

Transfer Learning Enables Predictions In Network Biology

Article

Transfer learning enables predictions in network biology

<https://doi.org/10.1038/s41586-023-06139-9>

Received: 29 March 2022

Accepted: 27 April 2023

Published online: 31 May 2023

 Check for updates

Christina V. Theodoris^{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,287,288,289,290,291,292,293,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,388,389,390,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,443,444,445,446,447,448,449,450,451,452,453,454,455,456,457,458,459,460,461,462,463,464,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479,480,481,482,483,484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,571,572,573,574,575,576,577,578,579,580,581,582,583,584,585,586,587,588,589,590,591,592,593,594,595,596,597,598,599,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985,986,987,988,989,990,991,992,993,994,995,996,997,998,999,1000}, Ling Xiao^{3,5}, Anant Chopra⁶, Mark D. Chaffin⁷, Zeina R. Al Sayed⁸, Matthew C. Hill^{9,10}, Helene Mantione¹¹, Elizabeth M. Brydon¹², Zexian Zeng¹³, X. Shirley Liu^{14,15} & Patrick T. Ellinger^{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,287,288,289,290,291,292,293,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,388,389,390,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,443,444,445,446,447,448,449,450,451,452,453,454,455,456,457,458,459,460,461,462,463,464,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479,480,481,482,483,484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,571,572,573,574,575,576,577,578,579,580,581,582,583,584,585,586,587,588,589,590,591,592,593,594,595,596,597,598,599,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985,986,987,988,989,990,991,992,993,994,995,996,997,998,999,1000}

Mapping gene networks requires large amounts of transcriptomic data to learn the connections between genes, which impedes discoveries in settings with limited data, including rare diseases and diseases affecting clinically inaccessible tissues. Recently, transfer learning has revolutionized fields such as natural language understanding^{1,2} and computer vision³ by leveraging deep learning models pretrained on large-scale general datasets that can then be fine-tuned towards a vast array of downstream tasks with limited task-specific data. Here, we developed a context-aware, attention-based deep learning model, Geneformer, pretrained on a large-scale corpus of about 30 million single-cell transcriptomes to enable context-specific predictions in settings with limited data in network biology. During pretraining, Geneformer gained a fundamental understanding of network dynamics, encoding network hierarchy in the attention weights of the model in a completely self-supervised manner. Fine-tuning towards a diverse panel of downstream tasks relevant to chromatin and network dynamics using limited task-specific data demonstrated that Geneformer consistently boosted predictive accuracy. Applied to disease modelling with limited patient data, Geneformer identified candidate therapeutic targets for cardiomyopathy. Overall, Geneformer represents a pretrained deep learning model from which fine-tuning towards a broad range of downstream applications can be pursued to accelerate discovery of key network regulators and candidate therapeutic targets.

Mapping the gene regulatory networks that drive disease progression enables screening for molecules that correct the network by normalizing core regulatory elements, rather than targeting peripheral downstream effectors that may not be disease modifying^{4,5}. However, mapping the gene network architecture requires large amounts of transcriptomic data to learn the connections between genes, which impedes network-correcting drug discovery in settings with limited data, including rare diseases and diseases affecting clinically inaccessible tissues. Although data remain limited in these settings, recent advances in sequencing technologies have driven a rapid expansion in the amount of transcriptomic data available from human tissues more broadly. Furthermore, single-cell technologies have facilitated the observation of transcriptomic states without averaging the expression of genes across multiple cells, potentially providing more precise data for inference of network interactions, especially in diseases driven by dysregulation of multiple cell types.

Recently, the concept of transfer learning has revolutionized fields such as natural language understanding^{1,2} and computer vision³ by leveraging deep learning models pretrained on large-scale general datasets that can then be fine-tuned towards a vast array of downstream tasks with limited task-specific data that would be insufficient

to yield meaningful predictions when used in isolation. Unlike modelling approaches that necessitate retraining a new model from scratch for each task^{6,7}, this approach democratizes the fundamental knowledge learned during the large-scale pretraining phase to a multitude of downstream applications distinct from the pretraining learning objective, transferring knowledge to new tasks (Fig. 1a and Extended Data Fig. 1a,b). The advent of the self-attention mechanism^{8,9} has further transformed the deep learning field by generating context-aware models that are able to pay attention to large input spaces and learn which elements are most important to focus on in each context, boosting predictions in a wide realm of applications^{10–15}. Gene regulatory network architectures are highly context-dependent, and attention-based models, known as transformers, may be exceptionally suited to context-specific modelling of network dynamics.

Here, we developed a context-aware, attention-based deep learning model, Geneformer, pretrained on large-scale transcriptomic data to enable predictions in settings with limited data. We assembled a large-scale pretraining corpus, Genecorpus-30M, comprising 29.9 million human single-cell transcriptomes from a broad range of tissues from publicly available data. We then pretrained Geneformer on this corpus using a self-supervised masked learning objective to gain a fundamental

¹Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA. ²Cardiovascular Disease Initiative and Precision Cardiology Laboratory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ³Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA. ⁴Harvard Medical School Genetics Training Program, Boston, USA. ⁵Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. ⁶Precision Cardiology Laboratory, Bayer US LLC, Cambridge, MA, USA. ⁷Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ⁸Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA. ⁹e-mail: christina.theodoris@dfci.harvard.edu; ellinger@genetics.harvard.edu

Transfer learning enables predictions in network biology, a field that has gained immense traction in recent years due to the explosive growth of biological data and advances in machine learning. Network biology examines the complex interactions between biological entities, such as genes, proteins, and metabolites, often represented as networks or graphs. The integration of transfer learning into this domain has proven to be a game-changer, allowing researchers to make predictions that were previously unattainable with traditional methods.

Understanding Transfer Learning

Transfer learning is a machine learning paradigm that focuses on leveraging knowledge gained from one task to improve performance in a related task. This method is particularly beneficial when the amount of labeled data for the target task is limited. By using a pre-trained model, which has been trained on a large dataset from a source domain, researchers can adapt this model to the target domain with significantly less data.

How Transfer Learning Works

The core idea behind transfer learning involves the following steps:

1. Pre-training: A model is trained on a large dataset (source domain) where abundant labeled data is available.
2. Fine-tuning: The pre-trained model is then adapted to a specific task or dataset (target domain) by further training it with a smaller amount of labeled data.
3. Prediction: Once fine-tuned, the model can make predictions on new, unseen data related to the target domain.

This process allows for the effective transfer of knowledge across domains, which is particularly useful in fields like network biology, where obtaining labeled data can be challenging.

Applications of Transfer Learning in Network Biology

The integration of transfer learning in network biology has led to several promising applications, each geared toward understanding biological networks and enhancing predictive capabilities. Below are some of the key applications:

1. Gene Function Prediction

Gene function prediction is a critical aspect of understanding biological networks. Traditional methods often rely on experimental data, which can be sparse and time-consuming to obtain. Transfer learning can significantly improve gene function prediction by:

- Leveraging existing knowledge: By using models trained on large datasets of genes with known functions, researchers can predict the functions of poorly characterized genes in different species.

- Cross-species predictions: Transfer learning allows for the transfer of functional information across species, enabling predictions even when extensive data is not available for a particular organism.

2. Protein-Protein Interaction Prediction

Protein-protein interactions (PPIs) are fundamental in understanding cellular processes. Transfer learning can enhance PPI prediction through:

- Integrating diverse data sources: Models can be pre-trained on extensive datasets containing known PPIs. By fine-tuning these models on specific subsets of data, researchers can improve prediction accuracy.
- Handling noisy data: Biological data often contain noise and inconsistencies. Transfer learning helps in building more robust models that can generalize better despite the presence of such noise.

3. Disease Prediction and Classification

Transfer learning plays a pivotal role in predicting diseases based on network biology. This includes:

- Identifying disease biomarkers: By utilizing models trained on large-scale genomic and proteomic datasets, researchers can identify potential biomarkers for various diseases.
- Classifying disease subtypes: By adapting models to specific disease subtypes, transfer learning enables more accurate classification, aiding in personalized medicine approaches.

4. Network Reconstruction

Reconstructing biological networks from high-dimensional data is a key challenge in network biology. Transfer learning assists in:

- Imposing prior knowledge: By using pre-trained models that encapsulate existing biological knowledge, researchers can reconstruct networks more accurately.
- Reducing dimensionality: Transfer learning can help simplify complex biological data, making network reconstruction more feasible.

Challenges and Considerations

While transfer learning has opened new avenues in network biology, several challenges remain. Addressing these challenges is crucial for maximizing the

potential of transfer learning in this field.

1. Domain Shift

One of the primary challenges in transfer learning is the issue of domain shift, where the source and target domains differ significantly. Such disparities can lead to suboptimal model performance. Researchers must carefully select source domains that are closely related to the target tasks to mitigate this issue.

2. Data Imbalance

Biological datasets are often imbalanced, with a few classes having a large number of samples while others are underrepresented. This imbalance can affect the performance of transfer learning models. Techniques such as oversampling, undersampling, and using weighted loss functions can be employed to address this challenge.

3. Interpretability

Machine learning models, particularly deep learning models used in transfer learning, are often criticized for their lack of interpretability. In network biology, where understanding the biological implications of predictions is crucial, developing interpretable models is essential. Researchers are exploring methods to enhance model interpretability, such as attention mechanisms and feature importance scores.

Future Prospects of Transfer Learning in Network Biology

The future of transfer learning in network biology appears promising, with several exciting developments on the horizon:

1. Integration with Multi-Omics Data

As biological research increasingly incorporates multi-omics data (genomics, transcriptomics, proteomics, etc.), transfer learning can play a vital role in integrating these diverse datasets. By leveraging transfer learning, researchers can create unified models that provide a more comprehensive view of biological systems.

2. Advances in Model Architectures

Recent advancements in model architectures, such as graph neural networks, offer new opportunities for transfer learning in network biology. These models are particularly suitable for representing biological networks, enabling more effective transfer of knowledge across domains.

3. Enhanced Collaborations Across Disciplines

The intersection of biology, computer science, and statistics is becoming increasingly important. Enhanced collaborations among researchers from these disciplines will foster the development of more sophisticated transfer learning approaches tailored to specific biological questions.

Conclusion

In summary, **transfer learning enables predictions in network biology** by providing powerful tools to leverage existing knowledge and improve predictive performance in various biological tasks. Its applications in gene function prediction, protein-protein interaction prediction, disease classification, and network reconstruction highlight its transformative potential. While challenges such as domain shift, data imbalance, and interpretability persist, ongoing research and advancements in this field promise to overcome these hurdles. As transfer learning continues to evolve, its integration into network biology will undoubtedly lead to new insights and breakthroughs in understanding the complex interplay of biological networks.

Frequently Asked Questions

What is transfer learning in the context of network biology?

Transfer learning in network biology refers to the technique of utilizing a pre-trained model on one biological network to improve predictions in another, potentially related network, thereby saving time and resources in model training.

How does transfer learning improve prediction accuracy in network biology?

Transfer learning enhances prediction accuracy by leveraging knowledge gained from previously learned tasks, allowing models to make more informed

predictions in new, but similar, biological contexts.

What types of data can benefit from transfer learning in network biology?

Types of data that can benefit from transfer learning in network biology include gene expression data, protein-protein interaction networks, and various omics datasets that share underlying biological patterns.

What challenges does transfer learning face in network biology?

Challenges in transfer learning for network biology include the difficulty in aligning different biological networks, variations in data quality, and the need for domain-specific adaptations of transfer learning techniques.

Can transfer learning be applied to predict disease outcomes in network biology?

Yes, transfer learning can be applied to predict disease outcomes in network biology by transferring knowledge from networks associated with one disease to enhance predictions for another disease, improving early diagnosis and treatment strategies.

What future advancements are expected in transfer learning for network biology?

Future advancements in transfer learning for network biology are expected to include improved algorithms for network alignment, better integration of multi-omics data, and the development of more sophisticated models that can generalize across diverse biological contexts.

Find other PDF article:

<https://soc.up.edu.ph/48-shade/Book?trackid=Nle36-3948&title=present-perfect-subjunctive-practice.pdf>

Transfer Learning Enables Predictions In Network Biology

Fußball-Transfers, Gerüchte, Marktwerte und News | Trans...

Transfermarkt - Das Fußball-Portal mit Transfers, Marktwerten, Gerüchten und Statistiken

WHO AM I? | Fußball-Quiz | Transfermarkt

Errate den gesuchten Spieler in 6 Versuchen! □ Nationalität, Liga, Verein, Position, Alter,

Rückenummer & ...

[Transfers & Gerüchte | Transfermarkt](#)

1 day ago · Top-Transfer-Statistiken Neueste Transfers weltweit Neueste Transfers Bundesliga Neueste ...

Gerüchteküche - Forum | Transfermarkt

Aug 13, 2009 · Dies ist die Übersicht aller Threads aus dem Forum Gerüchteküche, die in der ...

[Die neuesten Transfers | Transfermarkt](#)

Zuletzt bestätigte Transfers: weltweit Alter Verein Neuer Verein Ablöse nach gewünschtem Land filtern

[Fußball-Transfers, Gerüchte, Marktwerte und News | Transfermarkt](#)

Transfermarkt - Das Fußball-Portal mit Transfers, Marktwerten, Gerüchten und Statistiken

WHO AM I? | Fußball-Quiz | Transfermarkt

Errate den gesuchten Spieler in 6 Versuchen! □ Nationalität, Liga, Verein, Position, Alter, Rückennummer & Marktwert helfen dir auf die Spur Bist Du ein echter Experte?

Transfers & Gerüchte | Transfermarkt

1 day ago · Top-Transfer-Statistiken Neueste Transfers weltweit Neueste Transfers Bundesliga Neueste Transfers 2. Bundesliga Top-Transfers Auslaufende Verträge Vertragslose Spieler ...

[Gerüchteküche - Forum | Transfermarkt](#)

Aug 13, 2009 · Dies ist die Übersicht aller Threads aus dem Forum Gerüchteküche, die in der Transfermarkt-Community diskutiert werden.

Die neuesten Transfers | Transfermarkt

Zuletzt bestätigte Transfers: weltweit Alter Verein Neuer Verein Ablöse nach gewünschtem Land filtern

Transfers & Gerüchte | Transfermarkt

3 days ago · Aktuelle Meldungen zur Rubrik Transfers & Gerüchte auf Transfermarkt.

transport, transit, transmit, transfer□□□□□□□□ - □□

□□3□I will transfer the money tomorrow. □□4□The transfer of files to the new computer is complete. □□5□We transfer our job duties to the new employee. 2. □□transport ...

Transferticker | Transfermarkt

Der Transferticker von Transfermarkt gibt einen Überblick zu aktuellen Transferrmeldungen, Trainerentlassungen und Vertragsverlängerungen.

Bundesliga 25/26 | Transfermarkt

Bundesliga auf Transfermarkt mit Tabelle Ergebnissen Spielplan Liveticker Marktwerten Vereine Transfers Statistiken

SC Freiburg - Vereinsprofil | Transfermarkt

Alles zum Verein SC Freiburg (Bundesliga) aktueller Kader mit Marktwerten Transfers Gerüchte Spieler-Statistiken Spielplan News

Discover how transfer learning enables predictions in network biology

[Back to Home](#)