Original Article

EyeTFDB: a Curated Eye Disease Transcription Factor Web-Based Platform

Mahdiye Ghorbani [1], PhD; Alireza Meshkin [2,*], PhD; Ehsan Pournoor [3,*], PhD

1. Laboratory of Systems Biology and Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran.
2. Department of Computer Engineering, Damavand Branch, Islamic Azad University, Damavand, Iran.
3. Basir Eye Health Research Center, Iran University of Medical Sciences, Tehran, Iran.

***Corresponding Authors:** Alireza Meshkin; Ehsan Pournoor
E-mail: alireza.meshkin@iau.ac.ir; e.pournoor@ut.ac.ir

Abstract

Background: More than one billion people with vision impairment imply comprehensive optimistic disease datasets necessity. Transcription factors play essential roles in many biological and biomedical applications. A publicly available health dataset is a practical resource for health researchers. To render a database of genomic eye diseases requires related Transcription Factors (TF) data. The TFs identify as distinct DNA sequences to regulate transcription processes and chromatin forming.
Material and Methods: To address needs in multiple areas, we present the EyeTFDB (https://eyetfdb.databanks.behrc.ir/) web-based platform, a novel platform of curated data. It combines the eye disease-related pathways data from the KEGG and REACTOME and their associated mRNAs. Furthermore, we expanded our mRNAs data by extracting disease-related data from valuable databases like DisGeNET and OMIM. Eventually, marked gained mRNAs as transcription factors by looking in humanTFDB, JASPAR, cisbp_DB, animal_TF, animal_TF_COF databases.
Results: Despite the high importance of eye diseases and the global need for further research in this area, still, there is no remarkable comprehensive database for transcription factors of eye diseases. Here we integrated multiple heterogeneous datasets from different databases addressing this need. We collected 429 mRNAs from 1258 pathways for 25 types of eye disease. In addition, transcription factors are separated from this considerable amount of data. EyeTFDB user interface is designed to facilitate multiple types of the desired use: (i) the interactive investigations of individual entries on the level of a gene, transcription factor, pathway, eye disease, and (ii) the construction of highly customized datasets using advanced searching and filtering.
Conclusion: The status of eye TFs and their broad interactions could significantly influence eye health. EyeTFDB, as well as other established databases serve as rich resources for assisting the researchers in exploring eye TFs in the elevation of public health. The wealthy information generated from future investigations can be incorporated into EyeTFDB for better serving the eye TF research and exploration efforts.
Keywords: Eye; Optimistic Disease; Web-Based Platform; Transcription Factors (TFs), Transcriptional Cofactors (COFs).

Article Notes: Received: Feb. 27, 2020; Received in revised form: Apr. 1, 2020; Accepted: May. 20, 2020; Available Online: Sep. 22, 2020.

How to cite this article: Ghorbani M, Meshkin A, Pournoor E. EyeTFDB: a Curated Eye Disease Transcription Factor Web-based Platform. Journal of Ophthalmic and Optometric Sciences . 2020;4(4):1-9.

This work is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Introduction

Eye wellbeing is defined as ocular and functional health and consists of personal and social life aspects. Hence, focusing on eye disease becomes essential when considering sociodemographic status increment [1]. In 2020, there are an estimated 43 million blind people; a further 1.1 billion people have vision impairment, including 510 and 596 million who had uncorrected near and distance vision impairments, respectively. From these, it is determined that 295 million had upper than moderate and 258 million people with mild distance vision impairment. The latest prevalence figures imply 61 million blinds up to 2050 [2].

Fortunately, there were about one hundred open access eye health data sets up to 2020. It is one of the reasons for the 28.5 % global blindness decrease from 1990 up to 2020; also, as a fall of 25 % in vision impairment from 2010 to 2020. As concerned, many causes of vision impairment can be prevented or treated, so the development of validated genomic databases has been essential in the promotion of ophthalmological research [3].

To render a database of optimistic genomic diseases requires related Transcription Factors (TF) data. The TFs are a kind of proteins that interact with specific sequences of DNA with their DNA-Binding Domains (DBDs) and consequently tend to suppress or promote DNA transcription [4]. They play a significant role in biological processes by recognizing their related TF Binding Sites (TFBSs) detected at non-coding cis-regulatory regions like promoters, enhancers, and silencers [5]. Thereby, these TFBSs can be in, or nearby of a gene and regulate the transcription machine's assembly, leading to controlling gene expression.

The TFs can accomplish many functions such as gene selection, master regulators, cell types specifying, trans-differentiation, de-differentiation, and controlling pathways like immune responses. Also, their protein sequences, physiological roles, and DNA-binding domains are often profoundly conserved among Animals. By itself, it implies the conservation of gene regulatory networks [6]. just one TF can regulate various genes in diverse cell types, leading to dynamic regulatory networks, such as ESR1 in endometrial and breast cell lines [7]. The curated TFs can achieve by the DNA BINDING ASSAYS technique based on interaction with specific DNA sequences of enhancers and promoters. Employing reporter fusions like GFP or Luciferase is another method. Likewise, the electrophoresis mobility shift assay (EMSA) technique is a popular method for studying DNA-protein interactions; it is based on the principle that the DNA-protein complex has a slower migration on gel electrophoresis than free linear DNA [8]. Parallelly, in silico approaches can aim to the prediction of TFs. Overall, various pattern-matching and scanning specific DNA sequences with a position weight matrix (PWM) can be utilized [9-11].

Nowadays, an estimate indicates that just one percent of TFs are known despite their importance in regulatory activity [12-14]. As well, there are not very many databases for TFs that focus on a specific organ, especially the eyes. For instance, the current most usable was established by the JASPAR and TRANSFAC. However, some others, like the Human TFs and the AnimalTFDB do not focus on the organ filter feature, and some others, such as TF-Marker, have low resources for eyes [4, 15-17]. Nc2eye is another database that includes ncRNA of the eye but does not cover transcription factors [18].

In the present study, we utilized JASPAR,

This work is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Human TFs, Animal TFDB, and cisbp database to retrieve TFs and transcriptional cofactors (COFs). The COFs are cofactor proteins that bind to the DNA binding domain of TFs, so they play a significant role in regulating them [19].

Material and Methods

Data collection and preparation

A brief blueprint of the data collection and preparation of the web-based platform is illustrated in figure1. Accordingly, the flow of data collection was as follows:

Step 1: At first, we got (January 3, 2022) 25 Eye diseases from MESH (Medical Subject Headings) database (https://www.ncbi.nlm.nih.gov/mesh). MESH is the NLM-controlled vocabulary thesaurus used for indexing articles for PubMed. The retrieved keywords include: "cataract", "diabetic retinopathy (DR)", "myopia", "primary open-angle glaucoma (POAG)", "uveal melanoma (UM)", "retinal neovascularization", "corneal neovascularization", "retinoblastoma (RB)", "proliferative vitreoretinopathy (PVR)", "Exfoliation syndrome (XFS)", "primary Sjögren's syndrome (pSS)", "age-related cataract", "neuromyelitis optica (NMO)", "pterygium", "glaucoma", "Behcet's disease (BD)", "posterior capsule opacification (PCO)", "age-related macular degeneration (AMD)", "Vogt-Koyanagi-Harada disease", "strabismus", "keratoconus", "diabetic cataract", "herpes simplex virus keratitis", "retinitis pigmentosa (RP)" and, "Volkmann cataract".

Step 2: KEGG pathway and REACTOME are online databases for understanding high-level functions of biological systems using molecular data. As mentioned, eye disease keywords were searched in KEGG (January 3, 2022) PATHWAY database (https://www.genome.jp/kegg/pathway.html) [20, 21] and REACTOME (January 3, 2022) (https://reactome.org/). So, 1258 related pathways were extracted totally.

Step 3: Each fetched pathway was searched in KEGG and REACTOME for extracting associated curated mRNAs. Entirely 429 mRNAs (January 3, 2022) were retrieved.

Step 4: because of the importance of data reliability, we choose three important gene-disease databases, DisGeNET, OMIM, and KEGG, as another source of data integration. DisGeNET is a discovery platform containing one of the largest publicly available collections of genes and variants associated with human diseases. Online Mendelian Inheritance in Man (OMIM) is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. Every gathered mRNA was searched for collecting related information in DisGeNET (https://disgenet.org/), OMIM (https://www.omim.org/), and KEGG (January 3, 2022).

Step 5: In addition, gained mRNAs were searched (January 3, 2022) in NCBI (https://www.ncbi.nlm.nih.gov/gene/) ENSEMBLE (https://www.ensembl.org/), GENEBANK (https://www.ncbi.nlm.nih.gov/genbank/) for getting their sequences, ENTREZIDs and, descriptions.

Step 6: Eventually, accumulated mRNAs were searched (January 3, 2022) in JASPAR (https://jaspar.genereg.net/), humanTFDB (http://bioinfo.life.hust.edu.cn/HumanTFDB/), cisbp_DB (http://cisbp.ccbr.utoronto.ca/), animal_TF (https://www.deviantart.com/tag/animaltf) and, animal_TF_COF (https://pubmed.ncbi.nlm.nih.gov/28110180/) to determine the transcription factors among them. In the end, 97 curated transcription factors were identified.

This work is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

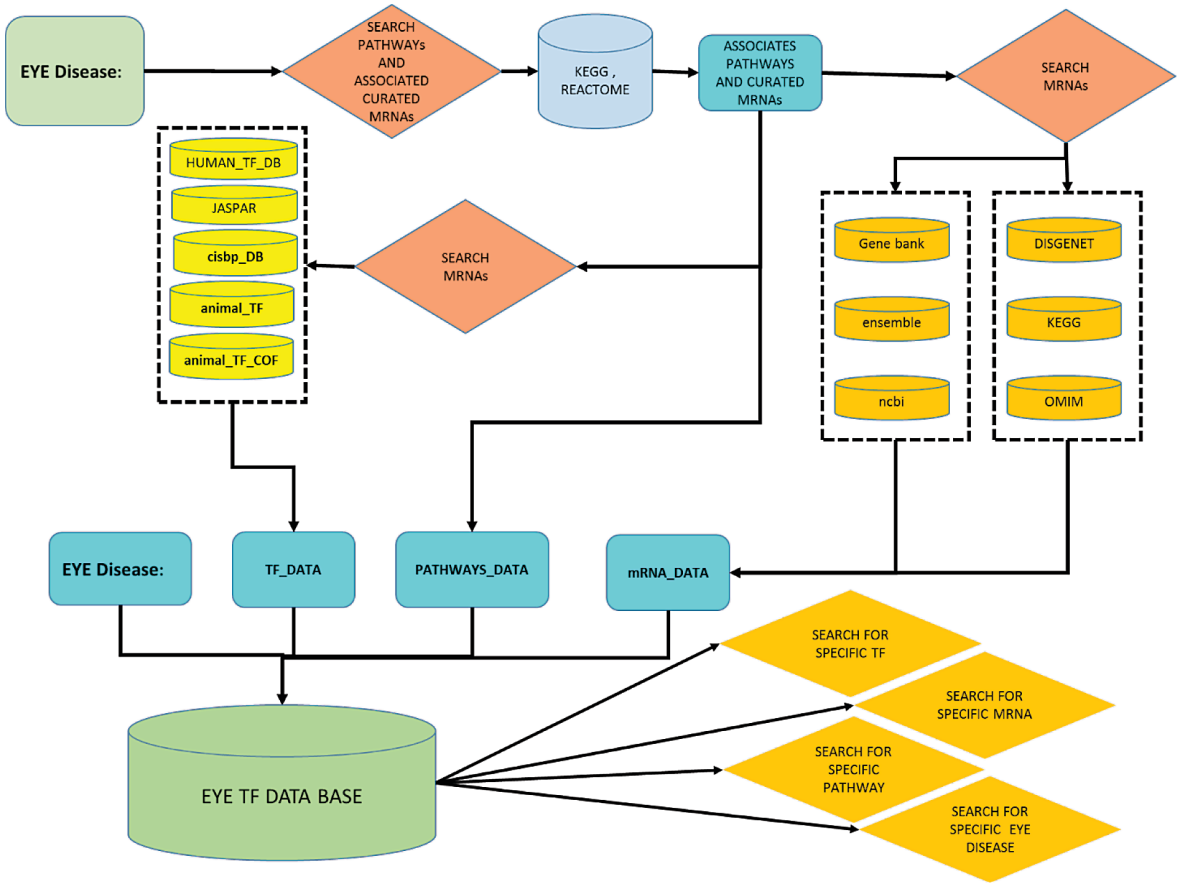


Figure 1: Data collection flowchart

Web portal

The EyeTFDB is created using the ASP.NET MVC framework. On the server-side, it is developed by C# on the server-side and the client-side, javascript, HTML5, and bootstrap. Its database technology is Microsoft SQLServer with the LINQ (ORM) and EntityFramework. The homepage of EyeTFDB is demonstrated in Fig. 2.

It is easy to research ocular disorders from several viewpoints throughout these three boxes of Eye Diseases, Genes, and Search TF. In the Eye Disease section (Figure 3), all eye diseases are presented with the possibility of a detailed review.

As stated formerly, the detailed info contains disease title, description, anchors to other bases, and its involved target genes, which can be viewed in detail (shown in figure 4).

Moreover, the search option contains the possibility of gene and Transcription factor search. As revealed in Fig. 5, inside the EyeTFDB, we provided the capability to search a TF using its Name, Ensembl ID, or Entrez ID.

Results

The integrated data consists of 429 mRNAs, 25 diseases, 1258 pathways, and 88 transcription factors. Fig. 6 illustrates the frequency of mRNAs among three reference databases, KEGG, OMIM, and DisGeNET. Some mRNAs are shared between two or three of these databases. Plot 7 shows the number of mRNAs per disease. Retinitis pigmentosa (RP) has the most mRNA frequency among these eye diseases.

The TFs part of the current platform contains

This work is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

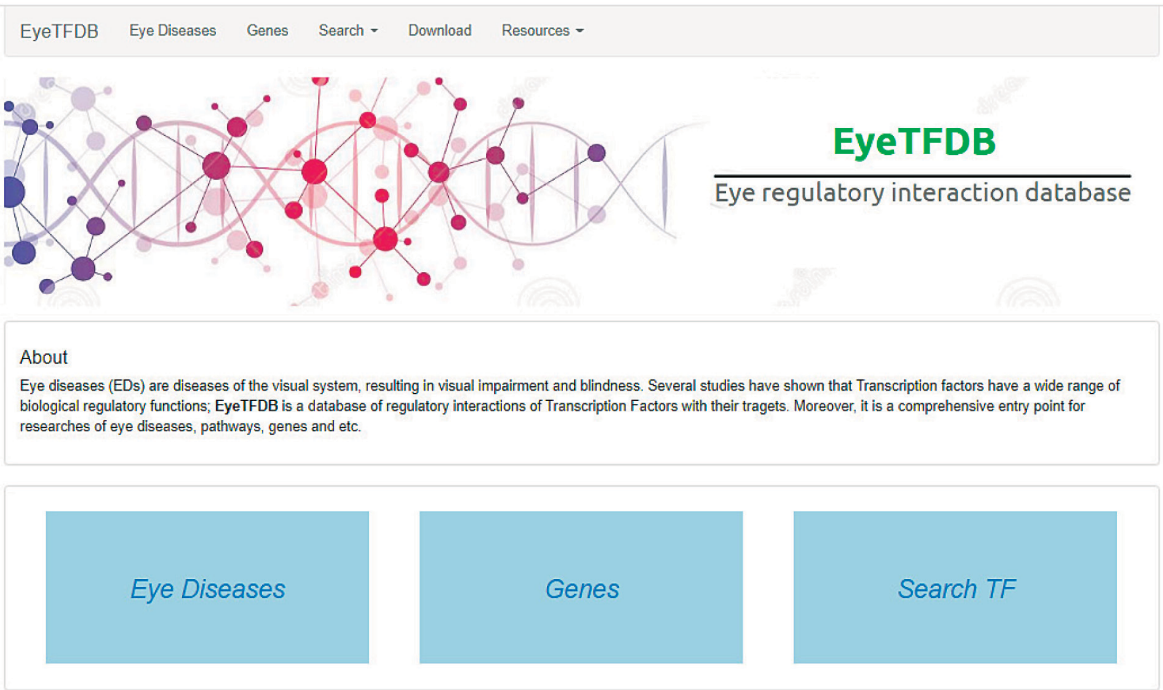
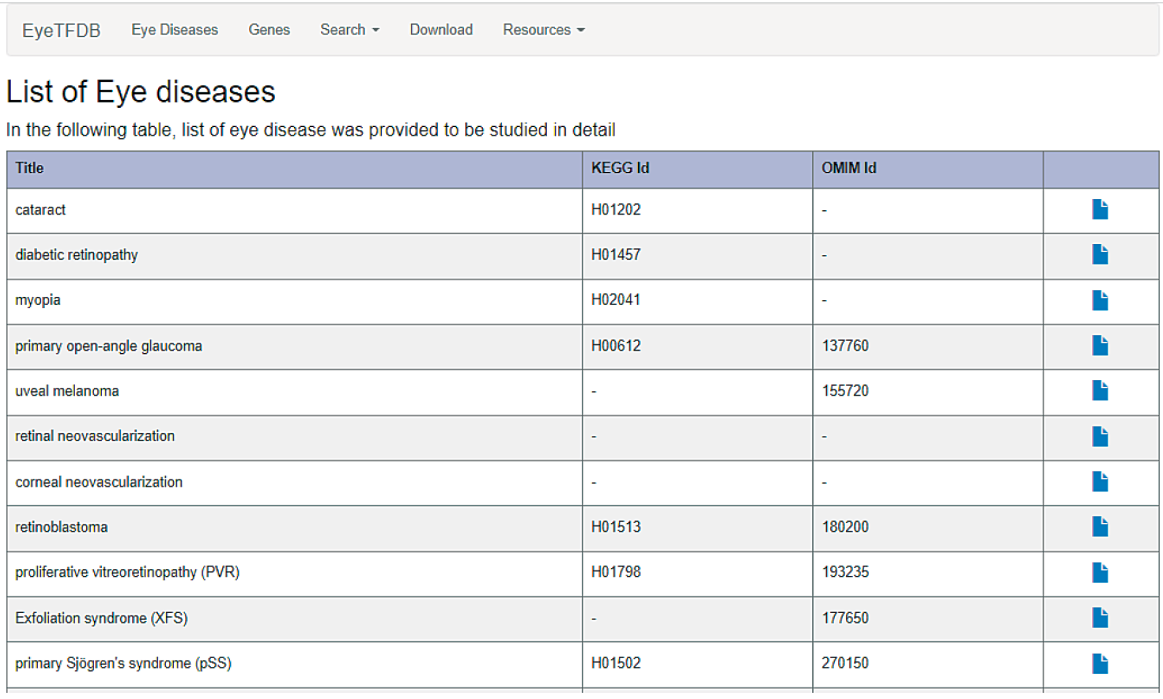


Figure 2: Home page of EyeTFDB

specific TFs whose human eye disease risk may alter their binding. One current challenge comprises addressing redundancy and synergy among multiple regulatory elements of a target gene. Predicting enhancer and promoter contacts leads to interpreting different epigenetic memory types. The current web-based platform provides valuable resources for



Title	KEGG Id	OMIM Id	
cataract	H01202	-	
diabetic retinopathy	H01457	-	
myopia	H02041	-	
primary open-angle glaucoma	H00612	137760	
uveal melanoma	-	155720	
retinal neovascularization	-	-	
corneal neovascularization	-	-	
retinoblastoma	H01513	180200	
proliferative vitreoretinopathy (PVR)	H01798	193235	
Exfoliation syndrome (XFS)	-	177650	
primary Sjögren's syndrome (pSS)	H01502	270150	

Figure 3: Eye Diseases list. For each row, the detailed info is delivered

This work is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

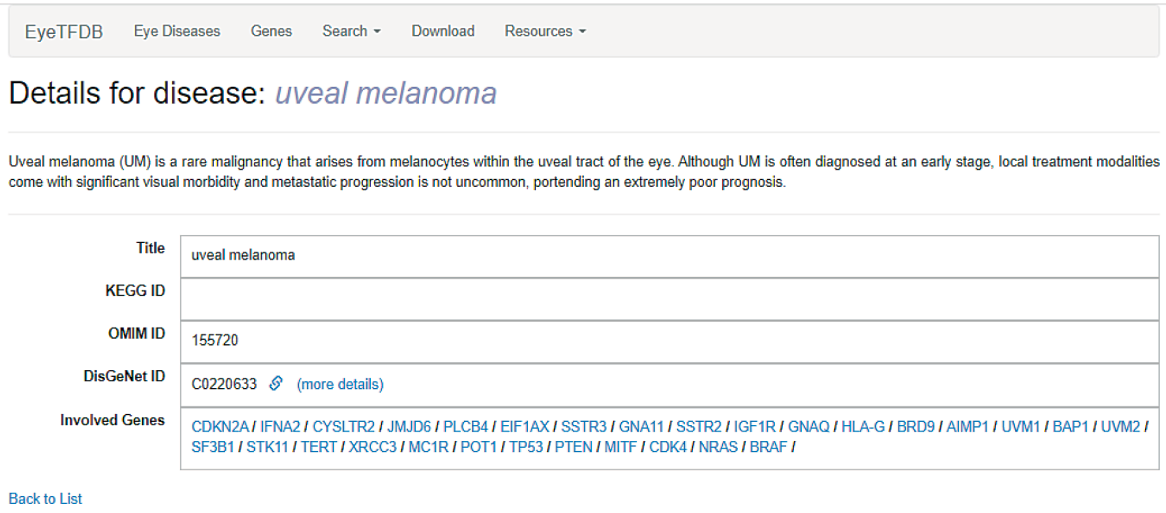


Figure 4: Disease Search Result page

ophthalmological scientists and researchers in human genetics who want to decode what TFs do. Also, the Protein-Protein Interactions (PPIs) network will be practical for a user to investigate TFs and COFs targets and their regulatory network.

Discussion

In comparison with other databases, EyeTFDB is an eye-specific TF-Target interaction data portal, which is a good entry point for those that start their research on eye diseases. We provided the capabilities to start from disease, gene, transcription factor, and pathway or search these entities and reach the other related information. In this case, links to original databases are also provided that aims scientific people to access more information on that topic.

Conclusion

Eye transcription factors play essential roles in

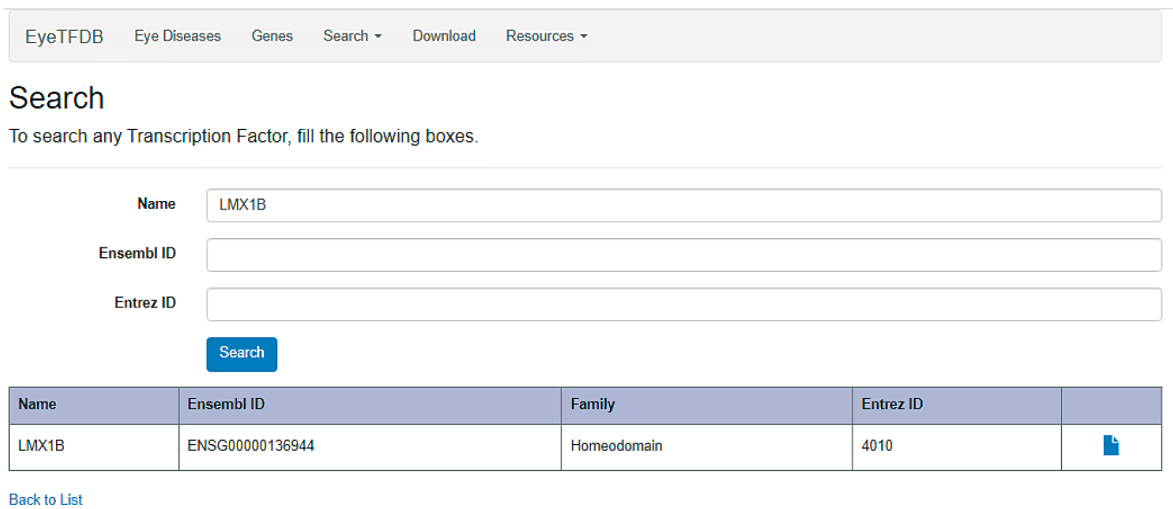


Figure 5: TF search page. Here, looking for miRNAs is provided through their Name, Ensembl ID, and Entrez ID

This work is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

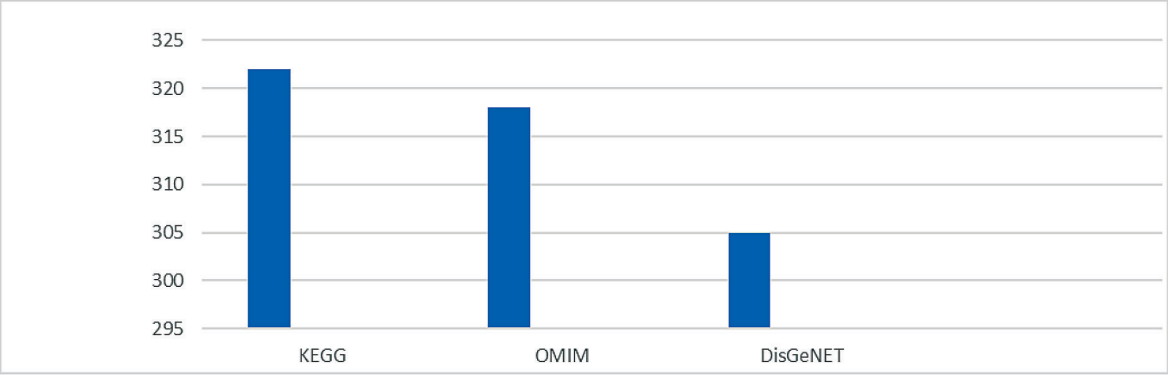


Figure 6: number of mRNAs per web-based platform. Are not disjoint and have overlap in mRNAs

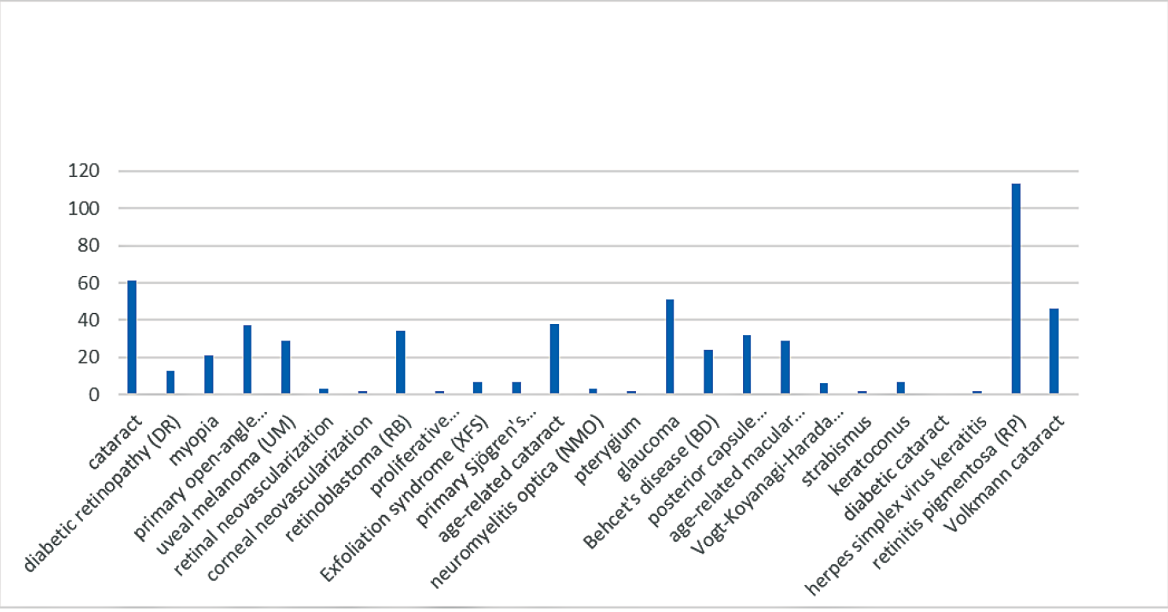


Figure 7: number of mRNAs per disease

human eye health, and their malfunction may lead to severe health problems. The status of eye TFs and their broad interactions could significantly influence eye health. EyeTFDB, as well as other established databases, serve as rich resources for assisting the researchers in exploring eye TFs in the elevation of public health. The wealthy information generated from future investigations can be incorporated into EyeTFDB for better serving the eye TF research and exploration efforts. We aim to update the newly emerging information into EyeTFDB regularly.

Authors ORCIDs

Mahdiye Ghorbani:
https://orcid.org/0000-0003-0237-1537
Ehsan Pournoor:
https://orcid.org/0000-0001-7609-1816
Alireza Meshkin:
https://orcid.org/0000-0002-7046-4404

This work is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Reference:

1. Burton MJ, Ramke J, Marques AP, Bourne RR, Congdon N, Jones I, et al. The Lancet global health Commission on global eye health: vision beyond 2020. The Lancet Global Health. 2021;9(4):e489-e551.

2. Bourne R, Steinmetz JD, Flaxman S, Briant PS, Taylor HR, Resnikoff S, et al. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. The Lancet global health. 2021;9(2):e130-e43.

3. Khan SM, Liu X, Nath S, Korot E, Faes L, Wagner SK, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. The Lancet Digital Health. 2021;3(1):e51-e66.

4. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The human transcription factors. Cell. 2018;172(4):650-65.

5. Smith NC, Matthews JM. Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. Current opinion in structural biology. 2016;38:68-74.

6. Weirauch MT, Hughes TR. Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. Trends in genetics. 2010;26(2):66-74.

7. Gertz J, Reddy TE, Varley KE, Garabedian MJ, Myers RM. Genistein and bisphenol A exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. Genome research. 2012;22(11):2153-62.

8. Yang VW. Eukaryotic transcription factors: identification, characterization and functions. The Journal of nutrition. 1998;128(11):2045-51.

9. Garcia-Alcalde F, Blanco A, Shepherd AJ. An intuitionistic approach to scoring DNA sequences against transcription factor binding site motifs. BMC bioinformatics. 2010;11(1):1-13.

10. Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, Masoudi-Nejad A. DrugR+: a comprehensive relational database for drug repurposing, combination therapy, and replacement therapy. Computers in biology and medicine. 2019;109:254-62.

11. Ahmadi H, Ahmadi A, Azimzadeh-Jamalkandi S, Shoorehdeli MA, Salehzadeh-Yazdi A, Bidkhori G, et al. HomoTarget: a new algorithm for prediction of microRNA targets in Homo sapiens. Genomics. 2013;101(2):94-100.

12. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. Cell. 2014;158(6):1431-43.

13. Najafi A, Bidkhori G, H Bozorgmehr J, Koch I, Masoudi-Nejad A. Genome scale modeling in systems biology: algorithms and resources. Current genomics. 2014;15(2):130-59.

14. Ghasemi M, Seidkhani H, Tamimi F, Rahgozar M, Masoudi-Nejad A. Centrality measures in biological networks. Current Bioinformatics. 2014;9(4):426-41.

15. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Berhanu Lemma R, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic acids research. 2022;50(D1):D165-D73.

16. Wingender E, Dietze P, Karas H, Knüppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic acids research. 1996;24(1):238-41.

17. Hu H, Miao Y-R, Jia L-H, Yu Q-Y, Zhang Q, Guo A-Y. AnimalTFDB 3.0: a comprehensive

This work is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

resource for annotation and prediction of animal transcription factors. Nucleic acids research. 2019;47(D1):D33-D8.
18. Zhang Y, Xue Z, Guo F, Yu F, Xu L, Chen H. Nc2Eye: a curated ncRNAomics knowledgebase for bridging basic and clinical research in eye diseases. Frontiers in Cell and Developmental Biology. 2020:75.
19. Haberle V, Arnold CD, Pagani M, Rath M, Schernhuber K, Stark A. Transcriptional cofactors display specificity for distinct types of core promoters. Nature. 2019;570(7759):122-6.
20. Masoudi-Nejad A, Goto S, Endo TR, Kanehisa M. KEGG bioinformatics resource for plant genomics research. Plant Bioinformatics: Springer; 2007. p. 437-58.
21. Masoudi-Nejad A, Goto S, Jauregui R, Ito M, Kawashima S, Moriya Y, et al. EGENES: transcriptome-based plant database of genes with metabolic pathway information and expressed sequence tag indices in KEGG. Plant Physiology. 2007;144(2):857-66.

Footnotes and Financial Disclosures

Conflict of interest:

The authors have no conflict of interest with the subject matter of the present manuscript.

This work is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)