High-Accuracy MicroRNA Target Prediction Using Disease Semantic Strings

Baiming CHEN[*1,2] and Hsi-Yuan Huang[†3]

[1]School of Medicine, The Chinese University of Hong Kong, Shenzhen, Longgang District, Shenzhen, Guangdong Province 518172, China
[2]Shenzhen X-Institute, Nanshan District, Shenzhen, Guangdong Province 518055, China

May 25, 2024

Abstract

MicroRNAs (miRNAs) are non-coding RNAs approximately 22 nucleotides in length. miRNAs regulate gene expression by binding to target mRNAs, inhibiting translation, or promoting mRNA degradation. miRNAs are of great importance in the development of disease. So far, a variety of miRNA target prediction tools have been available, which are able to identify and predict miRNA sequences and their target genes, providing key support for studying miRNA regulatory networks, resolving disease mechanisms and developing potential therapeutic strategies. In this paper, we propose a prediction method that predicts miRNA-gene relationships based on the disease semantic strings of miRNAs and their potential target genes, named the MicroRNA-Gene Connection Building Method. The method was trained on more than 1,900 pairs of experimental and predictive data consisting of 10 cancer-related genes and evaluated using stratified cross-validation with a model accuracy of more than 90%. Our method also demonstrates outstanding transferability, we used this model to classify 10 genes related to immunity and inflammation with a classification accuracy of over 85%.

1 Introduction

In 1993, miRNA (Lin-4) was first discovered in humans and found to be critical for the timing of embryonic development in C. elegans [1]. Since then, a large number of miRNAs have been discovered with the development of high-throughput bioinformatics, technologies [2]. microRNAs (miRNAs) belong to small ncRNAs, and miRNAs play a key role in regulating many physiological processes, such as cell cycle [3], cell growth, development, differentiation, and apoptosis [4], as well as pathological processes such as those associated with various cancers [5]. In addition, miRNAs may be promising biomarker candidates for early detection or prognosis of various diseases [6].

[*]Corresponding author: `baimingchen@link.cuhk.edu.cn`
[†]First corresponding author: `huanghsiyuan@cuhk.edu.cn`

Sentence-BERT (SBERT) fine-tunes BERT using siamese and triple network structures and is an extension of the BERT (Bidirectional Encoder Representations from Transformers) model. SBERT achieves good results in semantic similarity tasks [7]. The SBERT has faster speed compared to BERT [8] and RoBERTa [9] while maintaining accuracy.

There are many existing miRNA prediction tools. miRDB uses machine learning to predict human miRNA targets based on experimental data [10]. TargetScan predicts miRNA targets based on seed match and conservativeness analysis [11]. miRanda predicts miRNA targets based on the free energy of sequence matching [12]. PITA predicts miRNA targets by considering both the binding free energy and the accessibility of the target site within the mRNA secondary structure [13]. In this paper, we propose a prediction method based on the Sentence-BERT model to predict miRNA-gene relationships by miRNA and gene disease semantic similarity. We selected 10 genes highly related to cancer(TP53, PTEN, KRAS, BCL2, FOXO3, KIF2C, PLK1, CENPA, MYC, STAT3) to test the model [14–23]. The miRNAs corresponding to functional interaction with experiment verify for each gene, and the miRNAs predicted from the miRDB were input into the model, 1932 scores were obtained by model. The distribution of scores is shown in the box plot below, and the model has excellent performance.

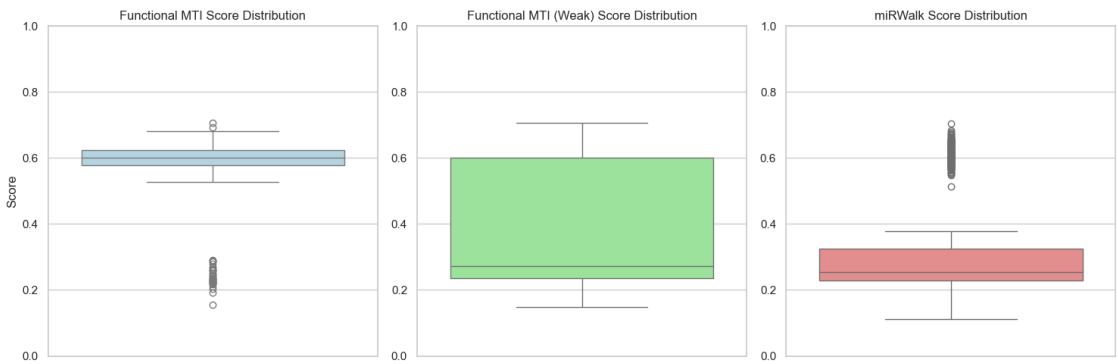

Figure 1: Score distribution for 10 cancer-related genes

2 Method

2.1 Fine-tune model

For the model to have the ability to handle the biomedical string semantic similarity computation task while maintaining its original performance, we generated a disease semantic similarity dataset to fine-tune the model. We downloaded the MeSH descriptors(https://www.nlm.nih.gov/mesh) from the National Library of Medicine and selected category C (diseases). To calculate the similarity between different disease semantic strings, we proposed a method that yields a distribution of values that is similar to the distribution of semantic similarity values output by the sentiment-transformers/multi-qa-MiniLM-L6-cos-v1 [24] model without fine-tuning. Both distribution types are log-normal distributions. The specific methods are as follows:

- Determine the depth of the lowest common ancestor (LCA):

$$D_{\text{LCA}} = \text{find_lca_depth}(\text{code1}, \text{code2}) \tag{1}$$

- Calculate the depth of each disease code:

$$D_1 = \text{len}(\text{code1.split('.')}) \tag{2}$$

$$D_2 = \text{len}(\text{code2.split('.')}) \tag{3}$$

- Calculate the maximum depth and total depth:

$$D_{\text{max}} = \max(D_1, D_2) \tag{4}$$

$$D_{\text{total}} = D_1 + D_2 - D_{\text{LCA}} \tag{5}$$

- Calculate the base similarity:

$$\text{similarity} = \frac{D_{\text{LCA}}}{D_{\text{total}}} \tag{6}$$

- Calculate the weighted similarity:

$$\text{weighted_similarity} = \text{similarity}^{1+\left(\frac{D_{\text{LCA}}^{\text{DecayFactor}}}{D_{\text{total}}}\right)} \tag{7}$$

In this paper, we chose the decay factor to be 0.8. Based on the formula, the weighted similarity between C18.452.648.398.500.440 Hypobetalipoproteinemias and C18.452.648.398.500.440 Abetalipoproteinemia is 0.781542212457061. The weighted similarity was transformed to match the model 'multi-qa-MiniLM-L6-cos-v1' output distribution to get the fine-tuned dataset. After fine-tuning the model with the dataset, we tested the model with the BIOSSES dataset, a dataset for evaluating biomedical text similarity measures. BIOSSES contains 100 pairs of biomedical statements, each with human-annotated similarity scores [25]. On this dataset, the 'Connection Building Method' model achieves good performance, demonstrating that after fine-tuning with MeSH data, the model retains the original pattern while learning new data.

The dataset and model have been uploaded to hugging face and named 'Connection_Building_Method_Disease'.

2.2 Model Design

To calculate the similarity between the miRNA- and gene-related diseases. We propose an algorithm to compare the similarity of two lists of disease strings. The algorithm combines clustering and semantic string similarity calculation to evaluate the overall similarity between miRNA- and gene-disease lists.

2.2.1 Semantic Coding and Similarity Calculation

First, each string in the list is encoded using the Sentence-BERT model. The similarity between two semantic vectors is obtained by calculating the cosine similarity:

$$\text{Cosine Similarity} = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\|\|\vec{B}\|}$$

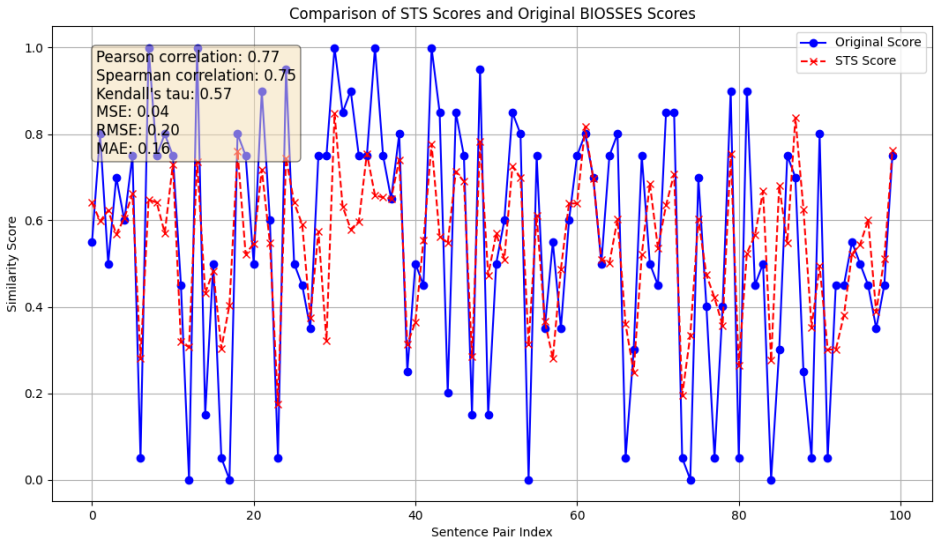

Figure 2: Scores from Connection Building Method vs. Human Scores on BIOSSES Dataset

2.2.2 Clustering

To classify miRNA-associated and gene-associated diseases into different groups to facilitate subsequent similarity calculations. First, merge the miRNA and gene diseases string list. The first element of the list automatically becomes the cluster center. As iterating through the remaining elements, if an element's semantic similarity to the cluster center exceeds the threshold, the element is added to that cluster. Otherwise, it becomes a new cluster center. After clustering, the elements in each class are traversed to find the element with the highest similarity sum as the center of the class.

$$c^* = \arg\max_{i \in C} \sum_{j \in C, i \neq j} \text{Cosine Similarity}(\vec{v_i}, \vec{v_j})$$

Parameter Definitions: where $\vec{v_i}$ and $\vec{v_j}$ represent the semantic vectors of elements i and j respectively, and Cosine Similarity$(\vec{v_i}, \vec{v_j})$ is the cosine similarity between these vectors.

2.2.3 Evaluation of Clustering Results

For the result after clustering, if the number of successfully clustered(successful clustering is defined that the elements within the cluster are not a subset of either miRNA disease list or gene disease list) clusters is greater than n(in this paper, we chose n=5). Accumulate the average similarity within each successful clustered cluster and divide by the number of successful clustered clusters to get the overall similarity between miRNA and gene disease list:

$$\text{Similarity}_{\text{successfully cluster}} = \frac{1}{|C|} \sum_{\text{(i, j) in C}} \text{Cosine Similarity}(\vec{v_i}, \vec{v_j})$$

Parameter Definitions: C denotes all pairwise elements within a cluster, and $\vec{v_i}$ and $\vec{v_j}$ represent the semantic vectors of the i-th and j-th elements, respectively.

If the number of successfully clustered clusters is less than n, the similarity between the two lists is calculated:

$$\text{Sum_1} = \sum_{i \in \text{miRNA_Disease}} \left(\frac{\sum_{j \in \text{Gene_Disease}} \text{Cosine Similarity}(\vec{v_i}, \vec{v_j})}{\text{len(Gene_Disease)}} \right)$$

$$\text{Sum_2} = \sum_{j \in \text{Gene_Disease}} \left(\frac{\sum_{i \in \text{miRNA_Disease}} \text{Cosine Similarity}(\vec{v_j}, \vec{v_i})}{\text{len(miRNA_Disease)}} \right)$$

$$\text{Similarity}_{\text{unsuccessfully cluster}} = \frac{\text{Sum_1} + \text{Sum_2}}{\text{len(miRNA_Disease)} + \text{len(Gene_Disease)}}$$

Parameter Definitions: $\vec{v_i}$ and $\vec{v_j}$ represent the semantic vectors of the i-th and j-th elements of the miRNA disease list and gene disease list, respectively.

2.2.4 Overall similarity measure

Ultimately, the value obtained from the above calculation is used as a measure of overall similarity between the two disease lists. With these steps, we are able to assess the semantic similarity between the miRNA-disease list and gene-disease list and thus the relevance between miRNA-related disease and gene-related diseases.

2.3 Data Process

To acquire datasets to test models and algorithms. We selected miRTarbase [26], miRWalk [27], HMDD [28] and DisGeNet [29]. miRTarbase provides experimentally validated miRNA-gene pairs. miRWalk provides predicted miRNA-gene pairs. HMDD provides microRNA-related disease data. DisGeNet provides gene-related disease data. Here, we processed the miRTarbase data by merging the -3p and -5p of the same miRNA, in order to match with the HMDD database.

We downloaded disease data related to 10 genes (TP53, PTEN, KRAS, BCL2, FOXO3, KIF2C, PLK1, CENPA, MYC, STAT3) from DisGeNET and chose the top 40 diseases with the highest scores for each gene to make up the gene's associated disease semantic string list.

Next, we searched for the miRNAs corresponding to these 10 genes in miRTarBase and miRWalk and obtained the list of related disease semantic strings for the miRNAs in the HMDD database. To filter the data from miRWalk, we set the following conditions: the prediction score is greater than 0.5, and the binding region is limited to the 3'UTR, in order to provide predictive data that is as inaccurate as possible, treating as Non-functional MTI data. In addition, to ensure the reasonableness of the data, we restricted the number of miRNAs from miRWalk to be at most twice as many as those from miRTarBase. Because if both were of the same size (i.e., 1:1 ratio), we would have difficulty in accurately distinguishing the difference between the predicted data and the experimental data (we assumed that the accuracy of the prediction database was any value between 0 and 1). Since the prediction database may contain both Functional MTIs and Non-Functional MTIs, by increasing the size of the prediction data, we can more clearly distinguish the difference between the scores of Functional MTIs and Non-Functional MTIs.

3 Results

Based on these 10 genes, the model outputs a total of 1932 data, of which 250 are labeled with 'Functional MTI', 394 are labeled with 'Functional MTI (Weak)' and 1288 are labeled with 'miRWalk'. Our classification using SVM on datasets with different levels of outlier removal shows impressive accuracy. The specific results are as follows:

1. **After removing the top and bottom 5% outliers for each gene**, the model achieved an accuracy of 82%.
2. **After removing the top and bottom 10% outliers for each gene**, the model's accuracy further increased to 90%.
3. **After removing the top and bottom 20% outliers for each gene**, the model achieved an accuracy of 98%.

To assess the performance of the model, we used the `StratifiedKFold` method for cross-validation with data after removing the top and bottom 10% outliers for each gene with the fold value set to 5. On each fold, we calculated the AUC (area under the curve) value of the ROC curve. The AUC values for each fold are as follows:

- Fold 1: 0.84
- Fold 2: 0.94
- Fold 3: 0.90
- Fold 4: 0.91
- Fold 5: 0.90

The average AUC value is 0.90 with a standard deviation of 0.03, and the decision boundary for the SVM classifier is 0.48705359.

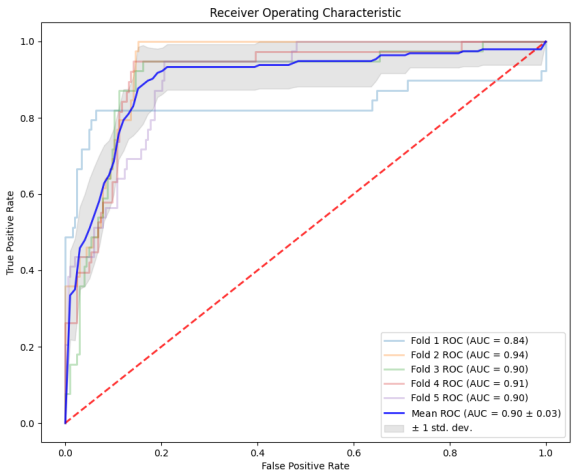

Figure 3: ROC Curve of Model

In addition, we plotted each of the 10 genes individually and the images are as follows, we found that the miRNA-gene pairings with strong experimental validation (validated by Reporter Assay or Western blot or qPCR), and with weak experimental validation (validated by Microarray, NGS and other methods) and with pure

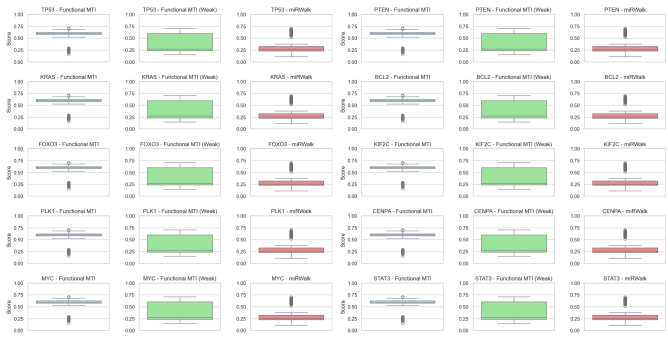

Figure 4: Each Gene Score Distribution

predictive database (miRWalk), the score intervals of these three types of data are from high to low. All the resulting distribution of the 10 genes matches biological evidence.

4 Discussion

In this paper, we propose a Connection Building Method for predicting miRNA-gene relationships. The method is based on disease semantic string list similarity computation of miRNA- and gene-related diseases to obtain miRNA-gene MTI scores. This experiment's results and the model's testing on the BIOSSES dataset show that the Natural Language Processing (NLP) method performs excellently in the biomedical domain, demonstrating that the semantic similarity algorithm is still highly accurate when dealing with biomedical vocabularies after specific fine-tuning.

The DisGeNET database provides 21,671 genes with associated disease types, and HMDD provides 1,890 microRNAs and their associated diseases. Some scholars estimate that the total number of human genes is around 22,333 [30], and the number of total human mature miRNAs is around 2,300 [31]. According to the estimate, the Connection Building Method is able to cover a large portion of miRNAs and genes. The data source used to calculate the disease semantic similarity is sufficient. We also have a question of whether only cancer-related genes have such a pattern. So we additionally selected 10 genes related to immune and inflammatory pathways, namely TNF, IL6, TGFB1, IL1B, IL10, IL17A, MMP9, CRP, TLR4, NLRP3 [32–41]. 342 miRNA gene pairwise scores were found using the Connection Building Method. The SVM classification model for cancer-related genes was used to classify these 342 sets of data, and after removing 10% outliers above and below each gene, the model achieved 86% accuracy, which suggests a high likelihood that the pattern is generalizable.

The most important point is that the actual accuracy of the model must exceed what we have gotten so far because we consider all the low threshold data from the prediction database to be Non-Functional MTI. In fact, according to the box plot, it is clear that there are some data distributions from miRWalk that are in the Functional MTI interval. This phenomenon also coincides with the fact that the predicted data has some accuracy. So we think these miRNA gene relationships may have Functional MTI. However, further confirmation can only be done by experimental methods.

In addition, we think that the results show that there may be a very strong correlation between miRNAs and diseases so such a high level of accuracy can be

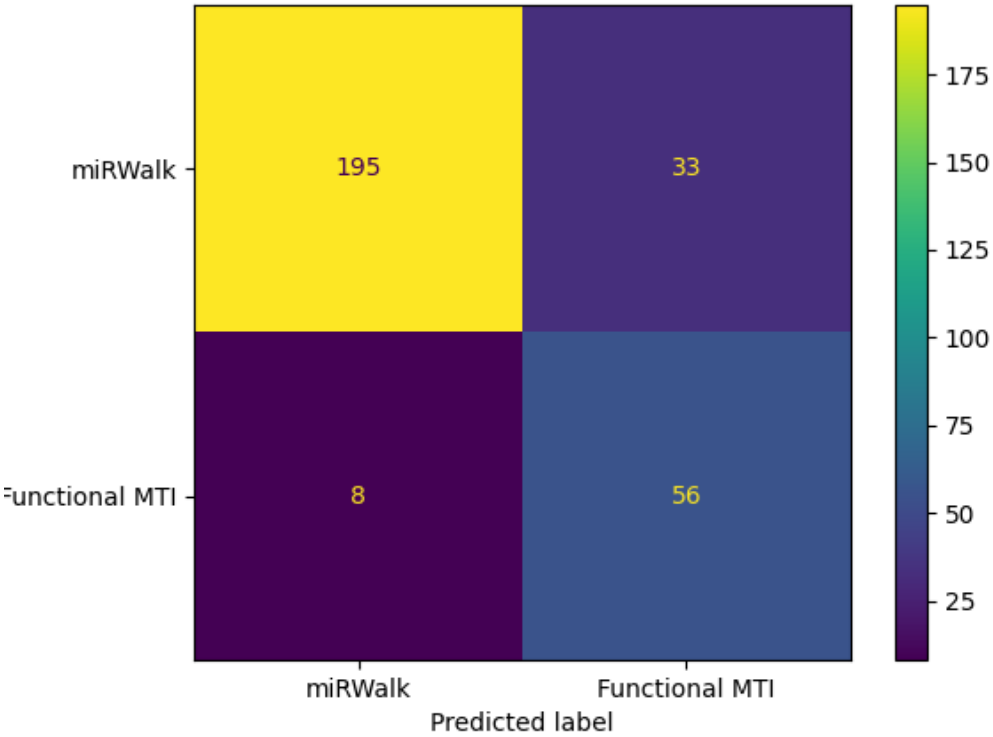

Figure 5: Confusion Matrix of Immunity- and Inflammation-related miRNA-gene Pairing Classification Result

achieved simply by calculating the semantic string similarity of diseases. However, limitations still exist. Although the method shows a consistent scoring trend in more than 2000 samples, more validation is needed to turn the present phenomenon into a general rule, including the validation of more genetic data.

5 Perspective

The Connection Building Method has promising scalability and can also be used to calculate the disease semantic similarity between other biomolecules as a way to explore the interactions between biomolecules. For example, except for the physical and chemical property prediction tools, the Connection Building Method can also provide data based on the disease semantic similarity dimension to optimize the prediction model. In addition, the method can also be used to calculate the functional semantic similarity between other biomolecules. Prospectively, we propose an application prospect whereby a large amount of miRNA-related disease data can be obtained by performing miRNA profiling on diseased tissue samples from clinical patients. A similar approach can also be used to obtain gene expression data. Using this method, it may be possible to build a prediction network with high accuracy, which is expected to provide strong support for future diagnosis and treatment.

More miRNA target predictions data based on the Connection Building Method using existing databases will be available soon.

Declarations

Funding

This research was partially funded by the Shenzhen X-institute ESRT Scholarship.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request. The scores for the 20 genes have been uploaded to the attachment. All the source code will be made publicly available after the paper is accepted by the journal.

Acknowledgments

I would like to thank Dr.Xiang for his valuable discussions and insights. My conversation with him inspired me to continue this project, which I had almost given up on. Thanks, Dr.Huang for letting me get to know about miRNA Target prediction, which is really fascinating. Thank you X-institute for changing my life. Last but not least, I want to thank the people around me, my family and friends, who have always supported me. My work wouldn't go so well without you guys.

Patent

A patent application related to the work presented in this paper has been filed Patent Application No.202410559682.3 Title: 基于文本的MicroRNA和基因关系挖掘方法、装置、设备及介质

References

[1] Rosalind C Lee, Rhonda L Feinbaum, and Victor Ambros. The c. elegans heterochronic gene lin-4 encodes small rnas with antisense complementarity to lin-14. *cell*, 75(5):843–854, 1993.

[2] Marieke Van Kouwenhove, Martijn Kedde, and Reuven Agami. Microrna regulation by rna-binding proteins and its implications for cancer. *Nature Reviews Cancer*, 11(9):644–656, 2011.

[3] Michael Carleton, Michele A Cleary, and Peter S Linsley. Micrornas and cell cycle regulation. *Cell cycle*, 6(17):2127–2132, 2007.

[4] Angie M Cheng, Mike W Byrom, Jeffrey Shelton, and Lance P Ford. Antisense inhibition of human mirnas and indications for an involvement of mirna in cell growth and apoptosis. *Nucleic acids research*, 33(4):1290–1297, 2005.

[5] Meredith Hill and Nham Tran. mirna interplay: mechanisms and consequences in cancer. *Disease models & mechanisms*, 14(4):dmm047662, 2021.

[6] Colin C Pritchard, Heather H Cheng, and Muneesh Tewari. Microrna profiling: approaches and considerations. *Nature Reviews Genetics*, 13(5):358–369, 2012.

[7] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[9] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[10] Yuhao Chen and Xiaowei Wang. mirdb: an online database for prediction of functional microrna targets. *Nucleic acids research*, 48(D1):D127–D131, 2020.

[11] Vikram Agarwal, George W Bell, Jin-Wu Nam, and David P Bartel. Predicting effective microrna target sites in mammalian mrnas. *elife*, 4:e05005, 2015.

[12] Anton Enright, Bino John, Ulrike Gaul, Thomas Tuschl, Chris Sander, and Debora Marks. Microrna targets in drosophila. *Genome biology*, 4:1–27, 2003.

[13] Michael Kertesz, Nicola Iovino, Ulrich Unnerstall, Ulrike Gaul, and Eran Segal. The role of site accessibility in microrna target recognition. *Nature genetics*, 39(10):1278–1284, 2007.

[14] A Petitjean, MIW Achatz, AL Borresen-Dale, P Hainaut, and M Olivier. Tp53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene*, 26(15):2157–2165, 2007.

[15] Nader Chalhoub and Suzanne J Baker. Pten and the pi3-kinase pathway in cancer. *Annual Review of Pathology: Mechanisms of Disease*, 4:127–150, 2009.

[16] Hua Yu, Drew Pardoll, and Richard Jove. Stats in cancer inflammation and immunity: a leading role for stat3. *Nature reviews cancer*, 9(11):798–809, 2009.

[17] Niki Karachaliou, Clara Mayo, Carlota Costa, Ignacio Magrí, Ana Gimenez-Capitan, Miguel Angel Molina-Vila, and Rafael Rosell. Kras mutations in lung cancer. *Clinical lung cancer*, 14(3):205–214, 2013.

[18] Anna Frenzel, Francesca Grespi, Waldemar Chmelewskij, and Andreas Villunger. Bcl2 family proteins in carcinogenesis and the treatment of cancer. *Apoptosis*, 14:584–596, 2009.

[19] Shang Yao, Lavender Yuen-Nam Fan, and Eric Wing-Fai Lam. The foxo3-foxm1 axis: A key cancer drug target and a modulator of cancer drug resistance. In *Seminars in cancer biology*, volume 50, pages 77–89. Elsevier, 2018.

[20] Shanshan Liu, Ziwei Ye, Vivian Weiwen Xue, Qi Sun, Huan Li, and Desheng Lu. Kif2c is a prognostic biomarker associated with immune cell infiltration in breast cancer. *BMC cancer*, 23(1):307, 2023.

[21] Zhixian Liu, Qingrong Sun, and Xiaosheng Wang. Plk1, a potential target for cancer therapy. *Translational oncology*, 10(1):22–32, 2017.

[22] Jing Han, Rongkai Xie, Ying Yang, Diangang Chen, Li Liu, Jiayang Wu, and Sufen Li. Cenpa is one of the potential key genes associated with the proliferation and prognosis of ovarian cancer based on integrated bioinformatics analysis and regulated by mybl2. *Translational cancer research*, 10(9):4076, 2021.

[23] Chi V Dang. Myc on the path to cancer. *Cell*, 149(1):22–35, 2012.

[24] sentence-transformers/multi-qa-minilm-l6-cos-v1 · hugging face.

[25] Gizem Soğancıoğlu, Hakime Öztürk, and Arzucan Özgür. Biosses: a semantic sentence similarity estimation system for the biomedical domain. *Bioinformatics*, 33(14):i49–i58, 2017.

[26] Hsi-Yuan Huang, Yang-Chi-Dung Lin, Jing Li, Kai-Yao Huang, Sirjana Shrestha, Hsiao-Chin Hong, Yun Tang, Yi-Gang Chen, Chen-Nan Jin, Yuan Yu, et al. mirtarbase 2020: updates to the experimentally validated microrna–target interaction database. *Nucleic acids research*, 48(D1):D148–D154, 2020.

[27] Carsten Sticht, Carolina De La Torre, Alisha Parveen, and Norbert Gretz. mirwalk: an online resource for prediction of microrna binding sites. *PloS one*, 13(10):e0206239, 2018.

[28] Zhou Huang, Jiangcheng Shi, Yuanxu Gao, Chunmei Cui, Shan Zhang, Jianwei Li, Yuan Zhou, and Qinghua Cui. Hmdd v3. 0: a database for experimentally supported human microrna–disease associations. *Nucleic acids research*, 47(D1):D1013–D1017, 2019.

[29] Janet Piñero, Àlex Bravo, Núria Queralt-Rosinach, Alba Gutiérrez-Sacristán, Jordi Deu-Pons, Emilio Centeno, Javier García-García, Ferran Sanz, and Laura I Furlong. Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research*, page gkw943, 2016.

[30] Mihaela Pertea and Steven L Salzberg. Between a chicken and a grape: estimating the number of human genes. *Genome biology*, 11:1–7, 2010.

[31] Julia Alles, Tobias Fehlmann, Ulrike Fischer, Christina Backes, Valentina Galata, Marie Minet, Martin Hart, Masood Abu-Halima, Friedrich A Grässer, Hans-Peter Lenhof, et al. An estimate of the total number of true human mirnas. *Nucleic acids research*, 47(7):3353–3364, 2019.

[32] Daniela N Mannel and Bernd Echtenacher. Tnf in the inflammatory response. *Chem Immunol*, 74(74):141–161, 2000.

[33] Toshio Tanaka, Masashi Narazaki, and Tadamitsu Kishimoto. Il-6 in inflammation, immunity, and disease. *Cold Spring Harbor perspectives in biology*, 6(10):a016295, 2014.

[34] John J Letterio and Anita B Roberts. Regulation of immune responses by tgf-β. *Annual review of immunology*, 16(1):137–161, 1998.

[35] Yuki Nakamura, Riho Aihara, Hisataka Iwata, Takehito Kuwayama, and Koumei Shirasuna. Il1b triggers inflammatory cytokine production in bovine oviduct epithelial cells and induces neutrophil accumulation via ccl2. *American Journal of Reproductive Immunology*, 85(5):e13365, 2021.

[36] Margarida Saraiva and Anne O'garra. The regulation of il-10 production by immune cells. *Nature reviews immunology*, 10(3):170–181, 2010.

[37] Yoichiro Iwakura, Susumu Nakae, Shinobu Saijo, and Harumichi Ishigame. The roles of il-17a in inflammatory immune responses and host defense against pathogens. *Immunological reviews*, 226(1):57–79, 2008.

[38] Xiaodong Wang, Yan Yiu Yu, Shirley Lieu, Frank Yang, Jeffrey Lang, Chuanyong Lu, Zena Werb, Diane Hu, Theodore Miclau, Ralph Marcucio, et al. Mmp9 regulates the cellular response to inflammation after skeletal injury. *Bone*, 52(1):111–119, 2013.

[39] Marco Del Giudice and Steven W Gangestad. Rethinking il-6 and crp: Why they are more than inflammatory biomarkers, and why it matters. *Brain, behavior, and immunity*, 70:61–75, 2018.

[40] Suprabhat Mukherjee, Subhajit Karmakar, and Santi Prasad Sinha Babu. Tlr2 and tlr4 mediated host immune responses in major infectious diseases: a review. *Brazilian Journal of Infectious Diseases*, 20:193–204, 2016.

[41] Fayyaz S Sutterwala, Stefanie Haasken, and Suzanne L Cassel. Mechanism of nlrp3 inflammasome activation. *Annals of the New York Academy of Sciences*, 1319(1):82–95, 2014.