Investigation on Genetic Mutations Leading to Monogenic and Polygenic Heart Diseases Using Publicly Available Disease Databases

Sanika Javeri

Indiana University, USA

ABSTRACT

An investigation was completed use publicly available NCBI and DISGNET databases to determine the genetic association to predominant heart diseases. Preliminary research was conducted to detect the two most common types of heart disease, hypertrophic cardiomyopathy (HCM) and coronary artery disease (CAD), and the respective monogenic and polygenic genes responsible for causing these diseases. The genetic etiology of the mutating genes was investigated with hopes that the phenotypic results could be established on publicly available databases. This would allow countless changes in medicine regarding heart disease such as accurate genetic counseling, treatment guidance, risk assessment, and preventative measures, such as living heart friendly lifestyle, can be taken to positively influence the disease course. This course of action would lead to early diagnosis in affected individuals and possibly preventing sudden cardiac death.

Introduction

Heart Disease, which can refer to several distinctive forms of heart conditions, is strikingly more dangerous than many presume. Inducing several deaths worldwide and being the number one killer in the U.S., heart disease is also a major cause of disability. According to the Centers for Disease Control and Prevention (CDC), heart disease has become the leading cause of death in the United States, approximately causing 1 in 4 deaths (Heron and Anderson 2016; CDC report). Although hereditary gene inheritance can be classified as a prominent reason for some of the severe heart diseases, in most scenarios, the root cause of the diseases relates to prior dietary habits, sedentary lifestyles, and other environmental factors. (Nystoriak and Bhatnagar 2018). Not only do more than 659,000 individuals die from the disease yearly, but the United States also suffers a great deal, due to loss of productivity, health care services, medicine, and monetary value. On average, the nation loses $363 million on heart disease and price is constantly increasing (Heidenreich et al. 2022). However, in many cases, research showcases that the key roles in the etiology of the heart diseases can furthermore be prevented through analyzing gene expression.

The two genetic forms of heart disease that exist are monogenic and polygenic. Monogenic is a result of a modification in one gene that obtains a strong influence in causing the disease, while polygenic is the result of modifications in more than one gene (Trinder et al. 2020). The more common the genetics variants in the sequence, the higher chance an individual has of being at risk for the disease. In the United States, the two most common type of heart diseases are hypertrophic cardiomyopathy (HCM) and coronary artery disease (CAD). Hypertrophic cardiomyopathy, a monogenic disease, has been cited as the most prevailing reason for sudden cardiac death in young people and athletes; both under the age of 35. In hypertrophic cardiomyopathy the heart muscle becomes perversely thick, making it difficult and demanding for the organ to continuously pump blood. This then can lead to blood clots, stroke, or heart failure. HCM is ordinarily caused by abnormal genes in the heart muscle. These atypical genes cause the walls of the heart chamber, specifically in the left ventricle, to contract vigorously and become thicker (Mayo Clinic 2020). On the other hand, coronary artery disease, a polygenic disease, affects approximately 7% of the population and is the leading

form of heart disease in the nation. The commonly acknowledged cause for CAD is atherosclerosis. Atherosclerosis is the buildup of plaque and other fatty material in the inner walls of the coronary arteries, causing them to narrow overtime. This dangerous process prevents consistent blood flow throughout the body, will eventually narrows the arteries, and leads to possible chest pain, heart attack, heart failure, or even abnormal heart rhythm, which is called arrhythmia (American Heart Association 2020). Both coronary artery disease and hypertrophic cardiomyopathy are serious cases, that when not given proper attention to or carefully observed, can result in death.

When diagnosing HCM and CAD, their risk factors can normally be classified into two categories: Hereditary and Non-hereditary. Non-hereditary traditional risk factors that are found to be responsible for HCM and CAD are obesity, hypertension, lack of exercise, diabetes mellitus, tobacco use, and hypercholesterolemia (Powell-Wiley et al. 2021; Roberts et al. 2022). However, recent research studies such as the "Hypertrophic Cardiomyopathy Overview of Genetics and Management" and "Genetic Risk in Coronary Artery Disease", show that there is a strong genetic predisposition on the causes of CAD and HCM (Teekakirikul et al. 2019; Khera et al. 2016). Controversy rises regarding diagnosis since it is believed that heart disease cannot be prevented, and even today in most cases of genetics, the patient is given insufficient choices; however, recent research has indicated that the identification can be determined of the mutated genes involved in the gene expression that evokes heart disease.

The investigation conducted studied the genetic etiology of hypertrophic cardiomyopathy (HCM) and coronary artery disease (CAD). It was hypothesized that a correlation would be found between the type of gene being mutated and the location of the variation on the chromosome that causes HCM and CAD respectively. A methodology was constructed to identify the genes involved in the monogenetic and polygenic diseases. The databases NCBI and DisGeNET were selected so that the clinical databases of genes and their mutations could be accessed. The purpose of the project was to identify the specific gene(s) that follow a pattern of randomly mutating and analyze the nature of the mutations so that methods of identification that are easily accessible to the public, such as the curated databases, UniProt, the GWAS Catalog, or ClinVar prevention could be include information regarding the variant-likely genes.

Methods

NCBI and DisGeNET Data Mining

Initially, primary research was conducted to determine the most prevalent genes present in both monogenic and polygenic diseases. The molecular consequences, variation types, and variant length of the genes were examined. Utilizing the United States National Library of Medicine and the ClinVar database, it was established that the most common gene that causes hypertrophic cardiomyopathy is MYH7 and most common gene that causes coronary artery disease is LPA, and these genes were applied as the model genes for the remainder of the project. These genes were then filtered through the NCBI database to determine the etiology of the disease and the location on the chromosome. NCBI, National Center for Biotechnology Information, is a branch of the National Institutes of Health that is reviewed by credible authors, approved by the National Library of Medicine, and funded by the U.S. government. Filters such as benign, single nucleotide, and variant length less than 51 bp, were used to narrow the results and find information regarding relevant clinical testing of single gene variants. After the genes were located on their respective chromosome, the variant-gene disease association was explored in more detail using DisGeNET. DisGeNET is another publicly available database that accommodates one of the largest collections of information regarding the association of genes and variants to diseases, also allowing the investigation of the molecular mechanisms underlying diseases of genetic origin. Using DisGeNET, the two types of heart disease, hypertrophic cardiomyopathy and coronary artery disease were explored. The results were narrowed down with filters such as the type of gene; MYH7 and LPA. This search was principal as it led to the understanding of the likelihood of the genes mutating as well as the chance MYH7 and LPA causes other diseases.

Results and Discussions

Using the NCBI database, the research administered indicated that there are two forms of heart disease; monogenic and polygenic. The prevalent heart diseases in the United States for monogenic and polygenic heart disease are hypertrophic cardiomyopathy (HCM) and coronary artery disease (CAD). With further use of the database, identification of the variating genes were located on the exact site on the chromosomes. The research indicated that there were at least 67 different sites on the DNA sequence that could increase the chances of an individual having a heart attack. Specific mutating genes were also identified for the monogenic and polygenic diseases. For hypertrophic cardiomyopathy, the MYH7 gene was stated to be responsible as it mutates and causes contractile velocity of the cardiac muscle myosin in the ventricles. The causative gene and the cytogenetic location for the mutation was determined using the NCBI database. The location was on 14q11.2, which is the long (q) arm of chromosome 14 at position 11.2 The Molecular Location is in between the base pairs 23,412,740 to 23,435,677 on chromosome 14 (Homo sapiens Updated Annotation Release 109.20200522, GRCh38.p13) (NCBI). To prevent any possible limitations that may occur, additional research was completed on the DisGeNET database to study the Variant Gene Disease Association. Reports found on DisGeNET found more than 400K gene-disease associations (GDAs), that linked roughly 18000 genes to 15000 diseases. The majority of the GDAs found through DisGeNET were supported by a research or clinical publication, allowing credibility for the selection of genes that were initially identified using the NCBI database. DisGeNET contained 72 870 variant-disease associations (VDAs), between 46 589 SNPs and 6356 diseases and phenotypes (Piñero, etal. 2020; Piñero,et al. 2017; Piñero, et al. 2015) . Next, Hypertrophic Cardiomyopathy (C0007194) was studied for variant disease association (VDA). From the 103 genes originally determined, there were 635 results for VDA. With this finding, the variant association between the MYH7 gene and the occurrence of HCM were observed and are reported in Table 1 and Table 2, where the top ten genes listed are showcased in Table 1. The variants of the MYH7 genes in DisGeNET were obtained from search in curated databases such as UniProt, ClinVar, the GWAS Catalog, and the GWAS. The variants, responsible for Hypertrophic Cardiomyopathy, were first annotated with the position in the chromosome 14 and were selected based on other criteria, such as the number of disease-causing variants (more than 5), disease specificity index (DSI; above 0.7) and the variable Disease Pleiotropy Index (DPI). It was found that the variants were mainly attributed to missense variation and approximately 80% of the genes only one variant causing HCM, and this information is shown in Table 2 and Table 3. Similar research was conducted for the variants of the LPA genes involved in coronary artery disease (CAD; C1956346) and they were also researched in DisGeNET Curated databases from UniProt, ClinVar, the GWAS Catalog, and GWAS. Once again, the variants, responsible for coronary artery disease, were annotated with the position in the chromosome 6 and criteria such as the number of disease-causing variants (more than 1), disease specificity index (DSI; above 0.6) and variable Disease Pleiotropy Index (DPI) was selected. From the 1380 genes, there were 1577 hits, and Table 1 showcases the top ten genes that appeared. The variants were mainly attributed to the missense variants, splice donor, and intron variants, with approximately 83% of the genes having ony one variant that was causing CAD. The data was then sorted for analysis of the variants of the major genes. In this work, variants associated with LPA genes that are associated with HCM occurrences are reported in Table 2 and Table 4.

Table 1. Top 10 genes and gene variants involved in Hypertrophic Cardiomyopathy and Coronary artery disease.

	Hypertrophic Cardiomyopathy		**Coronary Artery Disease**	
	Gene	Count of Genes variants	Gene	Count of Gene variants
1	MYBPC3	246	CDKN2B-AS1	41
2	**MYH7**	123	**LPA**	15
3	TNNT2	30	NOS3	14

4	TNNI3	23	LDLR	12
5	TPM1	23	LPL	12
6	MYL2	11	ABCA1	12
7	PRKAG2	10	SCARB1	9
8	LAMP2	9	MRAS	9
9	MHRT	8	CETP	9
10	TNNC1	7	PLPP3	8

Table 2. The count of consequence for polygenic CAD and Monogenic HCM differed for LPA and MYH7 genes. The polygenic LPA had more intron variants and less frameshift mutations**.**

	Count of Consequence for HCM	**Count of Consequence for CAD**
Intron Variant	10	687
Missense Variant	369	369
Upstream Gene Variant	0	104
3 Prime UTR Variant	1	91
Intergenic Variant	0	89
Synonymous Variant	2	53
Noncoding Transcript Exon Variant	2	51
Downstream Gene Variant	0	46
5 Prime UTR Variant	0	29
Regulatory Region Variant	0	24
Stop Gained	68	13
TF Binding Site Variant	0	5
Mature miRNA Variant	0	4
Frameshift Variant	110	3
Splice Region Variant	8	3
Splice Donor Variant	29	2
Start Lost	1	2
Stop Lost	0	1
Inframe Insertion	1	1
Splice Acceptor Variant	22	0
Inframe Deletion	12	0
Total	**635**	**1577**

Table 3. Variants of MYH7 genes in DisGeNET were obtained from searches in curated databases that mainly attributed to the missense variation.

Variant	N (diseases)	DSI	DPI	Position	Alleles
rs121913630	7	0.851	0.08	23425814	G/A;C
rs121913627	8	0.851	0.08	23427657	C/A;G;T
rs36211715	5	0.851	0.08	23424839	C/A;G;T
rs1566535410	5	0.851	0.08	23429297	T/C
rs606231324	5	0.851	0.08	23428505	C/G;T
rs267606910	6	0.807	0.08	23431589	C/T
rs863224900	6	0.807	0.16	23428534	A/C;G;T
rs3218714	9	0.763	0.16	23429279	G/A;C
rs121913628	10	0.763	0.16	23424059	C/G;T
rs267606908	9	0.763	0.16	23424112	T/C
rs3218713	10	0.763	0.16	23431468	C/A;T
rs371898076	9	0.763	0.16	23426833	C/T
rs397516127	9	0.763	0.16	23426834	G/A;C
rs397516171	9	0.763	0.16	23424041	C/G;T
rs397516264	9	0.763	0.16	23431602	C/T
rs3218716	17	0.716	0.28	23425316	C/A;G;T

Table 4. Variants of the LPA genes in DisGeNET were obtained from searches in the curated databases with multiple gene consequences.

Variant	N(diseases)	DSI	DPI	Position	Consequence	Alleles
rs10455872	33	0.662	0.32	160589086	intron variant	A/G
rs10755578	2	0.925	0.04	160548706	intron variant	C/G
rs3127599	2	0.925	0.04	160486102	intron variant	C/T
rs41272114	2	1	0.04	160585045	splice donor variant	C/A;T
rs74617384	5	0.925	0.08	160576086	intron variant	A/G;T
rs7767084	2	0.925	0.04	160541471	intron variant	T/C
rs12214416	1	1	0.04	160489485	intron variant	T/A
rs140570886	3	1	0.04	160591981	intron variant	T/C
rs147555597	2	1	0.04	160490564	intron variant	G/A
rs55730499	2	1	0.04	160584578	intron variant	C/T
rs6926458	2	1	0.04	160598834	intron variant	A/G;T
rs73596816	2	1	0.04	160596331	intron variant	G/A
rs7770628	2	1	0.04	160597142	intron variant	C/T
rs3798220	16	0.732	0.16	160540105	missense variant	T/C
rs1801693	1	1	0.04	160548597	missense variant	A/C;G
rs6415084	2	1	0.04	160559298	intron variant	T/C
rs7765781	1	1	0.04	160586464	missense variant	G/C

rs9364559	1	1	0.04	160555116	intron variant	A/G

Conclusions

In this project, the genes responsible for monogenic HCM and Polygenic CAD were researched using the publicly available database NCBI and then DisGeNET was accessed to mine for the top variants in the genes. Although, the information on mutants and mutations is key for clinical intervention, in today's world where cheap sequencing data is available, it has become simple to validate the mutations for their clinical relevance. For example, when the top mutant of the MYH7 gene (rs121913630) was checked for clinical relevance and evidence for any variant disease associations using the PMID 2010 year reports, it was found that the mutations were clinically relevant in multiple studies (García-Giustiniani et al. 2015; Olivotto et al. 2011). This study is relevant as it allows the genetic variation, cytogenic location, and information on phenotypic expression to be easily accessed and understood. This will permit researching the correct dietary and pharmaceutical molecules that could cure heart abnormalities. In the future, I would like to use this research to conduct genetic analysis for individuals who display the common phenotypic expression of mutations in MYH7 and LPA phenotype. This would be indicated by the thickening of the cardiac muscle, irregular heart rhythm, arrhythmia, shortness of breath, or dyspnea.

References

Centers for Disease Control and Prevention. About heart disease. Atlanta, Georgia. 2021. Available from: https://www.cdc.gov/heartdisease/about.htm.

Dai, Xuming, et al. "Genetics of Coronary Artery Disease and Myocardial Infarction." World Journal of Cardiology,Baishideng Publishing Group Inc, 26 Jan. 2016,

García-Giustiniani D, Arad M, Ortíz-Genga M, Barriales-Villa R, Fernández X, Rodríguez-García I, Mazzanti A, Veira E, Maneiro E, Rebolo P, Lesende I, Cazón L, Freimark D, Gimeno-Blanes JR, Seidman C, Seidman J, McKenna W, Monserrat L. Phenotype and prognostic correlations of the converter region mutations affecting the β myosin heavy chain. Heart. 2015 Jul;101(13):1047-53. doi: 10.1136/heartjnl-2014-307205. Epub 2015 May 2. PMID: 25935763; PMCID: PMC4484257.

Heron M, Anderson RN. Changes in the leading cause of death: Recent patterns in heart disease and cancer mortality. NCHS Data Brief, no 254. Hyattsville, MD: National Center for Health Statistics. 2016.

Heidenreich PA, Fonarow GC, Opsha Y, Sandhu AT, Sweitzer NK, Warraich HJ; HFSA Scientific Statement Committee Members Chair. Economic Issues in Heart Failure in the United States. J Card Fail. 2022 Mar;28(3):453-466. doi: 10.1016/j.cardfail.2021.12.017. Epub 2022 Jan 24. PMID: 35085762; PMCID: PMC9031347.

Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. Nat Rev Genet. 2017 Jun;18(6):331-344. doi: 10.1038/nrg.2016.160. Epub 2017 Mar 13. PMID: 28286336; PMCID: PMC5935119.

Murphy SL, Xu J, Kochanek KD, Arias E. Mortality in the United States, 2017 . NCHS data brief, no 328. Hyattsville, MD:National Center for Health Statistics; 2018.

Nystoriak MA, Bhatnagar A. Cardiovascular Effects and Benefits of Exercise. Front Cardiovasc Med. 2018 Sep 28;5:135. doi: 10.3389/fcvm.2018.00135. PMID: 30324108; PMCID: PMC6172294.

Olivotto I, Girolami F, Sciagrà R, Ackerman MJ, Sotgia B, Bos JM, Nistri S, Sgalambro A, Grifoni C, Torricelli F, Camici PG, Cecchi F. Microvascular function is selectively impaired in patients with hypertrophic cardiomyopathy and sarcomere myofilament gene mutations. J Am Coll Cardiol. 2011 Aug 16;58(8):839-48. doi: 10.1016/j.jacc.2011.05.018. PMID: 21835320.

Paula F. Martinez and Marina P. Okoshi Genetic Risk in Coronary Artery Disease. Arq Bras Cardiol. 2018 Jul; 111(1): 62–63.

Piñero, J.; Bravo, À.; Queralt-Rosinach, N.; Gutiérrez-Sacristán, A.; Deu-Pons, J.; Centeno, E.; García-García, J.; Sanz, F.; Furlong, L.I. DisGeNET: A Comprehensive Platform Integrating Information on Human Disease-Associated Genes and Variants. Nucleic Acids Res. 2017, 45, D833–D839.

Piñero, J.; Queralt-Rosinach, N.; Bravo, À.; Deu-Pons, J.; Bauer-Mehren, A.; Baron, M.; Sanz, F.; Furlong, L.I. DisGeNET: A Discovery Platform for the Dynamical Exploration of Human Diseases and Their Genes. Database 2015, 2015, bav028

Piñero, J.; Ramírez-Anguita, J.M.; Saüch-Pitarch, J.; Ronzano, F.; Centeno, E.; Sanz, F.; Furlong, L.I. The DisGeNET Knowledge Platform for Disease Genomics: 2019 Update. Nucleic Acids Res. 2020, 48, D845–D855.

Polakit Teekakirikul, Wenjuan Zhu, Helen C. Huang, and Erik Fung Hypertrophic Cardiomyopathy: An Overview of Genetics and Management Biomolecules. 2019 Dec; 9(12): 878.

Powell-Wiley TM, Poirier P, Burke LE, Després JP, Gordon-Larsen P, Lavie CJ, Lear SA, Ndumele CE, Neeland IJ, Sanders P, St-Onge MP; American Heart Association Council on Lifestyle and Cardiometabolic Health; Council on Cardiovascular and Stroke Nursing; Council on Clinical Cardiology; Council on Epidemiology and Prevention; and Stroke Council. Obesity and Cardiovascular Disease: A Scientific Statement From the American Heart Association. Circulation. 2021 May 25;143(21):e984-e1010. doi: 10.1161/CIR.0000000000000973. Epub 2021 Apr 22. PMID: 33882682; PMCID: PMC8493650.

Roberts R, Chavira J, Venner E. Genetic risk and its role in primary prevention of CAD. Journal of Translational Genetics and Genomics. 2022; 6(4): 388-402. http://dx.doi.org/10.20517/jtgg.2022.07

Teekakirikul P, Zhu W, Huang HC, Fung E. Hypertrophic Cardiomyopathy: An Overview of Genetics and Management. *Biomolecules*. 2019; 9(12):878. https://doi.org/10.3390/biom9120878

Trinder M, Francis GA, Brunham LR. Association of Monogenic vs Polygenic Hypercholesterolemia with Risk of Atherosclerotic Cardiovascular Disease. JAMA Cardiol. 2020 Apr 1;5(4):390-399. doi: 10.1001/jamacardio.2019.5954. Erratum in: JAMA Cardiol. 2020 Mar 11;: PMID: 32049305; PMCID: PMC7042820.