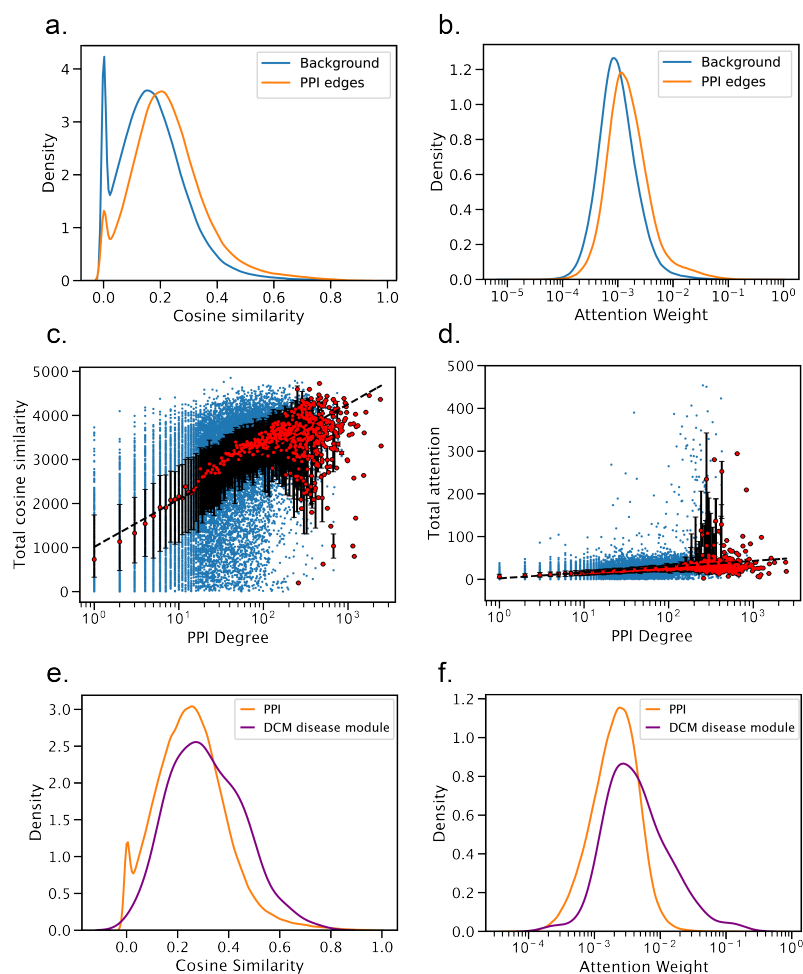


Figure 2: Genformer Prioritizes PPI Edges. (a) Using 10,000 cells from Genecorpus, we plot distributions of cosine similarities for the PPI and the fully connected background, observing a higher median weight in the PPI edges. (b) We plot the same distributions using attention weights and again find a higher median weight in the PPI edges. (c) Summing across rows of the cosine similarity matrix C , the total cosine similarity for a gene is plotted against its degree. We find that the total cosine similarity increases with the degree, and when modeled as a logistic regression, $\ln(k)$ explains 27% of the variation in C , the total cosine similarity. The red points with black error bars are the mean and standard deviation across all genes with a given degree k , and the black dashed line is the resulting fit from the logistic regression. The blue points represent individual genes. (d) Summing across rows of the attention matrix A , the total attention for a gene is plotted against its degree. We find that the total attention increases with the degree, and when modeled as a logistic regression, $\ln(k)$ explains 9% of the variation in A , the total attention. (e) Using 10,000 dilated cardiomyopathy (DCM) cardiomyocytes, we create a new cosine similarity matrix C and observe that cosine similarities within the largest connected component (LCC) of DCM disease genes in the PPI (also known as the disease module) have a higher median weight than other PPI edges. (f) We generate a new attention matrix A from these same 10,000 DCM cells, observing again that the DCM disease module edges have a higher median weight than other PPI edges.

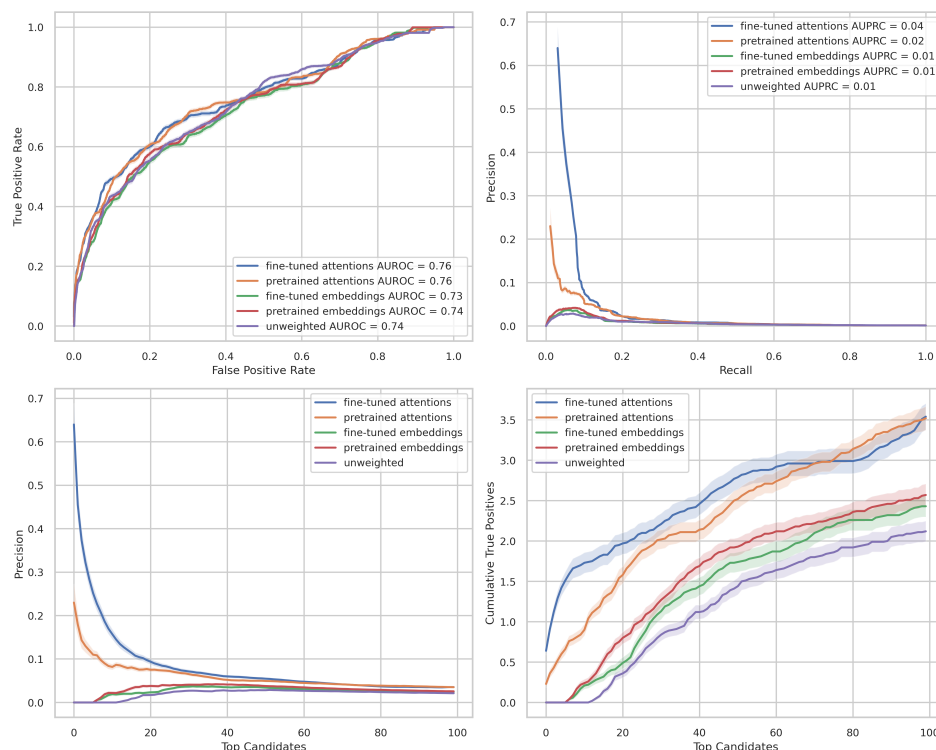


Figure 3: Geneformer Attention Weights Improve Capture-Recapture Analysis for Disease Module Discovery. (a) ROC curves are plotted for 4 different Geneformer weighted networks and a control unweighted network, comparing their recovery of removed disease module genes. The shaded ribbon indicates standard error in the mean of 100 trials with different seed genes. (b) The Precision-Recall curve is plotted for each network. The attention weighted networks have considerably higher initial precision than other networks, prompting us to look more closely at this region. (c) The precision is plotted for the top 100 candidates. The fine-tuned attention weighted network has higher precision for the top 100 candidates than any other network, followed by the pre-trained attention weighted network. (d) The cumulative true positives are plotted for the top 50 candidates. The attention weighted networks record the most true positives (between 3 and 4 on average), but the cosine similarity weighted networks also outperform the unweighted network.

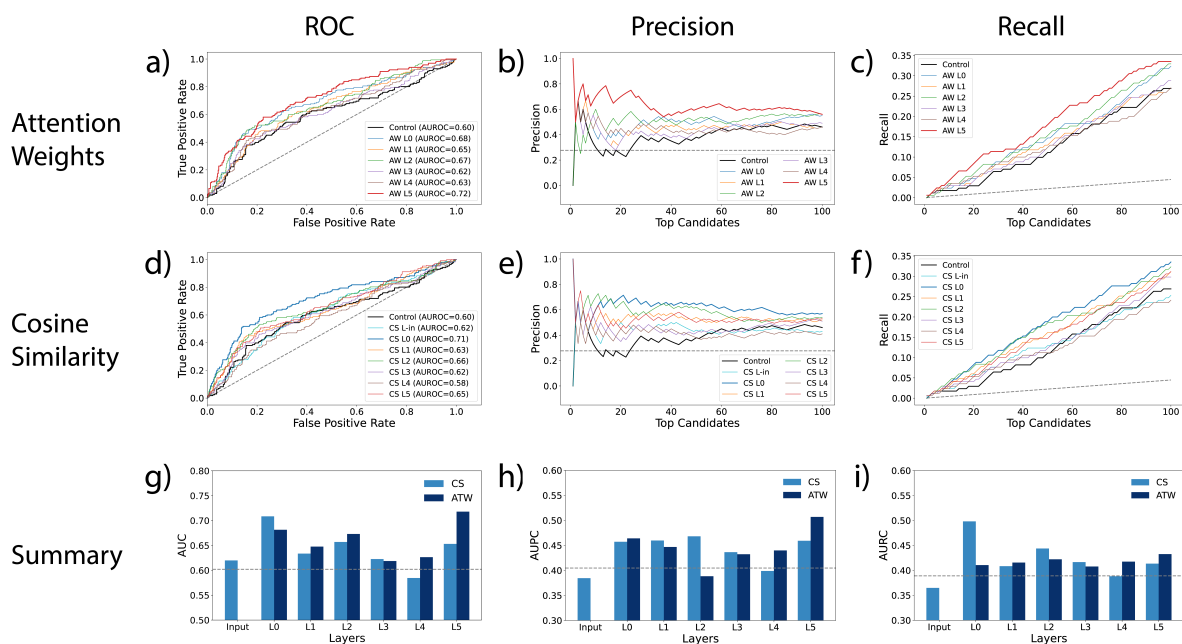


Figure 4: Layer-Wise Comparison of Geneformer Attention Weights and Cosine Similarity for Drug Repurposing in Dilated Cardiomyopathy. **a, d)** ROC curves comparing the drug repurposing performance using attention weights (AWs) and cosine similarities (CSs) from six Geneformer layers. AWs from layer 5 (AUROC = 0.72) and CSs from layer 0 (AUROC = 0.71) show the best predictive capacity in their respective sets, with both outperforming the unweighted control. **b, e)** Precision curves for the top 100 drug candidates. AWs from layer 5 (AUPC = 0.50) and CSs from layers 0 (AUPC = 0.45), 1 (0.45) and 2 (0.46) maintain the highest precision across most top candidates, indicating better predictive power compared to other layers and the control. **c, f)** Recall curves for the top 100 drug candidates using AWs and CSs from different layers. AWs from layer 5 (AURC = 0.43) and CSs from layer 0 (AURC = 0.50) achieve the highest recall, identifying more known cardiomyopathy drugs in the top candidates compared to other layers and the control network. **g-i)** AUROC, AUPC and AURC values comparing AWs and CSs across Geneformer layers. The gray dashed line represents the respective values derived from the unweighted control network (AUROC = 0.60, AUPC = 0.40, AURC = 0.39). In all cases, the results represent the average of ten independent runs, with a standard deviation below 0.02.

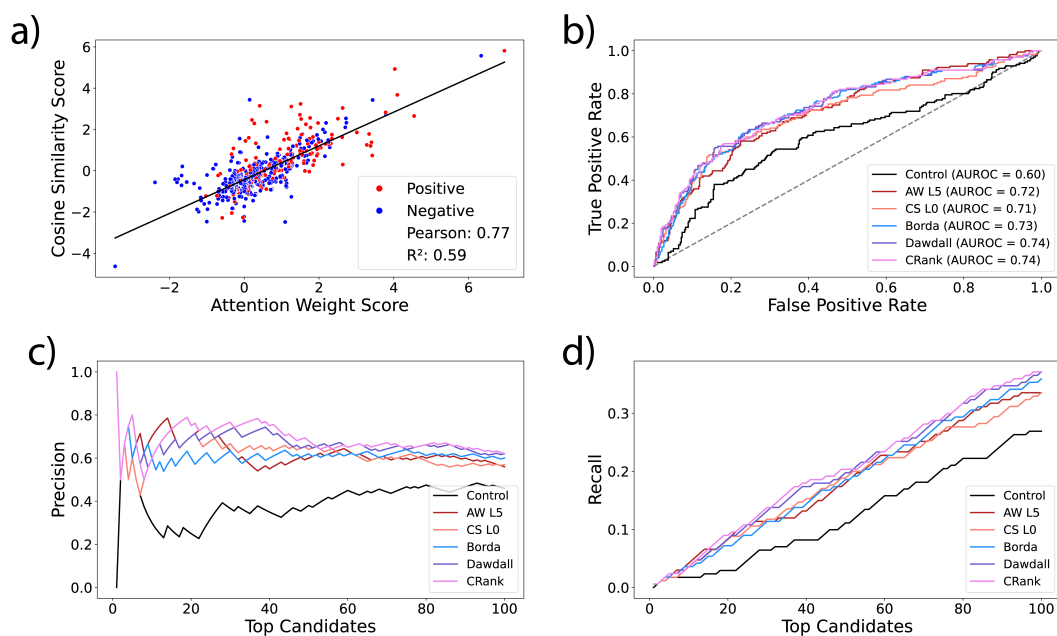


Figure 5: Correlation and Performance of Drug Repurposing Scores Derived from Cosine Similarity (CSs) from Layer 0 and Attention Weights (AWs) from Layer 5 Using Combined Ranking Methods. **a)** Correlation between drug repurposing scores derived from CSs from layer 0 and AWs from layer 5. The score is calculated as $-1 \times \text{proximity}(\text{DCM}, D)$, where DCM represents dilated cardiomyopathy disease genes, and D represents the drug targets. Blue points correspond to negative drugs, and red points correspond to positive drugs used to treat cardiomyopathy. The R^2 value is 0.59. **b)** Receiver Operating Characteristic (ROC) curves comparing the performance of drug repurposing using individual scores from CSs from layer 0, AWs from layer 5, and combined scores generated using Borda Count, Dowdall Count, and CRank methods. The combined ranking methods show improved performance, with CRank (AUROC = 0.74) and Dowdall (AUROC = 0.74) achieving the highest AUROC values. **c)** Precision curves for the top 100 drug candidates ranked by proximity to cardiomyopathy genes. The combined ranking methods show improved precision over the individual CS and AW scores, with CRank maintaining higher precision across most top candidates. **d)** Recall curves for the top 100 drug candidates, comparing individual and combined scores. The combined scores, particularly CRank and Dowdall, achieve higher recall, identifying more known cardiomyopathy drugs in the top-ranked candidates. In all cases, the results represent the average of ten independent runs, with a standard deviation below 0.02.