Abstract text mining to create an exhaustive disease-disease correlation database

Suchir Misra

Abstract

Craniosynostosis, the second most common craniofacial abnormality, shares genetic mutations implicated in cancer progression, yet a correlation between the two has not been elucidated. Disease-disease correlations can assist in developing improved disease treatments, yet finding genetic loci used to establish such correlations is expensive and time-consuming. Databases enhance visualization of disease-disease correlations and disease-gene associations; however, current databases overlook rare diseases and important connections by limiting the pool of diseases studied. A computational approach was designed to create a database of disease-disease correlations such that correlations with rare diseases could be elucidated. Python programs were written to collect a list of abstract IDs for all genetic papers related to an extensive list of diseases (N = 1857), to sort the abstract IDs numerically and remove duplicates, and to extract gene names from the abstracts. A PostgreSQL database was used to store the data for efficient querying. Disease-disease correlations were determined based on gene overlaps. The top ten disease-disease connections overall have been previously elucidated, validating the effectiveness of the method used to create the database. Of the top ten disease-disease connections for craniosynostosis, four were newly elucidated. In the future, publications should denote mutation percentages of genes in their abstracts so the importance of genes mutated in a disease can be considered in future iterations of the program. This study provides a tool to find genetic loci and design improved disease treatments for both rare and common diseases.

Introduction

Craniosynostosis, the second most common craniofacial abnormality, is caused by premature fusion of one or more cranial sutures [1]. Over 180 types of craniosynostosis exist and can present as a syndromic disorder or an isolated, nonsyndromic disorder [1], [2]. Syndromic craniosynostosis makes up about 15% of all cases and presents with a variety of additional clinical symptoms including malformations of the hands and feet, as well as cardiovascular defects, while nonsyndromic craniosynostosis is associated with neurodevelopmental disorders but both have poorly identified genetic underpinnings [1], [3].

Genetic factors play a role in causing both types of abnormal suture fusion [1]. Syndromic craniosynostosis has been associated with specific gene mutations, and 20% of its cases are caused by frequently repeated mutations in Fibroblast Growth Factor Receptor 2 (FGFR2), Fibroblast Growth Factor Receptor 3 (FGFR3), and Twist Basic Helix-Loop-Helix Transcription Factor 1 (TWIST1) [2]. Conversely, nonsyndromic craniosynostosis is a multifactorial disorder caused by the interaction of genetic and environmental factors [1]. However, in recent years, whole exome sequencing (WES), whole genome sequencing (WGS), and genome-wide association studies (GWAS) have begun to suggest stronger influences between genetic variants and nonsyndromic cases [4]–[7]. Notable mutations include Bone Morphogenetic Protein (BMP) mutations associated with neural crest derived tissue and apoptosis in bones, mutations in hedgehog and TWIST1 associated with Axin2 (a stem cell marker) which have a role in maintaining cranial suture function, and mutations in the Wnt signaling pathway which cooperate with FGFR disturbances [1]. Both the Wnt signaling pathway and the BMP signaling pathway regulate cellular proliferation and differentiation, and when dysregulated, may cause a number of other diseases, including osteoarthritis, hypertension, and cancer [8], [9]. Craniosynostosis shares additional genetic similarities with cancer, as mutations in FGFRs and TWIST1 are common to both [10], [11].

FGFR2 and FGFR3 are connected to craniosynostosis and cancer through the Extracellular Signal-Regulated Kinase (ERK)-Mitogen Activated Protein Kinase (MAPK) pathway [12], [13] A mutation in either FGFR2 or FGFR3 causes a conformational change in the receptor, causing activation of Fibroblast Growth Factor Receptor Substrate 2 (FRS2) and Growth Factor Receptor Bound Protein 2 (GRB2), which activates a Son of Sevenless (SOS) protein, to activate a protein from the RAS subfamily, that activates the Rapidly Accelerated Fibrosarcoma (RAF) protein, to activate MAPK then ERK. The ensuing gene transcription leads to excessive differentiation, such that undifferentiated suture cells differentiate into osteoblasts to form bone. Excessive activation of the ERK-MAPK pathway also leads to excessive cellular proliferation and enhanced cell survival, again observing additional characteristics with cancer [13]. TWIST1 mutations also cause craniosynostosis because excessive differentiation causes the sutures to change into bone. However, the mechanism by which TWIST1 is connected to cancer differs, as TWIST1

is an important gene for the process of epithelial-mesenchymal transition, which causes cells to lose target site recognition [10], [14], [15].

Craniosynostosis and cancer have similar molecular underpinnings despite having different pathologies. Correlations of diseases with similar molecular underpinnings are desirable because disease-disease correlations are typically drawn based on phenotypic similarities, leading to diseases with different etiologies but similar treatment modalities [16]. Correlating diseases based on shared molecular mechanisms, such as craniosynostosis and cancer, allows for improvement of disease therapies through drug target identification and reposition of existing drugs while also advancing the fields of disease taxonomy and etiology [16]–[21]. Building a comprehensive database of disease-disease correlations based on similar molecular mechanisms would be beneficial [16].

Current databases of disease-disease correlations have reduced benefits as a lack of expression data causes rare diseases, such as craniosynostosis to be overlooked (Table 1). Using literature searches to build databases would remediate this issue as there is a breadth of literature available concerning rare diseases. Automating the literature search would expedite the process and remove inefficiencies that come with manually searching literature. Text mining is the optimal method to automate literature searches as it is the most efficient way to extract information from publications [22]. Text mining of abstracts rather than full papers allows for a wider range of studies to be accounted for, due to limited accessibility of full text papers [23]. The use of abstracts also allows for only the most important genes to be considered, since abstracts only include the most pertinent information. There are no current disease-disease databases that exclusively use text mining of abstracts to elucidate disease-disease correlations.

Table 1. Limitations of current disease-disease databases in studying craniosynostosis. Three of the most well-known disease-disease correlation databases that used shared genes to form their correlations are limited in their study of craniosynostosis. DNetDB doesn't include it, MalaCards's gene list is not exhaustive, while DisGeNET dilutes its gene list by returning genes not related to it. N = number of craniosynostosis-mutated genes as per database.

Database Name	Total Number of Diseases	Craniosynostosis Genes	Data Collection Method	Limitation
DNetDB	108	0	Gene co-expression analysis	Limited amount of co-expression data – less diseases studied/Microarrays used
MalaCards	16,919	31	Database Integration	Leaves out significant number of genes - manual search finds more
DisGeNET	15,093	213	Database Integration/ Text Mining	Errors in text mining - includes genes mutated in CRS(Chronic Rhinosinusitis)/Microarrays used

While investigating limitations associated with disease-disease correlations in commercially available databases, this study attempted to elucidate disease-disease correlations in an author designed database by: collecting a list of common diseases and conditions, collecting a list of abstracts for genetic papers for each disease, extracting gene names from all the abstracts on the abstract list, and then elucidating disease-disease correlations by calculating the significance of the gene overlap between all disease combinations. A connection between lesser known and well-known disorders could support development

of disease-gene correlations for drug target development based upon disease-disease correlations that improve new drug target identification.

Methodology

Creating a Disease List

The Mayo Clinic website's collection of alphabetized diseases and conditions was used to create a disease list, where duplicate diseases were removed. The completed disease list consisted of 1857 distinct diseases, and the word "genetics" was added to the end of each disease name to form a list of PubMed search terms. PubMed uses the Medical Subject Headings (MeSH), terms arranged in an online hierarchy, to curate its search function [39]. MeSH terms are manually assigned by employees of the National Library of Medicine to each new publication [39]. PubMed expands the hierarchy under each MeSH term in the search term, so all related papers under that hierarchical term will be returned [39]. The presence of the term "genetics" in the search term ensures that population genetics, human genetics, genomics, and proteomics publications will all be returned. These types of publications were necessary to build the database as they would include gene names in their abstracts.

Collecting Abstract IDs

To collect abstract IDs, a Python program was written such that each of the search terms on the search term list was searched on PubMed and the PubMed IDs for all of the returned publications were stored and filed with their corresponding search term (Table 2). To ensure restartability of the ensuing program, the IDs were copied into a separate file, duplicate IDs were removed, and the remaining IDs were sorted in ascending numerical order using an author-designed Python program (Table 3). After sorting, the number of IDs was reduced from 6,393,179 to 2,056,144. The list of sorted IDs was then used to extract gene names.

Table 2. Input and output of program used to extract PubMed IDs for all search terms in search term list. Max of 100,000 in program ensures that all IDs related to that search term will be written to an output file. Output file has two columns with the first being the disease name, and the second being the ID

Input	Output	
Get_abstract_ids(Disease 1)	Disease 1	36
	Disease 1	5
	Disease 1	5
	Disease 1	8
	Disease 1	25

Extracting Genes from Abstracts

Each of the abstracts was parsed for gene names using an author-designed Python program (Figure 1.). Due to the large number of abstracts that had to be parsed, the program was designed to be restartable. As each abstract was processed, the ID was written to a "Last Processed ID" file. If the program crashed, a binary search was conducted to find the position of the last processed ID, so the program could restart from there. Gene names were extracted by matching to a list of all gene names collected from the HUGO Gene Nomenclature Committee (HGNC) [40]. To ensure that only relevant genes were returned as an output, a file of genetic terms was created with terms ranging from "variant" to "WES" (Appendix 1). The gene name would only be returned if one of these genetic terms was within three words of the gene. There was a total of 218,436 non-distinct genes returned from all of the abstracts.

Table 3. Sample input and output of Python program used to sort all abstract IDs. IDs copied into separate file, where list is first sorted numerically. All duplicated IDs then removed, so file only contains distinct abstract IDs.

Input	Output
Sort_abstract_ids(36,5,5,8,25)	5 8 25 36

Database Creation

The data was loaded into a PostgreSQL database for efficient querying [41]. The completed database included the lists of all diseases, diseases to abstract IDs, sorted abstract IDs, and abstract IDs to genes. A table correlating disease to gene name was created by joining the table of diseases to abstract IDs with the table of abstract IDs to genes. The disease to gene table had a total of 866,903 disease-gene associations.

```
try:
    abstracts = fetch.article_by_pmid(pmid).abstract
    abstract_string = str(abstracts)
    gene_punctuation_remove = (abstract_string.replace('(', '
        ').replace(')', ' ').replace('-', ' ').replace('/', '
        ').replace(',',' '))
    gene_split = (gene_punctuation_remove.split())
    gene_overlap = (set(gene_split) & set(gene_list))
    gene_overlap_list = []
    gene_overlap_list.extend(gene_overlap)
    for gene in gene_overlap_list:
        indices = [i for i, x in enumerate(gene_split) if x == gene]
        for i in indices:
            overlap_indices = gene_split[i - 3:i + 4]
            if (str(set(overlap_indices) & set(genetic_terms))) !=
                'set()':
                absToGeneFile.write(pmid + "\t" + gene + "\n")
                break
```

Figure 1. Python program used to extract gene names from each abstract on the sorted ID list. All punctuation removed from abstracts so genes could be matched from abstract to list of all gene names. Gene name would only write to output file if it was within three words of a "genetic term". Output file had two columns – abstract ID and gene.

Statistical Analysis

With all of the data loaded into the database, gene lists were created for each disease by querying the disease-gene table. These gene lists were stored in a file and were transformed to one large dictionary. Of the 1857 diseases in the original disease list, only 1648 of them had genes associated with them, thus calculations were only performed for combinations between these diseases. To find disease-disease connections, the same program that was used to correlate craniosynostosis-mutated genes to MSigDB gene sets was used, and a Bonferroni Correction was performed to establish a threshold p-value (Figure 5). The

results of the calculations were stored in an output file with four columns: disease 1, disease 2, p-value, and odds ratio.

User Interface Creation

A user interface using various JavaScript tools was created following protocol as established by Uttariello, 2019 [27]. The interface was created to query the results from the output table from PostgreSQL for a user's chosen input disease.

Disease Network Creation

A disease network was created in order to cluster diseases as another metric to elucidate novel disease-disease relationships. R's multidimensional scaling function was used to plot a matrix of author-derived similarity scores for each disease-disease correlation.

Results/Discussion

Database Validation

A Python program was written to perform a Fisher's Exact Test on all disease-disease combinations. The output file contained the name of each disease, the p-value and the odds ratio. When sorted by p-value, the combination of prostate cancer and male breast cancer had the lowest p-value, and the combinations of colorectal cancer and male breast cancer and heart disease and arthritis had the second and third lowest p-values (Table 4).

Table 4 - Ten disease combinations with lowest p-values overall. Of the 1357128 disease-disease combinations, the ten with the most significant p-values are shown. P-values were calculated using Python program that performed Fisher's Exact Test. All combinations validate accuracy of data collection method used to compile database

Disease 1	Disease 2	P-value	Shared Genes	Major Shared Genes	Pathological Connection
Prostate Cancer	Male Breast Cancer	$4.9407*10^{-324}$	254	NF2, MXI1	Result from uncontrolled cell division
Colorectal Cancer	Male Breast Cancer	$9.8813*10^{-324}$	277	MTRR, NF2	Result from uncontrolled cell division
Heart Disease	Arthritis	$9.8813*10^{-324}$	455	THRA, MTRR	Crucial role of inflammation
Heart Attacks	Hypertension	$3.9525*10^{-323}$	329	MTRR, VKORC1	Cardiovascular system disorders
Cardiac Ischemia	Obesity	$4.4466*10^{-323}$	364	THRA, MTRR	Involve muscle weakening
Dementia	Self-Injury	$1.4328*10^{-322}$	463	FOXP2, CLN8	Neurological disorders
Anemia	Liver Disease	$1.8280*10^{-322}$	371	TFR2, NPM1	Involve nutritional deficiencies
Cirrhosis	Erectile Dysfunction	$5.8794*10^{-322}$	477	AQP2, POLD1	Involve hormonal dysfunction
Lymphoma	Breast Cancer	$8.4979*10^{-322}$	376	RAD50, CCR2	Result from uncontrolled cell division
Chronic Kidney Disease	Horner-Bernard Syndrome	$4.0909*10^{-321}$	390	GNB3, FSTL1	Involve nerve damage

All of the disease combinations with the ten lowest p-values validate the accuracy of the data collection method used to compile the disease-disease connection database [28]–[39]. Over 200 genes were shared between each of the diseases in Table 1, demonstrating that the significance of the connections was not due to a small sample size. Since the data collection method is validated, the database can be used to identify novel disease-disease correlations.

Database Reveals Novel Craniosynostosis Connections

When sorting all craniosynostosis combinations by p-value, the combination of craniosynostosis and synostosis had the lowest p-value, the combination of craniosynostosis and synostosis had the lowest p-value, followed by the combinations of craniosynostosis and Horner-Bernard Syndrome and craniosynostosis and self-injury (Table 5).

Table 5. Ten disease combinations with lowest p-value for craniosynostosis. All disease-disease combinations of which craniosynostosis was a part of were ranked by p-value, which were calculated using a Fisher's Exact Test performed by a Python program. Four of these combinations have not been previously elucidated, thus they may provide new diseases that may be related to craniosynostosis.

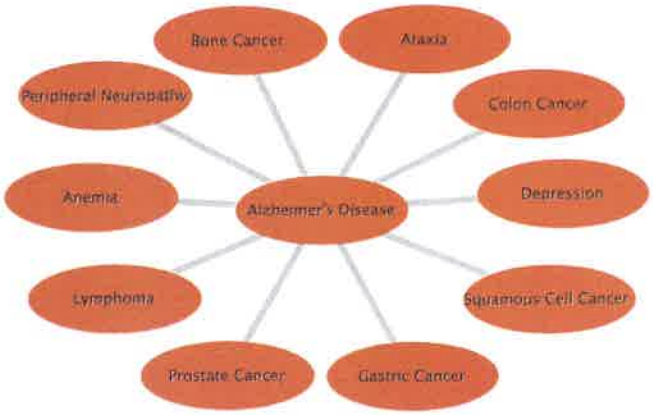
Disease 1	Disease 2	P-value	Shared Genes	Major Shared Genes	Pathological Connection
Craniosynostosis	Synostosis	$3.5842*10^{-203}$	99	FGFR2, FGFR3	Craniofacial disorders
Craniosynostosis	**Horner-Bernard Syndrome**	$6.9970*10^{-110}$	**114**	**FGFR2, FGFR3**	**Involve neurological dysfunction**
Craniosynostosis	**Self-Injury**	$6.6819*10^{-72}$	**87**	**FGFR2, FGFR3**	**Involve neurological dysfunction**
Craniosynostosis	Cardiovascular Disease	$3.8857*10^{-52}$	69	FGFR2, FGFR3	Share diabetes as a risk factor
Craniosynostosis	**Ataxia-Telangiectasia Syndrome**	$7.2962*10^{-48}$	**86**	**TWIST1, FGFR3**	**Involve neurological defects**
Craniosynostosis	**Syndrome X**	$6.5576*10^{-47}$	**42**	**FGFR2, FGFR3**	**Involve metabolic disorders**
Craniosynostosis	Heart Disease	$5.7907*10^{-44}$	70	FGFR2, FGFR3	Share diabetes as a risk factor
Craniosynostosis	Hearing Loss	$7.0746*10^{-41}$	40	FGFR2, FGFR3	Involve inner ear dysfunction
Craniosynostosis	Congenital Heart Defects in Children	$2.1270*10^{-40}$	36	FGFR2, FGFR2	Congenital defects
Craniosynostosis	Mental Illness	$1.0384*10^{-38}$	71	FGFR2, FGFR3	Involve neurological dysfunction

Of the craniosynostosis combinations with the ten lowest p-values, six have been previously elucidated [3], [40]–[42]. However, this leaves four potentially new disease-disease connections involving craniosynostosis. The four diseases are Horner-Bernard Syndrome, Self-Injury, Ataxia-Telangiectasia Syndrome, and Syndrome X. Horner-Bernard Syndrome is a condition that results from damage to the oculosympathetic pathway and presents with droopy eyelids, constricted pupils and a failure to sweat. Ataxia-telangiectasia is a rare autosomal recessive condition that results in a loss of coordination, and an increased susceptibility to cancer. Syndrome X is a syndrome made up of abnormalities that include hypoglycemia and glucose intolerance that lead to hypertension, obesity, and high cholesterol. These four diseases may present themselves later in the lives of craniosynostosis patients, and improved treatments for them can be designed based on their correlations to craniosynostosis.

Relationships Revealed by Disease Network

The disease network clusters diseases, such that similar clusters signify molecular similarities. A disease network for Alzheimer's Disease is shown in Figure 2. The ten diseases closest to Alzheimer's Disease on the genetic distance plot were added to the disease network. Alzheimer's Disease is closest to peripheral neuropathy, bone cancer, ataxia, colon cancer, depression, squamous cell cancer, gastric cancer, prostate cancer, lymphoma, and anemia. New drug targets for these diseases can be elucidated based on these connections.

Figure 2. Disease Network for Alzheimer's Disease. Disease network was created based on genetic distances of each disease-disease correlation. R's multidimensional scaling function used to create network from genetic distance matrix input. Disease network created by finding 10 diseases closest to Alzheimer's Disease from overall disease network plot. Users are able to hover over each line and a list of major shared genes and possible drug therapies will be displayed.



A disease network for craniosynostosis is shown in Figure 3. The ten diseases closest to craniosynostosis on the genetic distance plot were added to the disease network. Craniosynostosis is closest to obsessive compulsive disorder, eating disorders, kidney cysts, myoclonus, premature ovarian failure, hypopituitarism, atrial septal defect, dislocation, dysarthria, and detached retina. New drug targets for these diseases can be elucidated based on these connections.

Figure 3. Disease Network for Craniosynostosis. Disease network was created based on genetic distances of each disease-disease correlation. R's multidimensional scaling function used to create network from genetic distance matrix input. Disease network created by finding 10 diseases closest to Craniosynostosis from overall disease network plot. Users are able to hover over each line and a list of major shared genes and possible drug therapies will be displayed.

Conclusions

The disease-disease database is comprehensive, as it takes all genetic literature for 1648 diseases into account when determining the significance of disease-disease correlations. The new database is an improvement over previous databases due to the use of abstract parsing and the greater number of diseases included. The database can be used to design better pharmacological treatments, and to inform patients about future complications associated with their disease. Despite the new database being an improvement over previous ones, additional improvements will be made. Manually studying disease-disease correlations

allows the mutation percentages of genes to be considered, thus weighing the importance of each gene when establishing correlations. Many abstracts, however, do not contain the mutation percentages of genes, so their importance can't be accounted for when collecting gene lists. Future publications should include mutation percentages of genes in the abstracts, so future iterations of the program would be able to consider these when establishing disease-disease correlations. Another improvement to the database is to add a function such that the database is automatically up to date, and that data from new literature will not have to be manually loaded into the database. Expanding the database outside of the defined list from the Mayo Clinic will also be beneficial, so that even if a disease outside of the original list is searched, the database can still find disease-disease correlations for that disease. Using the disease-disease database, a disease network can be built to better visualize the relationships between diseases [43]. In the future, further development of disease-disease databases such as the one created in this study, will lead to physicians broadening their perspectives when studying diseases, to look at connections between multiple diseases, which will help develop treatments for life-threatening diseases that currently plague society

Appendix 1

Genetic Terms	
Mutation	Mutations
Mutant	Mutants
Variant	Variants
Missense	Nonsense
Deletion	Deletions
Insertion	Insertions
Genotype	Genotypes
Polymorphism	Polymorphisms
SNP	SNPs
Exon	Exons
GWAS	WES
Frameshift	Frameshifts

Appendix 1 – Genetic Terms List. Genes would only be written to the output file if one of these terms was within three words of the gene name to ensure that the gene was actually causative of the disease or mutated in it and not just important in a disease-related pathway.

References

[1] X. Wu and Y. Gu, "Signaling Mechanisms Underlying Genetic Pathophysiology of Craniosynostosis," *Int. J. Biol. Sci.*, vol. 15, no. 2, pp. 298–311, 2019, doi: 10.7150/ijbs.29183.

[2] S. R. F. Twigg and A. O. M. Wilkie, "A Genetic-Pathophysiological Framework for Craniosynostosis," *Am. J. Hum. Genet.*, vol. 97, no. 3, pp. 359–377, Sep. 2015, doi: 10.1016/j.ajhg.2015.07.006.

[3] K.-W. Shim, E.-K. Park, J.-S. Kim, Y.-O. Kim, and D.-S. Kim, "Neurodevelopmental Problems in Non-Syndromic Craniosynostosis," *J. Korean Neurosurg. Soc.*, vol. 59, no. 3, pp. 242–246, May 2016, doi: 10.3340/jkns.2016.59.3.242.

[4] M. Poot, "Structural Genome Variations Related to Craniosynostosis," *Mol. Syndromol.*, vol. 10, no. 1–2, pp. 24–39, 2019, doi: 10.1159/000490480.

[5] K. A. Miller *et al.*, "Diagnostic value of exome and whole genome sequencing in craniosynostosis," *J. Med. Genet.*, vol. 54, no. 4, pp. 260–268, Apr. 2017, doi: 10.1136/jmedgenet-2016-104215.

[6] E. L. Durham, R. N. Howie, and J. J. Cray, "Gene/environment interactions in craniosynostosis: A brief review," *Orthod. Craniofac. Res.*, vol. 20 Suppl 1, pp. 8–11, Jun. 2017, doi: 10.1111/ocr.12153.

[7] W. Lattanzi, M. Barba, L. Di Pietro, and S. A. Boyadjiev, "Genetic advances in craniosynostosis," *Am. J. Med. Genet. A.*, vol. 173, no. 5, pp. 1406–1429, May 2017, doi: 10.1002/ajmg.a.38159.

[8] Z. Steinhart and S. Angers, "Wnt signaling in development and tissue homeostasis," *Development*, vol. 145, no. 11, p. dev146589, Jun. 2018, doi: 10.1242/dev.146589.

[9] R. N. Wang *et al.*, "Bone Morphogenetic Protein (BMP) signaling in development and human diseases," *Genes Dis.*, vol. 1, no. 1, pp. 87–105, Sep. 2014, doi: 10.1016/j.gendis.2014.07.005.

[10] E. Cabrera Je, "The Role of Transcription Factor TWIST in Cancer Cells," *J. Genet. Syndr. Gene Ther.*, vol. 04, no. 01, 2013, doi: 10.4172/2157-7412.1000124.

[11] T. Helsten, S. Elkin, E. Arthur, B. N. Tomson, J. Carter, and R. Kurzrock, "The FGFR Landscape in Cancer: Analysis of 4,853 Tumors by Next-Generation Sequencing," *Clin. Cancer Res.*, vol. 22, no. 1, pp. 259–267, Jan. 2016, doi: 10.1158/1078-0432.CCR-14-3212.

[12] M. J. Pfaff, K. Xue, L. Li, M. C. Horowitz, D. M. Steinbacher, and J. V. P. Eswarakumar, "FGFR2c-mediated ERK-MAPK activity regulates coronal suture development," *Dev. Biol.*, vol. 415, no. 2, pp. 242–250, 15 2016, doi: 10.1016/j.ydbio.2016.03.026.

[13] C. Wellbrock and I. Arozarena, "The Complexity of the ERK/MAP-Kinase Pathway and the Treatment of Melanoma Skin Cancer," *Front. Cell Dev. Biol.*, vol. 4, p. 33, 2016, doi: 10.3389/fcell.2016.00033.

[14] J. Roche, "The Epithelial-to-Mesenchymal Transition in Cancer," *Cancers*, vol. 10, no. 2, Feb. 2018, doi: 10.3390/cancers10020052.

[15] Z. Zhao, M. A. Rahman, Z. G. Chen, and D. M. Shin, "Multiple biological functions of Twist1 in various cancers," *Oncotarget*, vol. 8, no. 12, pp. 20380–20393, Mar. 2017, doi: 10.18632/oncotarget.14608.

[16] C.-C. Liu *et al.*, "DiseaseConnect: a comprehensive web server for mechanism-based disease-disease connections," *Nucleic Acids Res.*, vol. 42, no. Web Server issue, pp. W137-146, Jul. 2014, doi: 10.1093/nar/gku412.

[17] J. Yang *et al.*, "DNetDB: The human disease network database based on dysfunctional regulation mechanism," *BMC Syst. Biol.*, vol. 10, no. 1, p. 36, May 2016, doi: 10.1186/s12918-016-0280-5.

[18] A. Guilhem *et al.*, "Intra-venous bevacizumab in hereditary hemorrhagic telangiectasia (HHT): A retrospective study of 46 patients," *PloS One*, vol. 12, no. 11, p. e0188943, 2017, doi: 10.1371/journal.pone.0188943.

[19] S. Lauro, C. E. Onesti, R. Righini, and P. Marchetti, "The use of bevacizumab in non-small cell lung cancer: an update," *Anticancer Res.*, vol. 34, no. 4, pp. 1537–1545, Apr. 2014.

[20] V. A. Venkatesha *et al.*, "P7170, a novel inhibitor of mTORC1/mTORC2 and Activin receptor-like Kinase 1 (ALK1) inhibits the growth of non small cell lung cancer," *Mol. Cancer*, vol. 13, p. 259, Dec. 2014, doi: 10.1186/1476-4598-13-259.

[21] Y. Zeng *et al.*, "MicroRNA-205 targets SMAD4 in non-small cell lung cancer and promotes lung cancer cell growth in vitro and in vivo," *Oncotarget*, vol. 8, no. 19, pp. 30817–30829, May 2017, doi: 10.18632/oncotarget.10339.

[22] N. Harmston, W. Filsell, and M. P. H. Stumpf, "What the papers say: text mining for genomics and systems biology," *Hum. Genomics*, vol. 5, no. 1, pp. 17–29, Oct. 2010.

[23] D. Westergaard, H.-H. Stærfeldt, C. Tønsberg, L. J. Jensen, and S. Brunak, "A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts," *PLoS Comput. Biol.*, vol. 14, no. 2, p. e1005962, 2018, doi: 10.1371/journal.pcbi.1005962.

[24] N. Baumann, "How to use the medical subject headings (MeSH)," *Int. J. Clin. Pract.*, vol. 70, no. 2, pp. 171–174, Feb. 2016, doi: 10.1111/ijcp.12767.

[25] B. Braschi *et al.*, "Genenames.org: the HGNC and VGNC resources in 2019," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D786–D792, Jan. 2019, doi: 10.1093/nar/gky930.

[26] M. Stonebraker, L. A. Rowe, and M. Hirohama, "The implementation of POSTGRES," *IEEE Trans. Knowl. Data Eng.*, vol. 2, no. 1, pp. 125–142, Mar. 1990, doi: 10.1109/69.50912.

[27] J. Uttariello, "Build a CRUD Template Using React, Bootstrap, Express & Postgres." Medium, 25-Mar-2019.

[28] C. T. Karlsson, B. Malmer, F. Wiklund, and H. Grönberg, "Breast cancer as a second primary in patients with prostate cancer--estrogen treatment or association with family history of cancer?," *J. Urol.*, vol. 176, no. 2, pp. 538–543, Aug. 2006, doi: 10.1016/j.juro.2006.03.036.

[29] W.-Y. Sun *et al.*, "Synchronous bilateral male breast cancer: a case report," *J. Breast Cancer*, vol. 15, no. 2, pp. 248–251, Jun. 2012, doi: 10.4048/jbc.2012.15.2.248.

[30] A. Ari, C. Tatar, K. Buyukasik, O. Segmen, C. Cakir, and S. Arikan, "Synchronous male breast and colon cancer presenting with ileus: A case report," *Int. J. Surg. Case Rep.*, vol. 28, pp. 31–33, 2016, doi: 10.1016/j.ijscr.2016.09.023.

[31] A. J. Karayiannakis *et al.*, "Synchronous breast and rectal cancers in a man," *Case Rep. Oncol.*, vol. 4, no. 2, pp. 281–286, May 2011, doi: 10.1159/000328995.

[32] S. Castaneda, C. Gonzalez-Juanatey, and M. A. Gonzalez-Gay, "Inflammatory Arthritis and Heart Disease," *Curr. Pharm. Des.*, vol. 24, no. 3, pp. 262–280, 2018, doi: 10.2174/1381612824666180123102632.

[33] V. Subramaniam and G. Y. H. Lip, "Hypertension to heart failure: a pathophysiological spectrum relating blood pressure, drug treatments and stroke," *Expert Rev. Cardiovasc. Ther.*, vol. 7, no. 6, pp. 703–713, Jun. 2009, doi: 10.1586/erc.09.43.

[34] M. Thomsen and B. G. Nordestgaard, "Myocardial infarction and ischemic heart disease in overweight and obesity with and without metabolic syndrome," *JAMA Intern. Med.*, vol. 174, no. 1, pp. 15–22, Jan. 2014, doi: 10.1001/jamainternmed.2013.10522.

[35] A. P. de Jonghe-Rouleau, A. M. Pot, and J. F. M. de Jonghe, "Self-injurious behaviour in nursing home residents with dementia," *Int. J. Geriatr. Psychiatry*, vol. 20, no. 7, pp. 651–657, Jul. 2005, doi: 10.1002/gps.1337.

[36] R. Gonzalez-Casas, E. A. Jones, and R. Moreno-Otero, "Spectrum of anemia associated with chronic liver disease," *World J. Gastroenterol.*, vol. 15, no. 37, pp. 4653–4658, Oct. 2009, doi: 10.3748/wjg.15.4653.

[37] R. Paternostro *et al.*, "Erectile dysfunction in cirrhosis is impacted by liver dysfunction, portal hypertension, diabetes and arterial hypertension," *Liver Int. Off. J. Int. Assoc. Study Liver*, vol. 38, no. 8, pp. 1427–1436, 2018, doi: 10.1111/liv.13704.

[38] N. Anne and R. Pallapothu, "Lymphoma of the breast: a mimic of inflammatory breast cancer," *World J. Surg. Oncol.*, vol. 9, p. 125, Oct. 2011, doi: 10.1186/1477-7819-9-125.

[39] K. V. Vinod, V. Sivabal, and M. V. Vidya, "Iatrogenic Horner's syndrome: A cause for diagnostic confusion in the emergency department," *World J. Emerg. Med.*, vol. 8, no. 3, pp. 235–236, 2017, doi: 10.5847/wjem.j.1920-8642.2017.03.014.

[40] D. Johnson and A. O. M. Wilkie, "Craniosynostosis," *Eur. J. Hum. Genet. EJHG*, vol. 19, no. 4, pp. 369–376, Apr. 2011, doi: 10.1038/ejhg.2010.235.

[41] I. Panigrahi, "Craniosynostosis genetics: The mystery unfolds," *Indian J. Hum. Genet.*, vol. 17, no. 2, pp. 48–53, May 2011, doi: 10.4103/0971-6866.86171.

[42] N. B. Agochukwu, B. D. Solomon, and M. Muenke, "Hearing loss in syndromic craniosynostoses: introduction and consideration of mechanisms," *Am. J. Audiol.*, vol. 23, no. 2, pp. 135–141, Jun. 2014, doi: 10.1044/2014_AJA-13-0036.

[43] E. Oerton, I. Roberts, P. S. H. Lewis, T. Guilliams, and A. Bender, "Understanding and predicting disease relationships through similarity fusion," *Bioinforma. Oxf. Engl.*, vol. 35, no. 7, pp. 1213–1220, Apr. 2019, doi: 10.1093/bioinformatics/bty754.