MethCard: A Cardiovascular Diseases Gene Methylation Database for Human

A Major project submitted in partial fulfilment of requirement for the degree of

MASTER OF TECHNOLOGY

IN

BIOINFORMATICS

Submitted by

AYUSHI GARG

(2K15/BIO/03)

Under the supervision of

Prof. Bansi D. Malhotra
Department of Biotechnology
Delhi Technological University
Delhi-110042, INDIA.

CERTIFICATE

This is to certify that the M.Tech dissertation entitled **"MethCard: A Cardiovascular Diseases Gene Methylation Database for Human",** submitted by **AYUSHI GARG (2K15/BIO/03)** in partial fulfilment of the requirements for the award of **Master of Technology in BIOINFORMATICS** and submitted to the Department of Biotechnology of Delhi Technological University (Formerly Delhi College of Engineering) is an authentic record of work carried out under the supervision of **Prof. Bansi D. Malhotra, Department of Biotechnology.**

The information and data presented in this dissertation has not been submitted for the award of any other degree elsewhere.

.

AYUSHI GARG
(Roll No. 2K15/BIO/03)

Prof. Bansi D. Malhotra
(Project Mentor)
Department of Biotechnology
Delhi Technological University
Delhi-110042.

ACKNOWLEDGEMENT

It is a fact that every mission needs a spirit of hard work & dedication but it also needs to be put on right path. The credit goes to a number of well-wishers who have helped me to complete my dissertation work successfully with profound appreciation.

I would like to acknowledge my deep sense of gratitude to **Prof. D. Kumar,** Head of Department of Biotechnology, Delhi Technological University for giving me opportunity to study and work in this prestigious institute.

I owe my inexhaustible gratitude to **Prof. Bansi D. Malhotra,** my mentor, Delhi Technological University who gave me a chance of doing this project work under his supervision. I am also thankful to all the faculty members and staff for their encouragement which helped me achieve success in my work.

I owe my sincere thanks to all the members of **Nanobioelectronics Laboratory** who guided me throughout the course of the project work and helped me complete it in a timely manner.

AYUSHI GARG
(Roll No. 2K15/BIO/03)

CONTENTS

ABBREVIATIONS USED

1)	CVD	Cardiovascular Diseases
2)	CAD	Coronary Artery Disease
3)	CTD	Comparative Toxicogenomics Database
4)	DDMGD	Dragon Database for Methylated Genes and Diseases
5)	HGNC	HUGO Gene Nomenclature Committee
6)	SNP	Single Nucleotide Polymorphism
7)	DNMT	DNA methyl-transferases
8)	MeSH	Medical Subject Headings
9)	HTML	Hypertext Markup Language
10)	CSS	Cascading Style Sheets

ABSTRACT

DNA methylation can be elucidated as the reversible addition of methyl (CH_3) group to C5 position of cytosine to form 5-methylcytosine that symbolizes an important regulatory layer associated with gene expression and regulation. It needs to be precisely regulated as any alteration in DNA methylation pattern may lead to various diseases. Recent studies have provided evidence of relation between DNA methylation and cardiovascular disease. Thus to facilitate improved diagnosis and better treatment decisions, the objective of this study is to create an efficient, effective and accurate systems to enable the extraction of the detailed information related to methylation of genes involved in cardiovascular disease development by contributing resources from various databases namely Coronary Artery Disease Gene Database (CADgene), DisGeNET, Comparative Toxicogenomics Database (CTD), DiseaseMeth, Dragon Database for Methylated Genes and Diseases (DDMGD), Database of CpG islands and Analytical Tool (DBCAT) etc**.** and various published literatures.

Keywords: DNA methylation, DiseaseMeth, DDMGD, CADgene, Epigenetics

INTRODUCTION

Cardiovascular Diseases (CVDs) are a set of diseases in which heart, blood vessels or both are affected either functionally or structurally. They have become a leading cause of deaths worldwide. Around 17.3 million people die from them every year which constitutes around 30 percent of global deaths [1][2]. In India also 272 deaths per 100,000 people are caused due to them. Genetic, environmental, behavioural, and clinical factors and their interactions are the major reasons behind such high propensity of CVD development. These reasons need to be studied thoroughly to address the significantly increasing burden of CVDs [3][4].

It has been estimated that about 90% of CVDs can be prevented but the mechanism under their development still remains to be a question [5]. In particular, there are many gaps in understanding how the interaction of genetic and environmental factors contribute to CVD development. Recently, epigenetic mechanisms have emerged as a link between internal genetic landscape and external environmental influences, thus may provide a better insight into the etiology of CVD [6]. Although DNA methylation, histone modification, and post-transcriptional silencing mediated by micro-RNAs can lead to epigenetic changes but nowadays most of the researchers are studying the impact change in DNA methylation as it is most feasible to measure in an epidemiological setting [7][8].

DNA methylation is a mechanism by which methyl (CH_3) group is added to DNA. It plays a critical role in gene expression and cell differentiation. The CpG dinucleotide are normally about 60%–90% methylated. Generally, hypo-methylation leads to gene activation whereas hyper-methylation causes gene repression but there are various exclusion to this trend [9]. Inconsistent DNA methylation patterns have been the most common feature of many human disorders such as cancer, cardiovascular diseases, skin diseases, autoimmune diseases, neurological disorders and many others. These DNA methylation patterns can be analysed using various techniques like pyro-sequencing, methylation-specific polymerase chain reaction and bisulfite sequencing [6].

Studies of DNA methylation in CVD patients have indicated both increased and decreased DNA methylation but the data about the methylation of various genes is scattered across a large number of electronic publications, several specialized databases and other repositories. This makes searching of useful information a difficult task. Thus for easy retrieval of information that can contribute to development of better therapeutics for CVD, the objective

of our study is to create an efficient, effective and accurate system which will enable the extraction of the detailed information related to methylation of genes.

MethCard facilitates access and analysis of the relationships asserted between methylation information, human single nucleotide polymorphisms (SNPs) and other gene details with the observed disease conditions. It has been created with the help of information extracted from various databases namely Coronary Artery Disease Gene Database (CADgene), DisGeNET, Comparative Toxicogenomics Database (CTD), DiseaseMeth, Dragon Database for Methylated Genes and Diseases (DDMGD), Humsavar database etc. and various published literatures. It currently comprises of 4176 entries of altered gene methylation patterns that may be linked to cardiovascular disease development. It currently has information on 2145 unique genes associated with 43 different CVDs. It provides detailed insight into various features of genes and disease by providing links to various other databases such as NCBI (National Center for Biotechnology Information), HGNC (HUGO Gene Nomenclature Committee), dbSNP (Single Nucleotide Polymorphism database), Ensembl, UCSC Genome Browser, Vega, OMIM (Online Mendelian Inheritance in Man) and UniProt which can be useful in various studies.

REVIEW OF LITERATURE

EPIGENETICS AND DNA METHYLATION

Epigenetics refers to any process involving alteration of gene activity without change in DNA sequence. It leads to modifications that can be transmitted and reversed [10]. These processes are essential for many functions of organism and if they occur improperly, it can lead to major adverse health and behavioural effects. There are various types of epigenetic processes like DNA methylation, histone modifications, nucleosome positioning etc. but the best known epigenetic process is DNA methylation as it is easiest to study with existing technology, is DNA methylation [11].

DNA methylation (DNAm) is a process that involves covalent transfer of a methyl group to the 5' position of the cytosine ring in DNA by DNA methyl-transferases (DNMTs). DNAm in mammalian genomes occurs almost exclusively on cytosine residues in CpG dinucleotides in somatic cells but it also occurs in non-CpG cytosines in embryonic stem cells [12]. CpG dinucleotides within promoters tend to be unmethylated permitting the transcription of the associated gene. DNAm plays a crucial role gene expression, cell differentiation and embryonic development. Thus any defect in DNAm can significantly alter the development process and lead to various diseases [13][14].

DNA METHYLATION IN CARDIOVASCULAR DISEASES

Cardiovascular diseases include various diseases like coronary artery disease, stroke, arrhythmia, congenital heart disease, myocardial infarction, heart failure, hypertensive heart disease, cardiomyopathy, aortic aneurysms etc [15]. The major causes of their development are physical inactivity, smoking, unhealthy eating habits, obesity, family history and hypertension [16]. Recently, the impact of alteration in DNA methylation has emerged as an important regulatory player in CVD at different levels from pathophysiology to therapeutics [17]. In a recent research when normal heart tissues were compared with tissues from heart failure patients, differential methylation in angiogenesis-related genes in diseased tissue was found. In blood also, change in methylation of repetitive sequences like *LINE-1* and *ALU* has been associated with CVD [6].

Recent studies suggest that altered DNA methylation mechanism might be involved in the pathways by which environmental and lifestyle factors contribute to CVD development. It has been found that based on maternal exposures, the risk factors operating during fetal life can lead to increased chances of particular disease incidence during adulthood. Studies in humans and animals indicate that dietary habits can also induce DNA methylation changes likely contribute to CVD. Animal studies have showed that high intake of foods that donate methyl groups can cause clinically relevant phenotypic changes by DNA methylation modifications. For example, in sheep a maternal diet deficient in vitamin B12, folate and methionine led to drastic changes in CpG island DNA methylation in the fetal liver. The offspring produced were heavier and had higher adulthood blood pressure as compared to control animals. Evidence has been found that smoking leads to DNA methylation pattern alterations in several tissues. Lower methylation was seen in the coagulation factor II receptor-like 3 gene of smokers when compared to non-smokers. Environmental exposures like air pollution, particulate matter, heavy metals and arsenic can also alter DNA methylation in repetitive elements as well as in specific genes [6].

ONLINE RESOURCES AND TECHNIQUES FOR DATABASE DEVELOPMENT

Lots of online resources and techniques have been used for extraction of information and development of an effective and elaborate database of human DNA Methylation for genes associated with cardiovascular diseases. These are as follows:

- **Coronary Artery Disease Gene Database (CADgene)**

  CADgene database (www.bioguo.org/CADgene/) is a comprehensive resource for detailed information of more than 600 genes associated to coronary artery disease and thus facilitate a deeper insight into its genetic basis. It offers a user friendly web interface with multiple search options and elaborate information on CAD associated SNPs, gene ontology annotations and protein-protein interactions [18].

- **DisGeNET**

  DisGeNET (www.disgenet.org/) is an excellent platform to address a variety of questions related to genetic basis of human diseases and generation of hypothesis related to therapeutic action of drug and its adverse effects. It provides one of the most comprehensive collections of human gene-disease associations and homogeneously derived ontologies by integrating highly curated data from various repositories, published literature, animal models and GWAS catalogues. It also prioritizes associations of genes and diseases based on the score derived from

supporting evidence. This database can be accessed through web interface, a Cytoscape app or R package. [19][20].

- **Comparative Toxicogenomics Database (CTD)**

  CTD (ctdbase.org) houses around 24 million manually curated toxicogenomics relations between genes, proteins, phenotypes, diseases, Gene Ontology annotations, chemicals/drugs, taxa, pathways and interaction modules. It facilitates elaborate representation of gene–disease, chemical–gene and chemical–disease interactions. It also provides large amount of information about response of various genes and proteins of different species to environmental toxic agents [21].

- **Dragon Database of Methylated Genes and Diseases (DDMGD)**

  DDMGD (www.cbrc.kaust.edu.sa/ddmgd/) is a comprehensive repository of associations between methylated gene and diseases in various species which are ranked according to their confidence scores. It aims to facilitate disease diagnosis and treatment decisions by contributing to research in the field of epigenetics. It also allows submission of newly found associations and consists of most recent association data as compared to other databases [22].

- **DiseaseMeth**

  DiseaseMeth (bioinfo.hrbmu.edu.cn/diseasemeth) is a human disease methylation database which includes literature-based experimental information as well as large-scale methylation data. It consists of around 14,530 entries on the experimental information which was extracted through text mining from more than 25,000 published research paper on DNA methylation in PubMed. It holds about 175 large-scale methylation data sets for around 72 diseases, which were primarily collected from various websites and institutes namely ArrayExpress, Gene Expression Omnibus (GEO) and UCSC etc. and also allows analysis and visualization of these datasets [23].

- **Humsavar**

  It is a file that consists of 71,795 entries of all the single amino acid variants annotated in human UniProtKB/Swiss-Prot. These variants are classified into three categories which include disease associated mutations (26,874), neutral polymorphism (38,105) and unclassified variants (6,816). It also provides information about sequence position and amino acid substitution of these variants [24].

- **HUGO Gene Nomenclature Committee (HGNC)**

  HGNC (www.genenames.org/) has been created with the aim to provide a unique name and symbol to every human gene and currently consists of around 24000 genes. It provide detailed information related to gene like Gene Id, alias name, gene family name, karyotype, Pubmed Id, Ensembl Id, OMIM Id etc [25].

- **Database of CpG islands and Analytical Tool (DBCAT)**

  DBCAT (dbcat.cgm.ntu.edu.tw/) is a comprehensive database of characterization records of DNA methylation profiles of different human cancers. It is a web-based application that enables investigation into epigenetic regulation of human diseases by providing various user friendly tools. It consists of three parts ie. genome query browser, a CpG island finder, and methylation microarray data analyzer. It allows comparison of changes in methylation status of genes from different microarrays and also enables functional analysis. It provides information related to Gene, Chromosome, Entrez GeneID, Refseq ID, Description, Biological Process, Molecular Function, KEGG Pathway and Genome Position [26].

- **ENSEMBL**

  Ensembl (www.ensembl.org/) is a versatile browser which houses genomes of around 87 vertebrates thus aids research in field of evolution, comparative genomics, sequence polymorphisms and transcriptional regulation. It enables diverse features such as gene annotation, multiple alignment computation, regulatory function prediction and disease data collection. It not only allows known gene annotation but also provides novel gene prediction by incorporation of numerous resources and by using various methods like homology predictions and ab-initio predictions The tools offered by it includes Biomart, BLAT, BLAST, BLAT and the Variant Effect Predictor (VEP) [27].

- **Hypertext Markup Language (HTML)**

  It is most widely used markup language for developing web pages and web applications. Its elements act as the building blocks of HTML pages. It can be used to embed images and other interactive objects in the web page using various tags which are enclosed by angular brackets. These tags can also include various attributes which enables creation of schematic documents by denoting structural semantics for text such as headings, paragraphs, links, lists etc. Web browsers use these HTML tags to interpret the content of the page. With the inclusion of Cascading Style Sheets (CSS)

and JavaScript along with HTML looks, layout and content of web pages can be better controlled [28].

- **JavaScript (JS)**

  It is a high level, interpreted computer programming language which is employed and supported by most of the websites and web browsers. It was earlier used as a part of web browsers so that client-side could interact with user, control the browser and make the required changes in the content of the document that was displayed. It is object oriented scripting language which is dynamic, untyped, and consists of first-class functions. Its syntax was influenced by the language C. Although, it copies some names and conventions from Java, but both the languages are otherwise unrelated and highly differ in semantics. It is a multi-paradigm language, supporting object-oriented, imperative, and functional programming styles [29].

- **Cascading Style Sheets (CSS)**

  It is a style sheet language which aims to provide greater control over presentation and is used for describing the presentation of a web page. It is not only used to set the visual style of web pages and user interfaces written in HTML, XHTML and XML but also enables speech or other media rendering**.** It can be used to separate the content and presentation of the document including aspects such as the background images, colours, fonts and textures. It enables separation of presentation instructions from the HTML content in a separate file or section. It can also be used to adjust the web page according to the screen size of the device on which it is being viewed [30].

MATERIAL AND METHODS

Database Construction

MethCard is a human DNA Methylation database for genes associated with CVDs. It has been created with the help of information extracted from various databases and published literatures. The gene disease association data has been integrated from mainly three database namely Coronary Artery Disease Gene Database (CADgene), DisGeNET and Comparative Toxicogenomics Database (CTD). Dragon Database for Methylated Genes and Diseases (DDMGD) provides data about methylation type, Pubmed IDs and Gene expression and Diseasemeth provides information about methylation (%). SNP data has been retrieved from Humasavar database. The gene/disease ontology data ie. Pathway and molecular function has been integrated mainly from Database of CpG islands and Analytical Tool (DBCAT). Ensemble and HGNC provides information related to percentage of GC Content, Gene ID, RefSeq ID, Karyotype, Genome Position, Accession Numbers, Enzyme IDs, Ensembl Gene ID, UCSC ID, Vega ID, Gene Family ID, Gene Family Name, Accession Numbers, OMIM ID, HGNC ID, UniProt ID and ALIAS.

The data from different resources were heterogeneous so a uniform format to describe the diverse data was created and then the extracted information was stored in excel files. Data modelling was done to design a conceptual model of relation of different data items. The redundant and irrelevant data was removed and missing data was searched for separately. The excel file containing data from different sources were merged and the three final excel sheet namely Gene Methylation, Gene SNP data, General Gene Information were created with relevant information using various excel functions.

Gene Id	Gene	Disease	MESH ID (Disease ID)	GC Content (%)	Methylation (%)	Methylation	Gene Expression	Pathway ID (Pathway Name)
2	A2M	Cardiomegaly	<a href=\https://www.ncbi.nlm.	37.18	36.6	NA	NA	KEGG:04610 (Complement and coagulation cascades)
2	A2M	Ischemic stroke	<a href=\https://www.ncbi.nlm.	37.18	36.6	NA	NA	KEGG:04610 (Complement and coagulation cascades)
8086	AAAS	Aortic Aneurysm	<a href=\https://www.ncbi.nlm.	49.39	49.1	NA	NA	KEGG:04520 (Adherens junction), REACT:115566 (Cell Cycle)
8086	AAAS	Tachycardia	<a href=\https://www.ncbi.nlm.	49.39	49.1	NA	NA	KEGG:04920 (Adipocytokine signaling pathway), REACT:115566 (Cell Cycle)
13	AADAC	multiple myeloma	<a href=\https://www.ncbi.nlm.	35.33	43.6	demethylation	expression	KEGG:05221 (Acute myeloid leukemia)
13	AADAC	myeloma	<a href=\https://www.ncbi.nlm.	35.33	43.6	methylation	expression	KEGG:05221 (Acute myeloid leukemia)
19	ABCA1	Aortic Aneurysm	<a href=\https://www.ncbi.nlm.	43.42	50.2	methylation	expression	KEGG:04520 (Adherens junction), KEGG:02010 (ABC transporters)
19	ABCA1	Atherosclerosis	<a href=\https://www.ncbi.nlm.	43.42	50.2	methylation	expression	KEGG:02010 (ABC transporters)
19	ABCA1	coronary artery disease	<a href=\https://www.ncbi.nlm.	43.42	50.2	methylation	expression	KEGG:02010 (ABC transporters)
19	ABCA1	Ischemic stroke	<a href=\https://www.ncbi.nlm.	43.42	50.2	methylation	expression	KEGG:02010 (ABC transporters)
19	ABCA1	Myocardial Ischemia	<a href=\https://www.ncbi.nlm.	43.42	50.2	methylation	expression	REACT:1762 (3' -UTR-mediated translational regulation), KEGG:02010 (ABC transporters)
24	ABCA4	Congestive heart failure	<a href=\https://www.ncbi.nlm.	46.44	49.3	NA	NA	KEGG:02010 (ABC transporters)
5243	ABCB1	Ischemic stroke	<a href=\https://www.ncbi.nlm.	37.04	33.6	methylation	expression	KEGG:02010 (ABC transporters)
5243	ABCB1	multiple myeloma	<a href=\https://www.ncbi.nlm.	37.04	33.6	methylation	expression	KEGG:02010 (ABC transporters)
5243	ABCB1	Myocardial Ischemia	<a href=\https://www.ncbi.nlm.	37.04	33.6	methylation	expression	REACT:1762 (3' -UTR-mediated translational regulation), KEGG:02010 (ABC transporters)
5243	ABCB1	preeclampsia	<a href=\https://www.ncbi.nlm.	37.04	33.6	methylation	expression	KEGG:04960 (Aldosterone-regulated sodium reabsorption), KEGG:02010 (ABC transporters)
5243	ABCB1	Vasculitis	<a href=\https://www.ncbi.nlm.	37.04	33.6	methylation	expression	KEGG:02010 (ABC transporters)
10058	ABCB6	Cardiomyopathy	<a href=\https://www.ncbi.nlm.	54.46	51.8	NA	NA	KEGG:02010 (ABC transporters)
4363	ABCC1	Hypertension	<a href=\https://www.ncbi.nlm.	48.48	42.9	NA	NA	KEGG:02010 (ABC transporters)
85320	ABCC11	Vasculitis	<a href=\https://www.ncbi.nlm.	45.62	35.8	NA	NA	KEGG:02010 (ABC transporters)
8714	ABCC3	coronary artery disease	<a href=\https://www.ncbi.nlm.	51.95	52.1	NA	NA	KEGG:02010 (ABC transporters)
10057	ABCC5	Cardiomyopathy	<a href=\https://www.ncbi.nlm.	44.7	54.2	NA	NA	KEGG:02010 (ABC transporters)
368	ABCC6	Ischemic stroke	<a href=\https://www.ncbi.nlm.	49.86	56.3	NA	NA	KEGG:02010 (ABC transporters)
368	ABCC6	Myocardial Ischemia	<a href=\https://www.ncbi.nlm.	49.86	56.3	NA	NA	REACT:1762 (3' -UTR-mediated translational regulation), KEGG:02010 (ABC transporters)
6833	ABCC8	Vasculitis	<a href=\https://www.ncbi.nlm.	50	63.8	NA	NA	KEGG:02010 (ABC transporters)
10060	ABCC9	Cardiomyopathy	<a href=\https://www.ncbi.nlm.	36.12	35	NA	NA	KEGG:02010 (ABC transporters)
10060	ABCC9	Long QT Syndrome	<a href=\https://www.ncbi.nlm.	36.12	35	NA	NA	KEGG:04920 (Adipocytokine signaling pathway), KEGG:02010 (ABC transporters)
10060	ABCC9	Myocardial infarction	<a href=\https://www.ncbi.nlm.	36.12	35	NA	NA	KEGG:04920 (Adipocytokine signaling pathway), KEGG:02010 (ABC transporters)
10060	ABCC9	Tachycardia	<a href=\https://www.ncbi.nlm.	36.12	35	NA	NA	KEGG:04920 (Adipocytokine signaling pathway), KEGG:02010 (ABC transporters)
9619	ABCG1	Myocardial Ischemia	<a href=\https://www.ncbi.nlm.	51.02	49.6	NA	NA	REACT:1762 (3' -UTR-mediated translational regulation), KEGG:02010 (ABC transporters)
9429	ABCG2	Cardiomyopathy	<a href=\https://www.ncbi.nlm.	41.4	43.6	methylation	up-regulated	KEGG:02010 (ABC transporters)
9429	ABCG2	Ischemic stroke	<a href=\https://www.ncbi.nlm.	41.4	43.6	methylation	up-regulated	KEGG:02010 (ABC transporters)
9429	ABCG2	multiple myeloma	<a href=\https://www.ncbi.nlm.	41.4	43.6	methylation	up-regulated	KEGG:05221 (Acute myeloid leukemia), KEGG:02010 (ABC transporters)

Excel sheet for Methylation Data

Gene	Gene ID	dbSNP	AA-change
A2M	2	rs1800434	Arg704His
A2M	2	rs1800433	Cys972Tyr
A2M	2	rs669	Ile1000Val
A2M	2	rs226405	Asn639Asp
A2M	2	rs3180392	Leu815Gln
AAAS	8086	rs121918549	Gln15Lys
AAAS	8086	rs121918550	Ser263Pro
AAAS	8086	rs13330	Lys108Met
AADAC	13	rs1803155	Val281Ile
ABCA1	19	rs9282543	Val399Ala
ABCA1	19	rs2853574	Arg587Trp
ABCA1	19	rs137854496	Trp590Ser
ABCA1	19	rs2853578	Gln597Arg
ABCA1	19	rs28937313	Asn935Ser
ABCA1	19	rs137854495	Ala937Val
ABCA1	19	rs137854500	Asp1289Asn
ABCA1	19	rs137854494	Cys1477Arg
ABCA1	19	rs146292819	Asn1800His
ABCA1	19	rs2230806	Arg219Lys
ABCA1	19	rs9282541	Arg230Cys
ABCA1	19	rs758100110	Ala255Thr
ABCA1	19	rs2066718	Val771Met

Excel sheet for SNP Data

	A	B	C	D	E	F	G	H	I	J	K	L
1	Gene	Karyotype	Chromosome Numbe	Gene I	Name	Genome Position	Accession Numbers	Ensembl Gene ID	UCSC ID	OMIM ID	Vega ID	UniProt ID
2	A2M	12p13.31	12	<a href=\htt	alpha-2-macroglobul	chr12: 9153325-9161325	BX647329, X68728, M11313	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
3	AAAS	12q13	12	<a href=\htt	achalasia, adrenocort	chr12: 52001141-520031	AJ289841	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
4	AADAC	3q25.1	3	<a href=\htt	arylacetamide deacet	chr3: 153013051-153015	L32179	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
5	ABCA1	9q31.1	9	<a href=\htt	ATP-binding cassette	chr9: 106729757-10673	AJ012376	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
6	ABCA4	1p22.1	1	<a href=\htt	ATP-binding cassette	chr1: 94358793-943607	U88667	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
7	ABCB1	7q21.12	7	https://www	ATP-binding cassette	chr7: 87180000-871820	M14758	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
8	ABCB6	2q35	2	<a href=\htt	ATP binding cassette	chr2: 219791416-219793	AF070598	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
9	ABCC1	16p13.11	16	<a href=\htt	ATP-binding cassette	chr16: 15949435-159514	L05628	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
10	ABCC11	16q12.1	16	<a href=\htt	ATP-binding cassette	chr16: 46826089-46828	AF367202	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
11	ABCC3	17q21.33	17	<a href=\htt	ATP binding cassette	chr17: 46065727-46067	Y17151	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
12	ABCC5	3q27.1	3	<a href=\htt	ATP binding cassette	chr3: 185217321-185219	AF104342	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
13	ABCC6	16p13.11	16	<a href=\htt	ATP-binding cassette	chr16: 16224315-162263	AF076622	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
14	ABCC8	11p15.1	11	<a href=\htt	ATP-binding cassette	chr11: 17454525-174565	L78207	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
15	ABCC9	12p12.1	12	<a href=\htt	ATP-binding cassette	chr12: 21980395-219823	AF061323	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
16	ABCG1	21q22.3	21	<a href=\htt	ATP binding cassette	chr21: 42511577-425135	U34919	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
17	ABCG2	4q22.1	4	<a href=\htt	ATP-binding cassette	chr4: 89238535-893005	AF103796	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
18	ABCG5	2p21	2	<a href=\htt	ATP-binding cassette	chr2: 43918962-439209	T93792	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
19	ABCG8	2p21	2	<a href=\htt	ATP-binding cassette	chr2: 43918107-439201	AF320294	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
20	ABL1	9q34.12	9	<a href=\htt	ABL proto-oncogene	chr9: 132699152-132701	M14752	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
21	ABO	9q34.2	9	<a href=\htt	ABO blood group (trar	chr9: 135139951-135141	AF134415	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
22	ACADVL	17p13.1	17	<a href=\htt	acyl-CoA dehydroger	chr17: 7062377-706437	BC012912	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
23	ACAT1	11q22.3	11	<a href=\htt	acetyl-CoA acetyltrar	chr11: 107495368-10749	D90228	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
24	ACAT2	6q25.3	6	<a href=\htt	acetyl-CoA acetyltrar	chr6: 160101575-160103	AF356877	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
25	ACD	16q22	16	<a href=\htt	ACD, shelterin comple	chr16: 66251714-662537	AF070535	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
26	ACE	17q23.3	17	<a href=\htt	angiotensin I convertir	chr17: 58906666-58908	J04144	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
27	ACE2	Xp22	X	<a href=\htt	angiotensin I convertir	chrX: 15528558-155305	AF291820	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
28	ACOT2	14q24.3	14	<a href=\htt	acyl-CoA thioesterase	chr14: 73104032-731060	AY005822, AK001939	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
29	ACOT4	14q24.1	14	<a href=\htt	acyl-CoA thioesterase	chr14: 73126663-731286	BC031799	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
30	ACP1	2p25	2	<a href=\htt	acid phosphatase 1, s	chr2: 253396-255396	M87546	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
31	ACP5	19p13.2	19	<a href=\htt	acid phosphatase 5, t	chr19: 11548996-115509	X14618	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
32	ACSM3	16p13.11	16	<a href=\htt	acyl-CoA synthetase	chr16: 20681313-206833	D16350	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
33	ACTA1	1q42.13	1	<a href=\htt	actin, alpha 1, skeletal	chr1: 227635966-22763	J00068	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w
34	ACTA2	10q23.31	10	<a href=\htt	actin, alpha 2, smooth	chr10: 90701991-907035	X13839	<a href=\http://www.ense	<a href=\http:/	<a href=\http://	<a href=\http	<a href=\http://w

Excel sheet for General Gene information

	M	N	O	P	Q	R
	HGNC ID	Gene Family ID	PubMed II	RefSeq	Molecular Function	ALIAS
	<a href=\http:/	<a href=\http://www.ge	21383588	NM_000014	GO:0042803(protein homodir	FWP007, S8
	<a href=\http:/	<a href=\http://www.ge	19183758	NM_015665	GO:0005515(protein binding)	AAA; AAASb
	<a href=\http:/	<a href=\http://www.ge	11979358, 1222	NM_001086	GO:0003824(catalytic activity	DAC; CES5A
	<a href=\http:/	<a href=\http://www.ge	16414398	NM_005502	GO:0005524(ATP binding)	TGD; ABC1; C
	<a href=\http:/	<a href=\http://www.ge	15585077, 196	NM_000350	GO:0003674(molecular_func	FFM; RMP; A
	<a href=\http:/	<a href=\http://www.ge	16087761, 163	NM_000927	GO:0008559(xenobiotic-tran	CLCS; MDR1;
	<a href=\http:/	<a href=\http://www.ge	5647419	NM_005689	GO:0000166(nucleotide bindi	ABC; LAN; PF
	<a href=\http:/	<a href=\http://www.ge	17272743	NM_019899	GO:0016787(hydrolase activit	MRP; ABCC;
	<a href=\http:/	<a href=\http://www.ge	8038483	NM_032583	GO:0008514(organic anion tra	WW; EWWD; I
	<a href=\http:/	<a href=\http://www.ge	6973297	NM_003786	GO:0015238(drug transmemb	MLP2; MRP3
	<a href=\http:/	<a href=\http://www.ge	5647419	NM_005688	GO:0005524(ATP binding)	MRP5; SMRF
	<a href=\http:/	<a href=\http://www.ge	16806698	NM_001171	GO:0046982(protein heterodi	ARA; PXE; MI
	<a href=\http:/	<a href=\http://www.ge	11955656, 147	NM_000352	GO:0005524(ATP binding)	HI; SUR; HHF
	<a href=\http:/	<a href=\http://www.ge	5647419	NM_020298	GO:0005515(protein binding)	SUR2; ABC3
	<a href=\http:/	<a href=\http://www.ge	25890853, 255	NM_207174	GO:0046982(protein heterodi	ABC8; WHITE
	<a href=\http:/	<a href=\http://www.ge	10620750	NM_004827	GO:0042803(protein homodir	MRX; MXR; A
	<a href=\http:/	<a href=\http://www.ge	11138003	NM_022436	GO:0046982(protein heterodi	STSL
	<a href=\http:/	<a href=\http://www.ge	16632123	NM_022437	GO:0005515(protein binding)	GBD4
	<a href=\http:/	<a href=\http://www.ge	10850458, 1146	NM_005157	GO:0005515(protein binding)	ABL; JTK7; p
	<a href=\http:/	<a href=\http://www.ge	19591532, 2182	NM_020469	GO:0046872(metal ion bindin	GTB; NAGAT
	<a href=\http:/	<a href=\http://www.ge	11455412	NM_0010338	GO:0050660(flavin adenine d	ACAD6; LCAI
	<a href=\http:/	<a href=\http://www.ge	11340680, 252	NM_000019	GO:0005154(epidermal growt	T2; MAT; ACA
	<a href=\http:/	<a href=\http://www.ge	11724102	NM_005891	GO:0005515(protein binding)	NA
	<a href=\http:/	<a href=\http://www.ge	26345285	NM_022914	GO:0005515(protein binding)	Ptop, Pip1, Tp
	<a href=\http:/	<a href=\http://www.ge	10069205, 106	NM_000789	GO:0016787(hydrolase activit	ACE1, CD143
	<a href=\http:/	<a href=\http://www.ge	25301841	NM_021804	GO:0005515(protein binding)	ACEH
	<a href=\http:/	<a href=\http://www.ge	6723573	NM_006821	GO:0005102(receptor binding	Mte1, ZAP128
	<a href=\http:/	<a href=\http://www.ge	2004323	NM_152331	GO:0005515(protein binding)	FLJ31235, P1
	<a href=\http:/	<a href=\http://www.ge	20471113, 222(	NM_004300	GO:0005515(protein binding)	HAAP, LMW-
	<a href=\http:/	<a href=\http://www.ge	830309	NM_001611	GO:0016787(hydrolase activit	TRAP, HPAP
	<a href=\http:/	<a href=\http://www.ge	12390063, 124	NM_005622	GO:0016787(hydrolase activit	SA
	<a href=\http:/	<a href=\http://www.ge	5647419	NM_001100	GO:0005524(ATP binding)	NEM3
	<a href=\http:/	<a href=\http://www.ge	10231640	NM_001613	GO:0005515(protein binding)	ACTSA

Remaining Part of Excel sheet for General Gene Information

User Interface Construction

User Interface was constructed using the HTML, JavaScript and CSS. Four pages namely Home page, Search page, Information page and Contact page were constructed with the help of text editor Brackets.

Home page contains the Welcome page containing introduction to MethCard, Cardiovascular diseases and Methylation and Data source.

```
39          <!-----------------------------------------------Home Page Content-----------------------------------------------
            ---->
40          <div id="Hcontent" class="clearfix">
41              <div id="main-content">
42                  <h2> MethCard </h2>
43                  <p>
44                      MethCard is a comprehensive database of human DNA methylation for the genes associated with
                        cardiovascular diseases. The methylation data collected is an amalgamation of numerous published
                        literature, several databases and other resources. The database currently comprises of 4176 enteries of
                        altered gene methylation patterns that may be linked to cardiovascular disease development.
45                  </p>
46                  <h2> Cardiovascular Diseases and Methylation </h2>
47                  <p>
48                      Cardiovascular Diseases are a leading cause of deaths worldwide. Around 17.5 million people die from
                        CVDs every year that constitutes  around 30 percent of global deaths. Although various factors like
                        genetic, environmental, behavioural and clinical factors contribute to CVD development but recently
                        epigenetics has emerged as an important regulatory player in disease development.
49
50                      <br>
51                      Nowadays researchers are focussing on study of the impact change in DNA methylation as it is most
                        feasible to measure in an epidemiological setting. It plays a major role at different levels from
                        pathophysiology to therapeutics as it is associated to gene expression, gene regulation and cellular
                        specification.
52
53                  </p>
54                  <h2> Data Source </h2>
55                  <p>
56                      MethCard organizes the associations in a summary table that include Gene, Name, Gene ID, PubMedID,
                        Disease Name, MeSH ID, Methylation, Methylation(%), Gene Expression, GC Content(%), dbSNP ID (AA-
                        change), RefSeq ID, Karyotype, Genome Position, Accession Numbers, Enzyme IDs, Ensembl Gene ID, UCSC
                        ID, Vega ID, Gene Family ID, Gene Family Name, Accession Numbers, OMIM ID, HGNC ID, UniProt ID,
                        Molecular Function, Pathway, ALIAS from various database like DiseaseMeth, CADgene, DisGeNET, Dragon
                        Database for Methylated Genes and Diseases (DDMGD).
57      .
58                  </p>
59
```

Search is divided into three categories ie. Methylation search, SNP search and General Gene details. Database can be searched either by selecting gene name or by selecting disease.

```
                                <text class="dropbtn">Search</text>
                                <div class="dropdown-content">
                                    <a href="searchMethylation.html">Methylation Search</a>
                                    <a href="searchSNiPs.html">SNP Search</a>
                                    <a href="searchGeneInformation.html">Gerneral Gene Details</a>
                                </div>
                            </div>
                        </li>
                        <li> <a href="information.html">Information</a></li>
                        <li> <a href="contact.html">Contact</a> </li>
                        </h3>
                </ul>
            </div>
        </div>
        <!-----------------------------------------SearchContent---------------------------------------------------->
        <div id="GScontent" class="clearfix">
            <!---------------------------------GeneSearch---------------------------------------------->
            <select class="searchInput" id="myGeneInput">
                <option value="">Select Gene</option>
                <option value="A2M">A2M</option>
<option value="AAAS">AAAS</option>
<option value="AADAC">AADAC</option>
<option value="ABCA1">ABCA1</option>
<option value="ABCA4">ABCA4</option>
<option value="ABCB1">ABCB1</option>
<option value="ABCB6">ABCB6</option>
<option value="ABCC1">ABCC1</option>
<option value="ABCC11">ABCC11</option>
<option value="ABCC3">ABCC3</option>
<option value="ABCC5">ABCC5</option>
<option value="ABCC6">ABCC6</option>
<option value="ABCC8">ABCC8</option>
```

Information page contains detailed information about Methylation Type, Gene Expression, Methylation (%), Molecular Function, KEGG Pathway, REACTOME Pathway and MeSH ID.

```
        <!-----------------------------------------------------Information Content-----------------------------------------
        ------>
        <div id="Icontent" class="clearfix">
            <h2>Search Result </h2>
            <p>Search result is divided into 3 categories namely Methylation Search, SNP Search and General Gene details
            Search. These categories include information about Gene, Name, Gene ID, PubMed ID, Disease Name, MeSH ID,
            Methylation Type, Methylation(%), Gene Expression, GC Content(%), dbSNP ID (AA-change), RefSeq ID, Karyotype,
            Genome Position, Accession Numbers, Enzyme IDs, Ensembl Gene ID, UCSC ID(supplied by UCSC), Vega ID, Gene
            Family ID, Gene Family Name, Accession Numbers, OMIM ID(supplied by OMIM), HGNC ID, UniProt ID(supplied by
            UniProt), Molecular Function, Pathway, ALIAS.
            </p>
            <ul>
                <li>
                    <p> <b> Methylation Type</b> - It provides the type of methylation for each gene queried and has 7
                    types of methylation categories like demethylation, hypomethylation, hypermethylation, methylation,
                    unmethylation, dimethylation and trimethylation. </p>
                </li>
                <li>
                    <p><b> Gene Expression </b> - It provides the associations between gene methylation and gene expression
                    for the gene queried in a particular disease.</p>
                </li>
                <li>
                    <p> <b> Methylation(%) </b> - It provides the degree of methylation in percentage for each gene queried
                    in a particular disease.</p>
                </li>
                <li>
                    <p><b>  Molecular Function </b> - It is related to the activities that can be performed by individual
                    gene products or assembled complexes of gene products at the molecular level like binding or catalysis.
                    </p>
                </li>
                <li>
                    <p><b>KEGG Pathway  </b> - It provides manually drawn pathways representing molecular interaction and
                    reaction networks for Metabolism, Genetic Information Processing, Environmental Information Processing,
                    Cellular Processes, Organismal Systems and Human Diseases.</p>
                </li>
                <li>
                    <p><b>REACTOME Pathway</b> - It provides biological interactions networks like classical intermediary
                    metabolism, signaling, innate and acquired immune function, transcriptional regulation, apoptosis and
                    disease. These networks are formed by reactions between proteins, nucleic acids, complexes, small
                    molecules etc. </p>
```

Contact page contains information to contact for any queries.

```
                        <li> <a href="information.html">Information</a></li>
                        <li> <a href="contact.html">Contact</a> </li>
                        </h3>
                </ul>
            </div>
        </div>
        <!-----------------------------------------------------Content-------------------------------------------------
        >
        <div id="Ccontent" class="clearfix">
            <div id="contact-content">
                <h3> Contact :</h3>
                <b>Prof. Bansi D. Malhotra</b>
                <br> Department of Biotechnology
                <br> Delhi Technological University (formerly Delhi College of Engineering)
                <br> Shahbad Daulatpur, Main Bawana Road,
                <br> Delhi - 42, India
                <br> email: <i>bansi.malhotra@dce.ac.in</i>
                <hr>
                <b>Ayushi Garg</b>
                <br> M.tech 2nd year
                <br> Department of Biotechnology
                <br> Delhi Technological University (formerly Delhi College of Engineering)
                <br> Shahbad Daulatpur, Main Bawana Road,
                <br> Delhi - 42, India
                <br> email: <i>ayushi.garg206@gmail.com</i>
            </div>
        </div>
        <!-----------------------------------------------------Footer-------------------------------------------------->
        <div class="footer" class="clearfix">
            <div id="footernavigation">
                <ul>
                    <h4>
                        <li> <a href="home.html">Home</a> </li>
                        <li> <a href="#">Search</a> </li>
                        <li> <a href="information.html">Information</a> </li>
```

The formatting of visual styles like background, colours, fonts and images of database pages was done using Cascading Style Sheets (CSS).

```
body {
    background-image: url(background.jpg);
    background-attachment: fixed;
    background-position: center;
    background-repeat: repeat;


    width: 99%;
}


/************************ Header ************************/

.header {
    background: #43cea2;
    background: linear-gradient(to left, #9999ff, #e699ff);
    position: fixed;
    width: 1111px;
    height: 100px;
    margin: 0 auto;
    background-color: brown;
    padding-top: 10px;
}

.header #heading {
    text-align: center;
    margin-top: -10px;
    color: white;
}

.header #main-navigation {
    width: 100%;
    display: inline-block;
    margin-top: -45px;
}

.header #main-navigation ul {
    padding-bottom: 25px;
        margin-top: 35px;
}

```

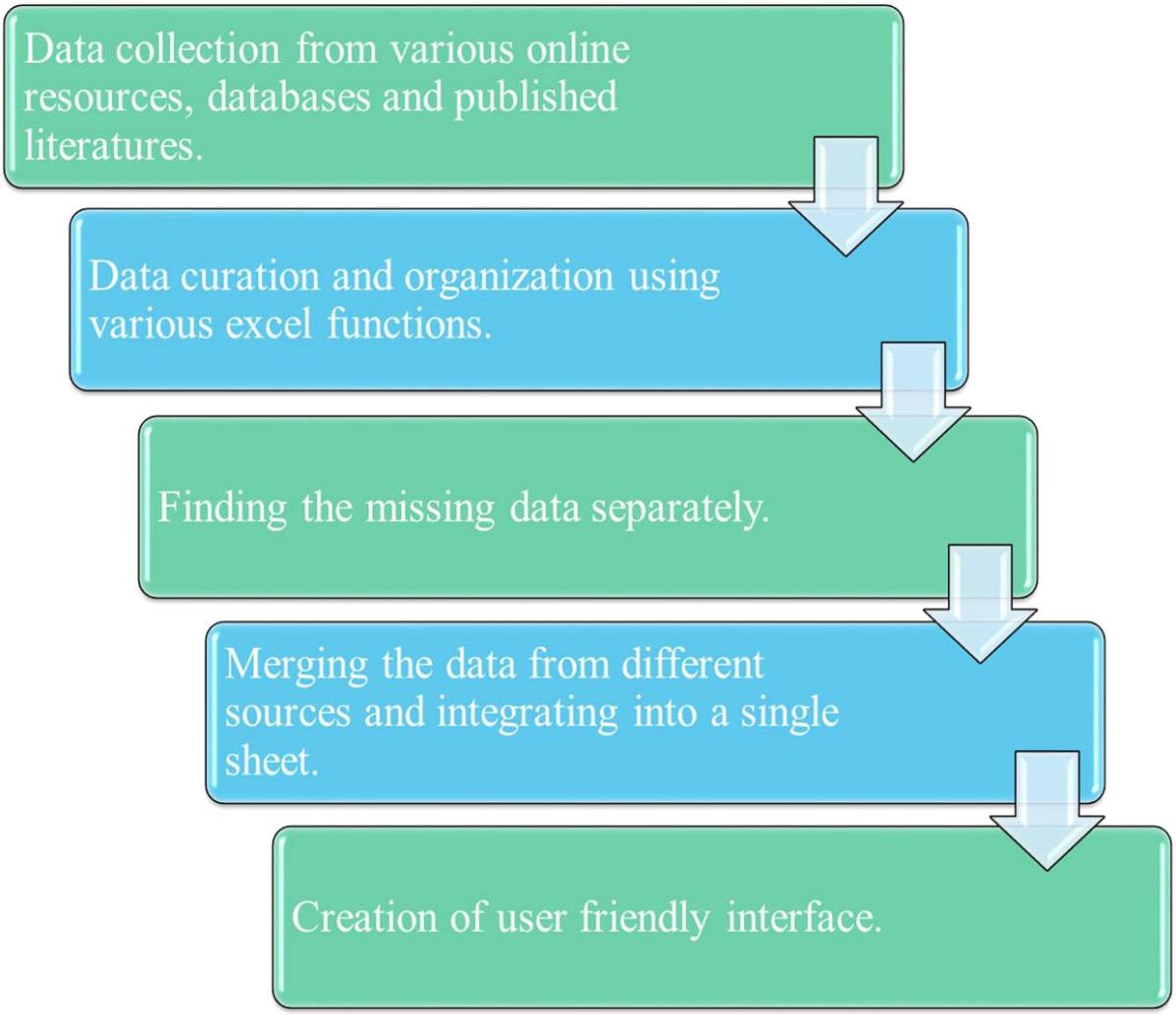
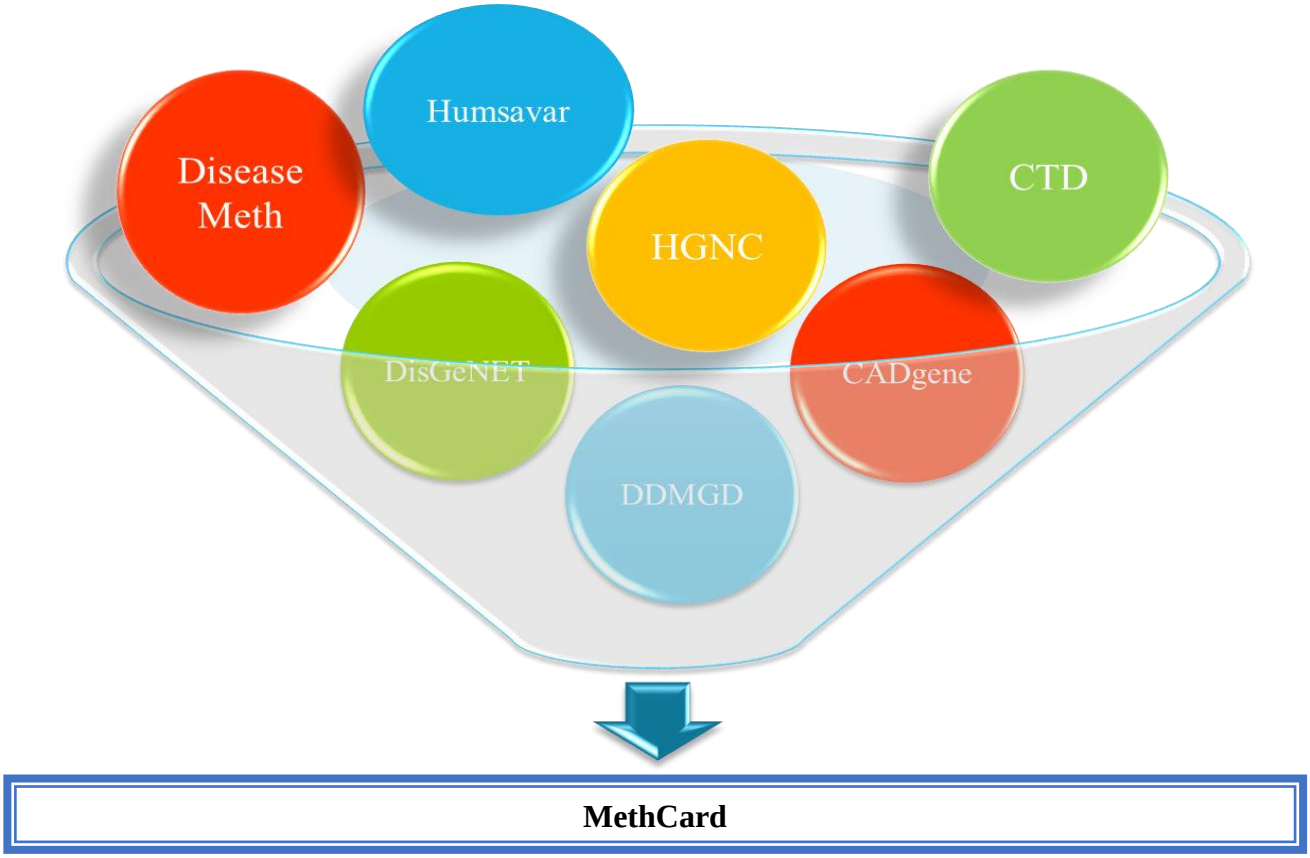


Figure 1: Flow of Methodology



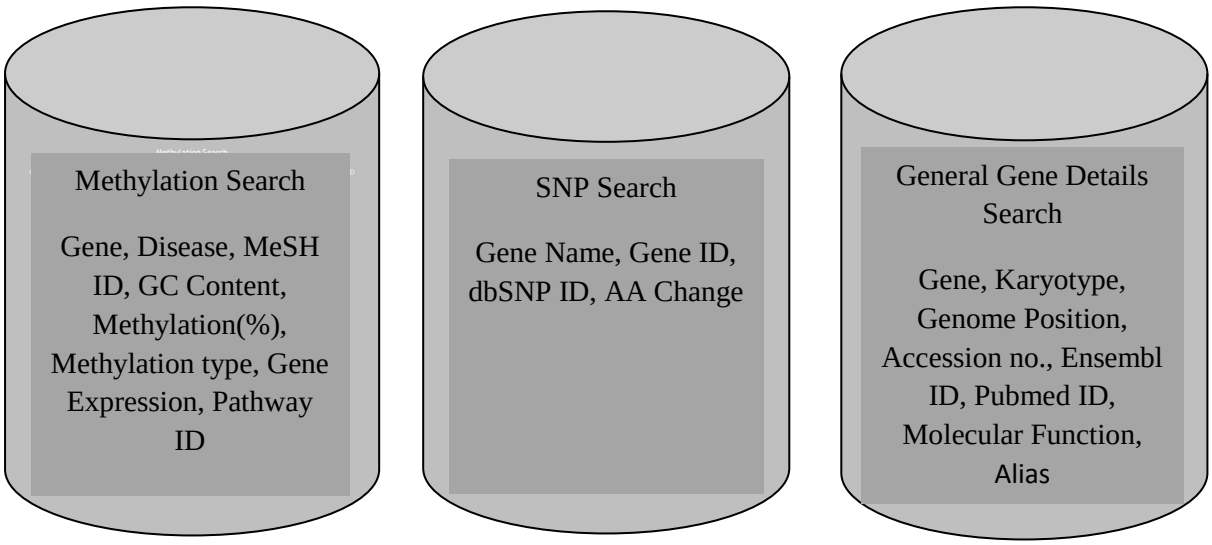


Figure 2: Schematic Representation of MethCard

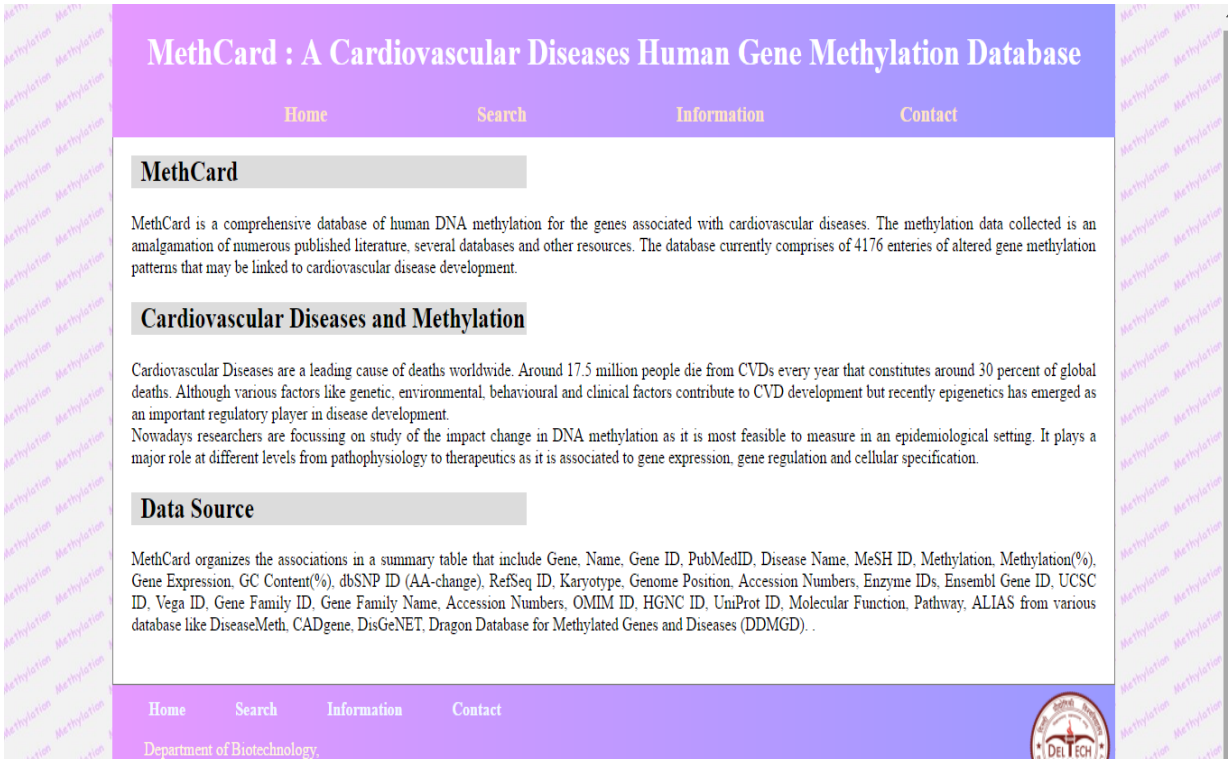
RESULTS

MethCard catalogues information about many of the cardiovascular diseases that are believed to be wide spread in the population across the world. It consists of methylation Patterns associated with CVDs like Coronary artery disease, Cardiomyopathy, Stroke, arrhythmia, congenital heart disease, myocardial infarction, heart failure, aortic aneurysms etc. The methylation data collected is an amalgamation of numerous published literature, several databases and other resources. The database currently comprises of 4176 entries of altered gene methylation patterns that may be linked to cardiovascular disease development. It currently has information on 2145 unique genes associated with 43 different CVDs.

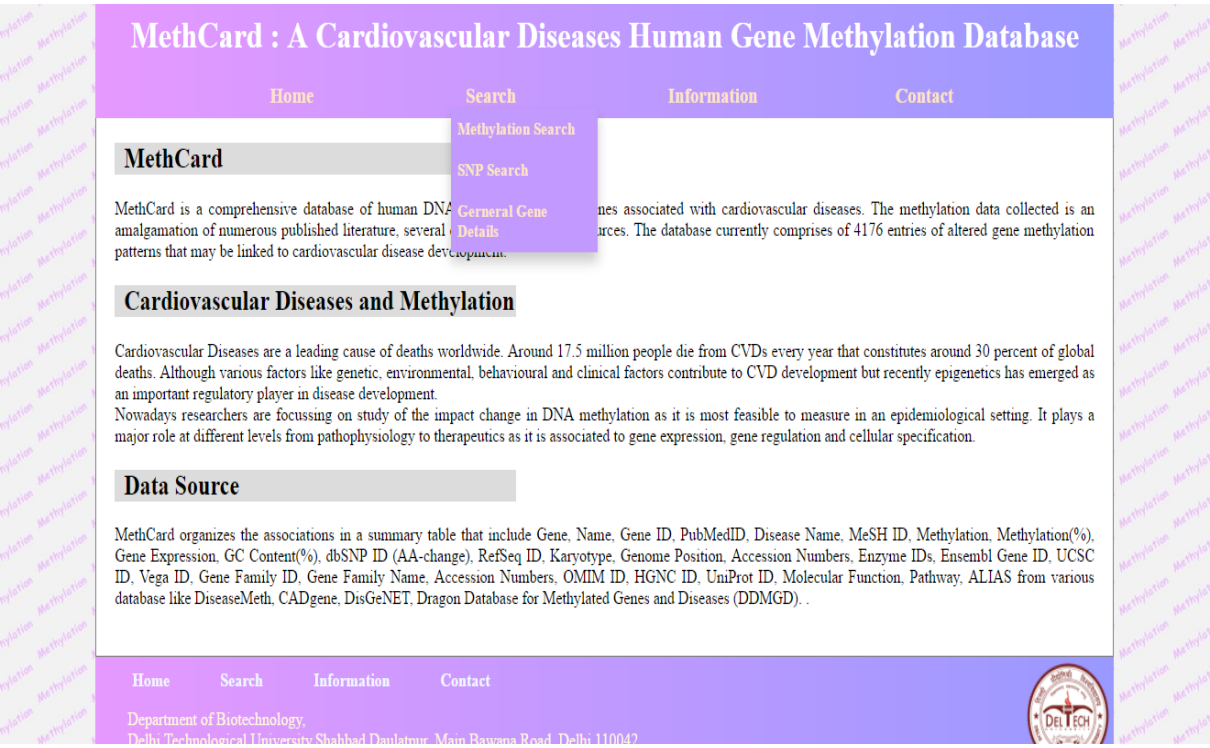
User Interface

It provides a user-friendly interface to query detailed information on each gene associated to CVDs. The result page also provide direct links to the source of information.

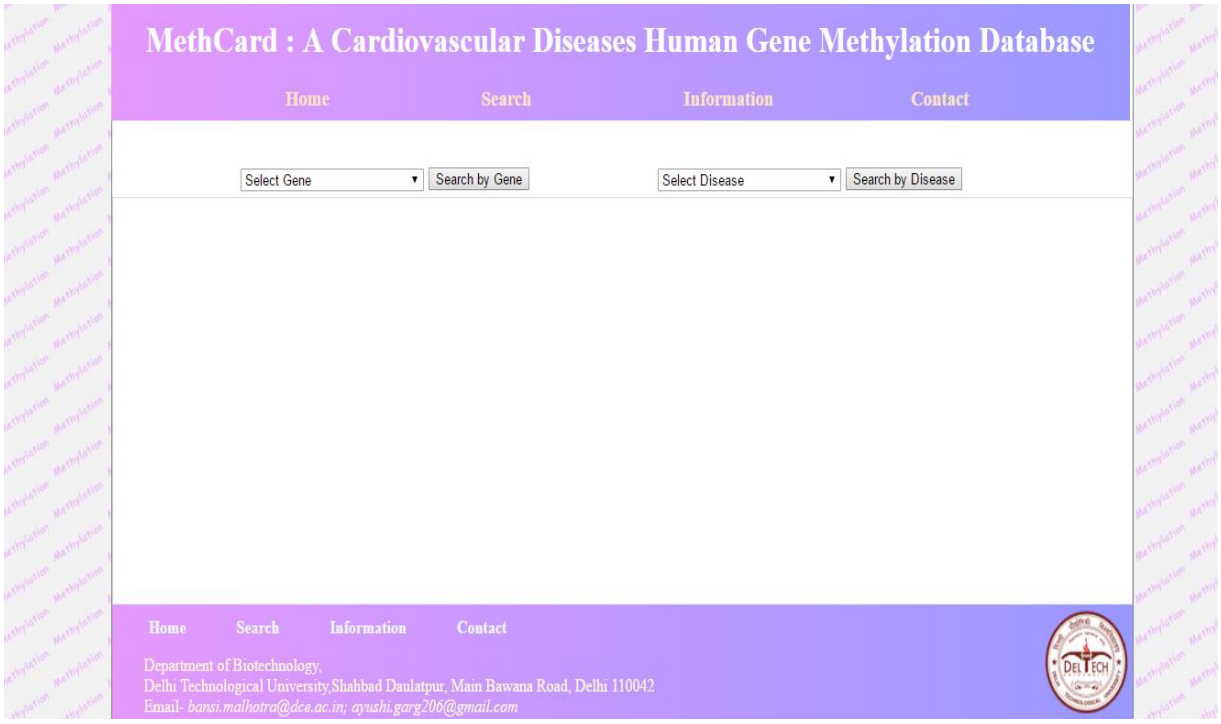
Home page consists of Welcome page containing introduction to MethCard, Cardiovascular diseases and Methylation and Data source.

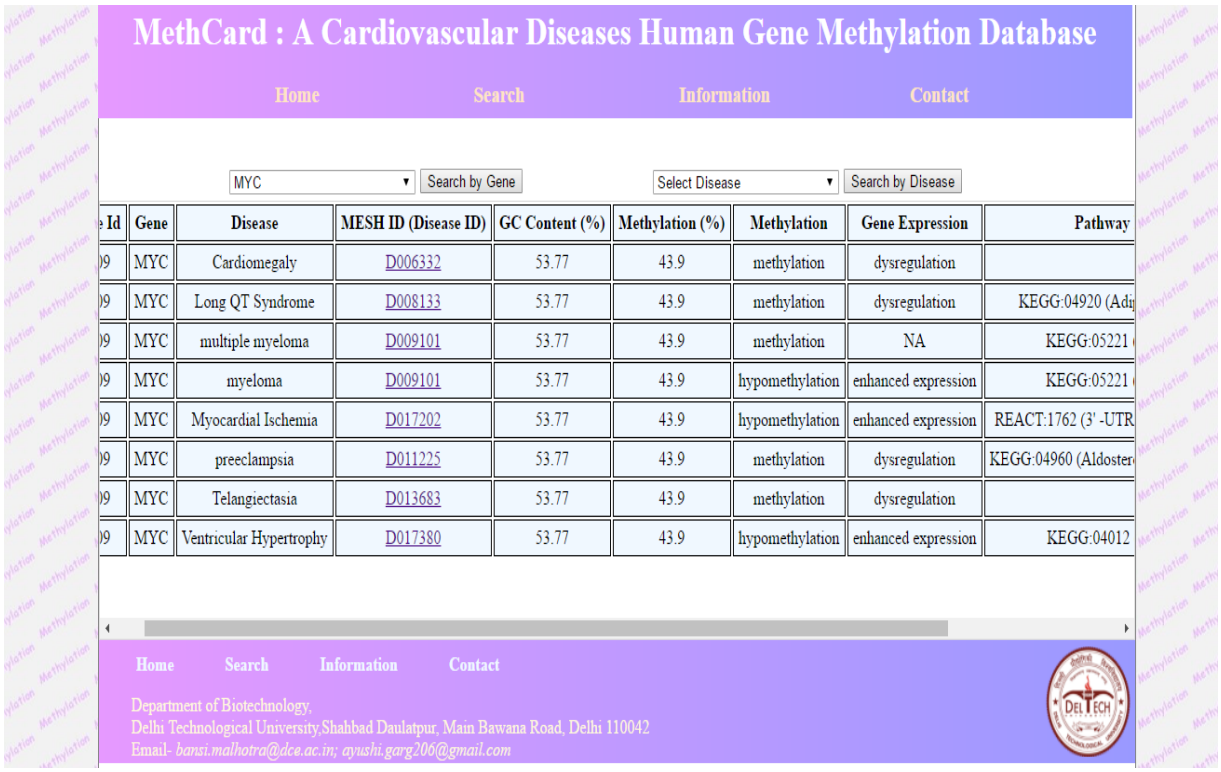


Search is divided into three categories ie. Methylation search, SNP search and General Gene details.



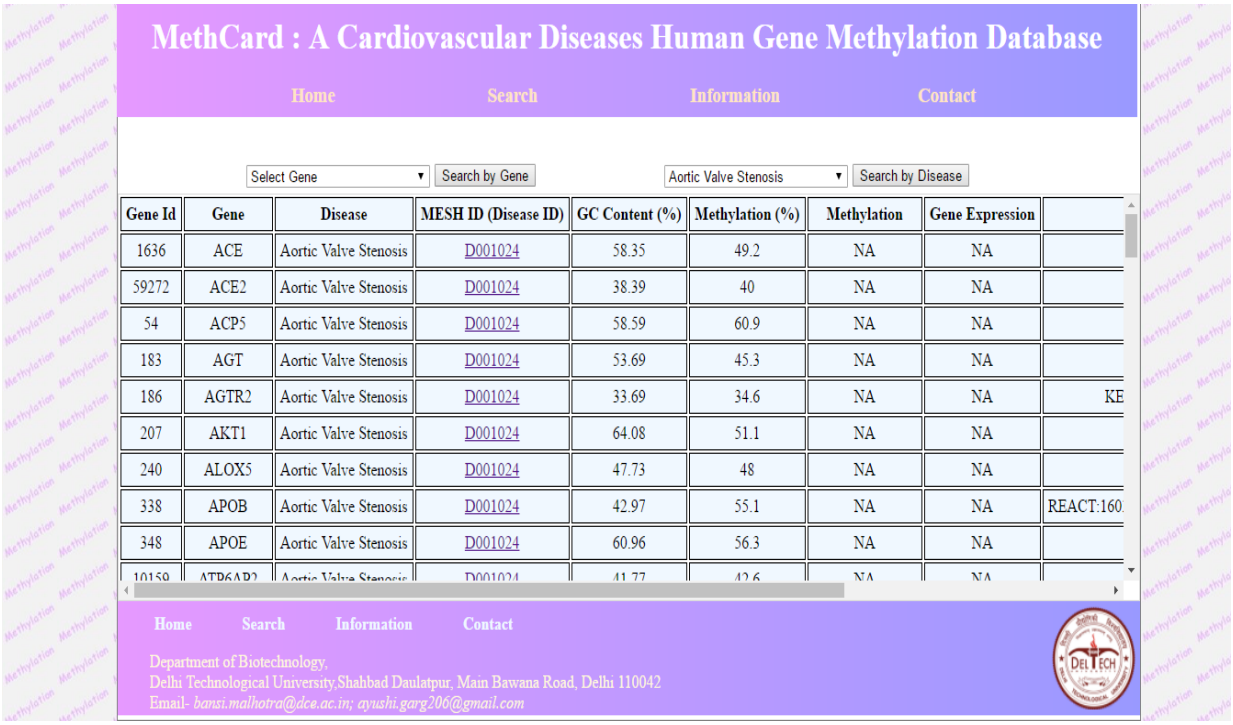
Database can be searched to retrieve the methylation data either by selecting gene name or by selecting disease. Methylation search provides the type of methylation for each gene queried. It has six methylation types categories namely demethylation, hypomethylation, hypermethylation, methylation, dimethylation and trimethylation and gives information about the degree of methylation in percentage for each gene in a particular disease. It provides the associations between gene methylation and gene expression for the gene in a particular disease. The result table also contains MeSH ID (Disease ID), GC Content (%) and the pathway (KEGG and REACTOME pathway) involved in disease development. KEGG (Kyoto Encyclopedia of Genes and Genomes) provides manually drawn pathways representing molecular interaction and reaction networks for Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Organismal Systems and Human Diseases [31]. REACTOME Pathway provides various biological interactions networks of signalling, metabolism, innate and acquired immunity, transcriptional regulation, disease and apoptosis. These networks are formed by reactions between proteins, nucleic acids, complexes, small molecules etc [32].





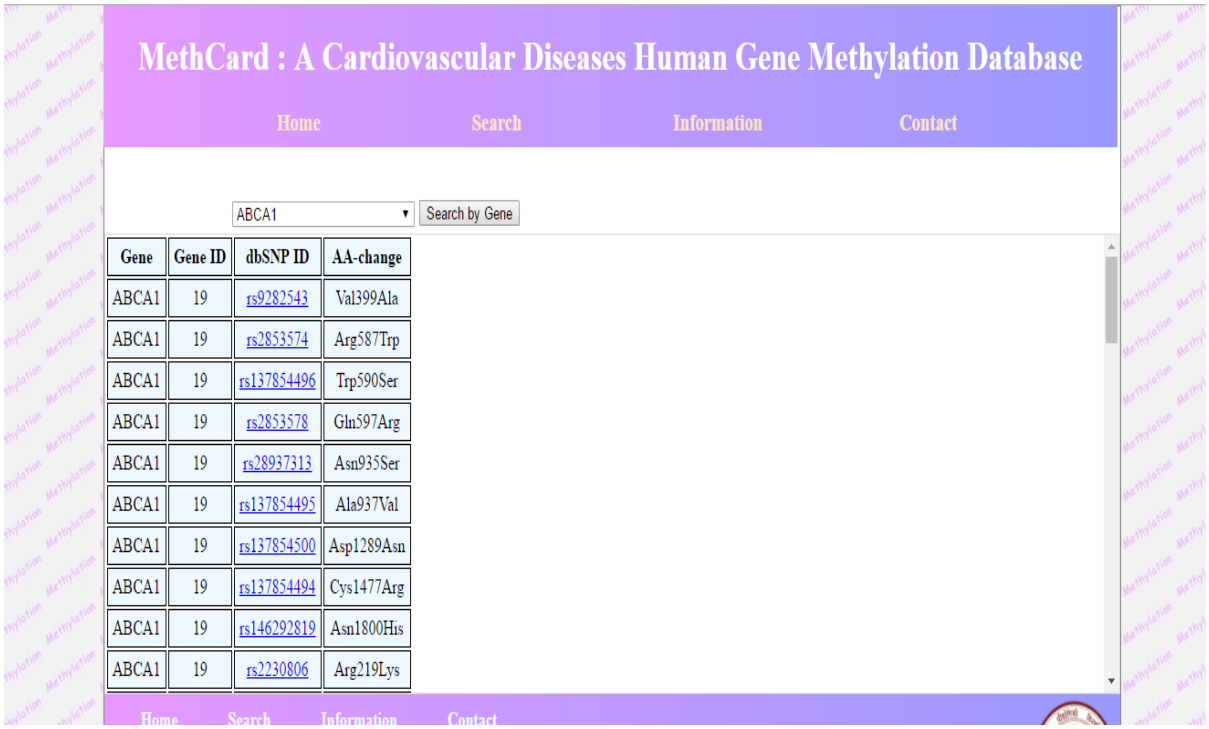
e Id	Gene	Disease	MESH ID (Disease ID)	GC Content (%)	Methylation (%)	Methylation	Gene Expression	Pathway
)9	MYC	Cardiomegaly	D006332	53.77	43.9	methylation	dysregulation	
)9	MYC	Long QT Syndrome	D008133	53.77	43.9	methylation	dysregulation	KEGG:04920 (Adi
)9	MYC	multiple myeloma	D009101	53.77	43.9	methylation	NA	KEGG:05221
)9	MYC	myeloma	D009101	53.77	43.9	hypomethylation	enhanced expression	KEGG:05221
)9	MYC	Myocardial Ischemia	D017202	53.77	43.9	hypomethylation	enhanced expression	REACT:1762 (3' -UTR
)9	MYC	preeclampsia	D011225	53.77	43.9	methylation	dysregulation	KEGG:04960 (Aldoster
)9	MYC	Telangiectasia	D013683	53.77	43.9	methylation	dysregulation	
)9	MYC	Ventricular Hypertrophy	D017380	53.77	43.9	hypomethylation	enhanced expression	KEGG:04012

Result for Methylation data retrieved from gene search

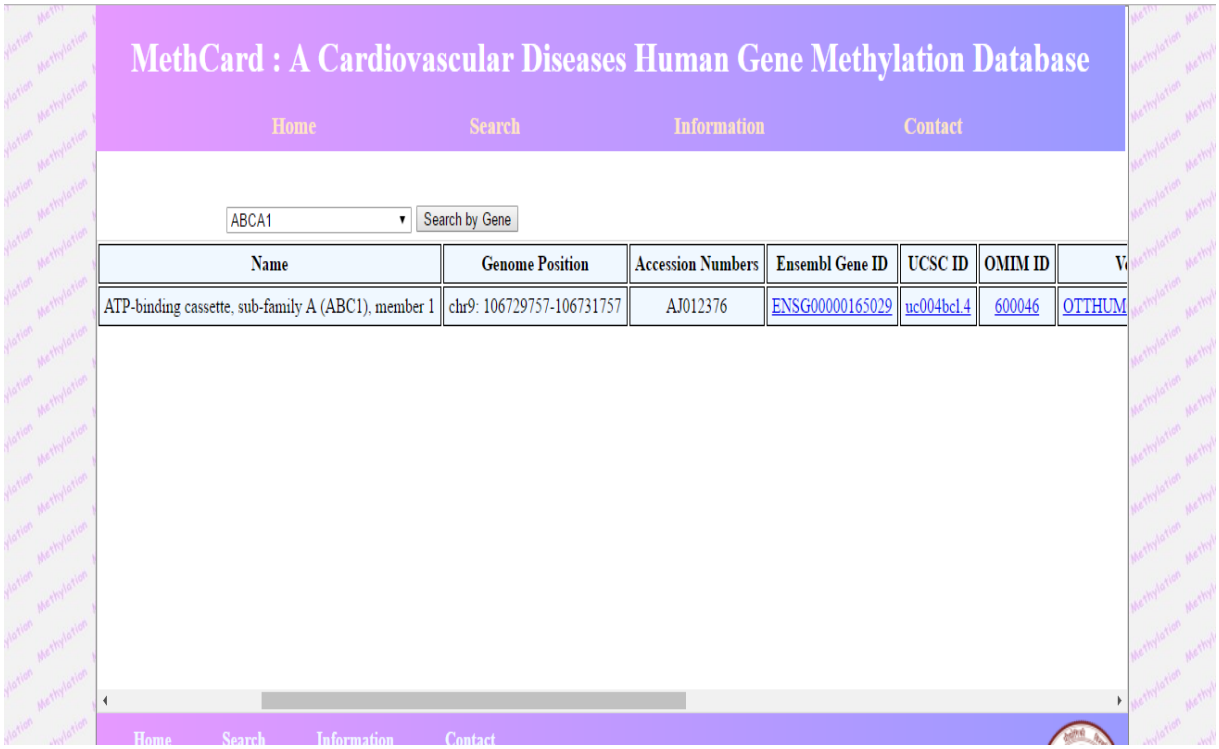


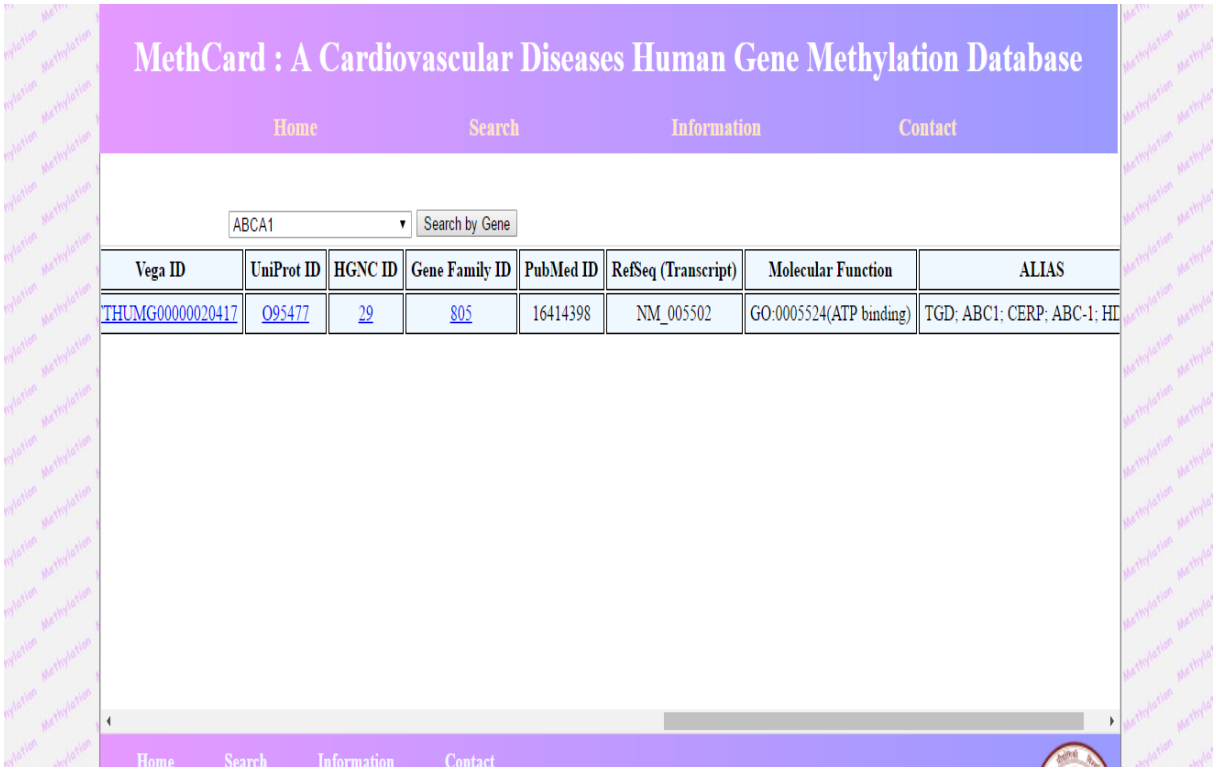
Result for Methylation data retrieved from disease search

There are about 10 million SNPs (single nucleotide polymorphisms) in human genome. They can be neutral ie. have no implications on gene function if present outside the coding region of gene or may play a critical role in disease development by affecting the protein function if present in coding or regulatory region of the gene. They as biomarkers by facilitating the location of genes associated with disease. They aid research in the field of pharmacogenomics by helping in prediction of an individual's response to drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases. They can also be used to track the inheritance of diseased genes [33][34]. MethCard's SNP search gives information about rs IDs, amino acid change and position of SNPs. It provides links to dbSNP thus giving deeper insight into SNP characteristics like clinical significance, minor allele frequency, neighbouring SNPs, DNA sequence etc.

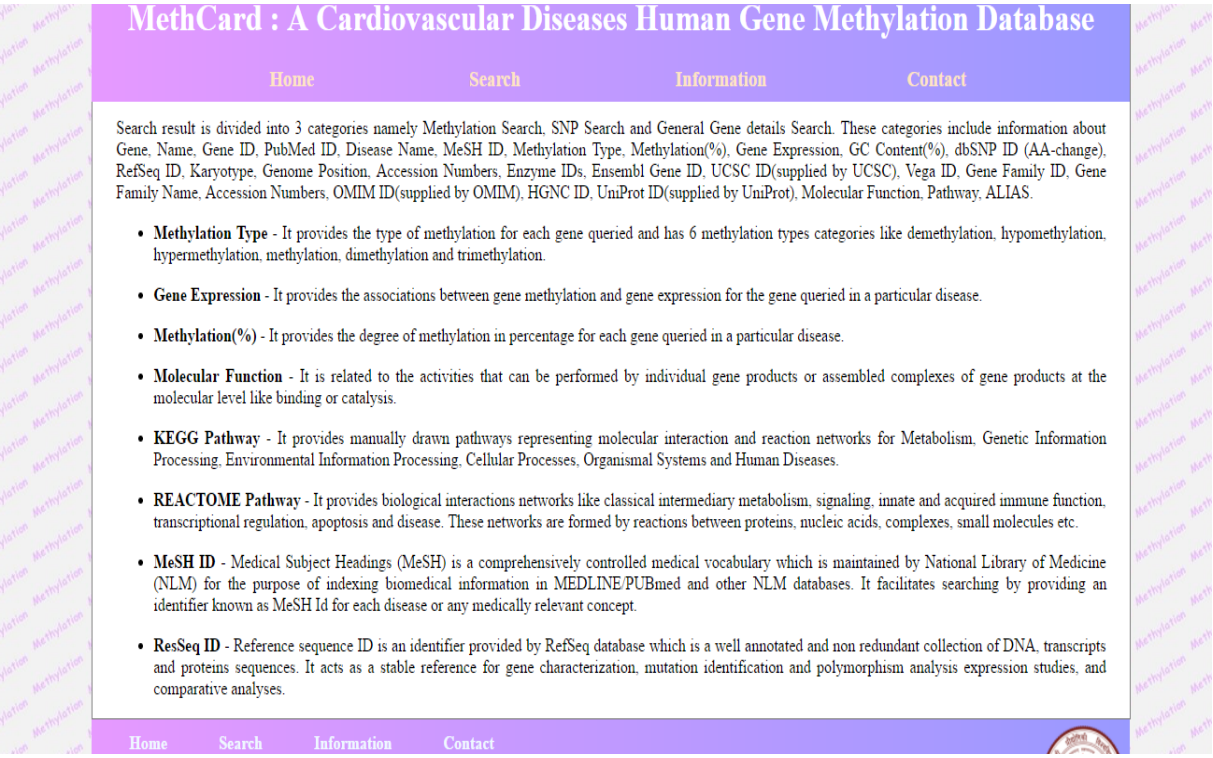


General gene details search result table contains various sections, namely Gene name, Gene ID, Karyotype, Chromosome Number, Genome Position, Molecular Function, HGNC ID, Ensembl ID, Vega ID, OMIM ID, Entrez Gene ID, UniProt ID, Gene Family ID, UCSC ID, Enzyme IDs, Alias/Synonyms, Refseq ID, PubMed IDs and Accession Numbers. Molecular function tells about the activities that can be performed by individual gene products or assembled complexes of gene products at the molecular level like binding or catalysis [35]. Reference sequence ID is an identifier provided by RefSeq database which is a well annotated and non-redundant collection of DNA, transcripts and proteins sequences enabling gene characterization, mutation identification etc [36]. It also gives links to databases which provide the above mentioned IDs thus providing greater insight into the gene information like sequence, annotation, gene products, gene disease associations, gene family name and other members and published literature.


MethCard : A Cardiovascular Diseases Human Gene Methylation Database
Home
Search
Information
Contact
ABCA1
Search by Gene
Name
Genome Position
Accession Numbers
Ensembl Gene ID
UCSC ID
OMIM ID
ATP-binding cassette, sub-family A (ABC1), member 1
chr9: 106729757-106731757
AJ012376
ENSG00000165029
uc004bcl.4
600046
OTTHUM
Home
Search
Information
Contact

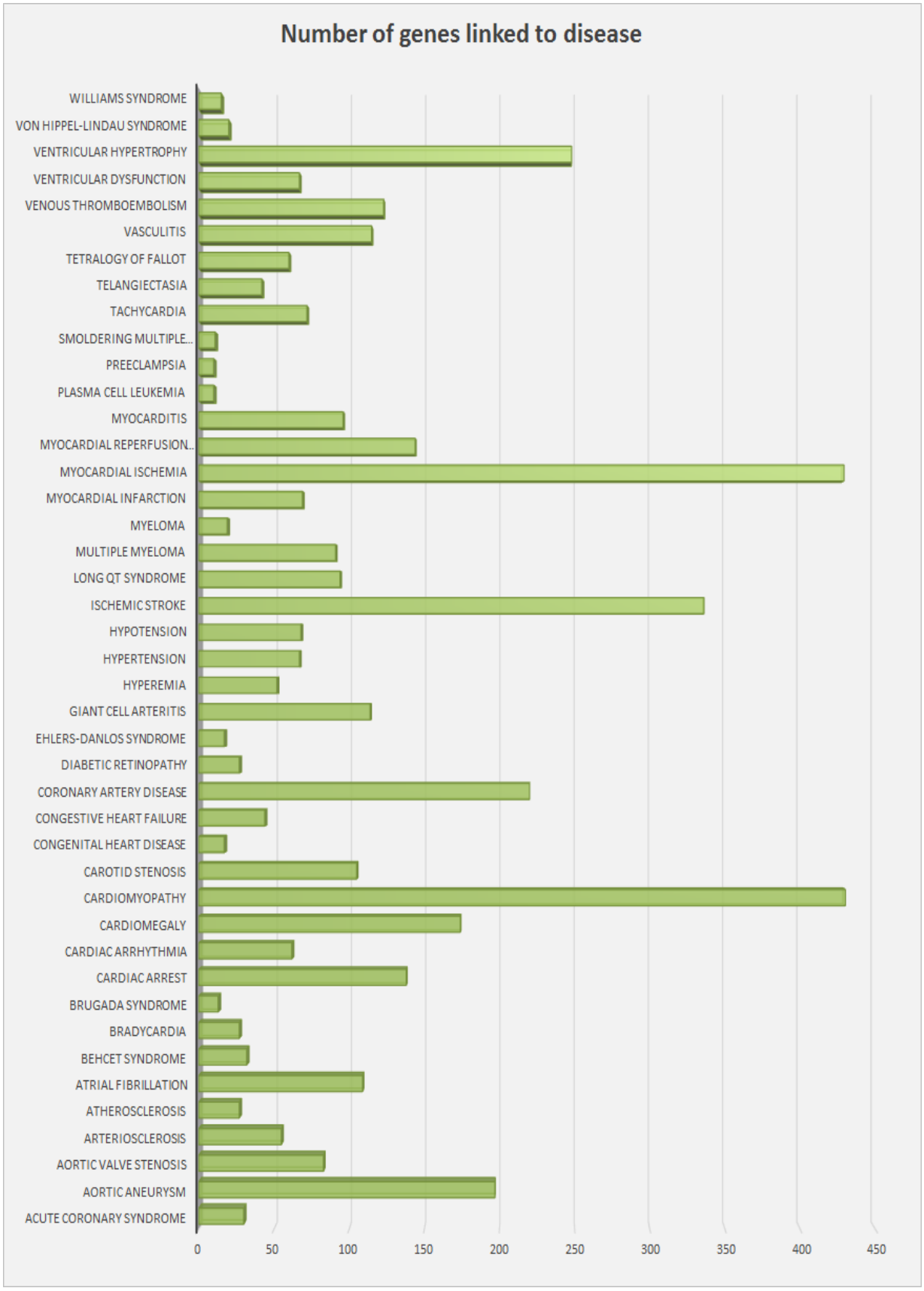

MethCard : A Cardiovascular Diseases Human Gene Methylation Database
Home
Search
Information
Contact
ABCA1
Search by Gene
Vega ID
UniProt ID
HGNC ID
Gene Family ID
PubMed ID
RefSeq (Transcript)
Molecular Function
ALIAS
THUMG00000020417
O95477
29
805
16414398
NM_005502
GO:0005524(ATP binding)
TGD; ABC1; CERP; ABC-1; HD
Home
Search
Information
Contact

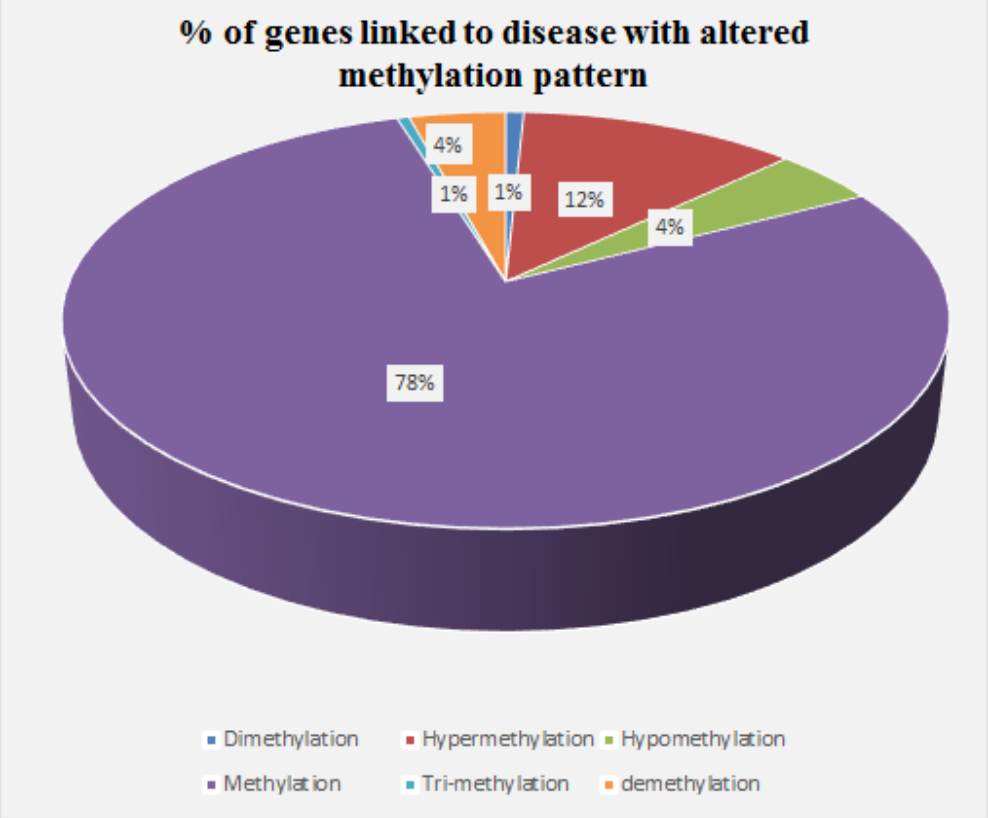
Information page contains detailed information about Methylation Type, Gene Expression, Methylation (%), Molecular Function, KEGG Pathway, REACTOME Pathway, MeSH ID and RefSeq ID.

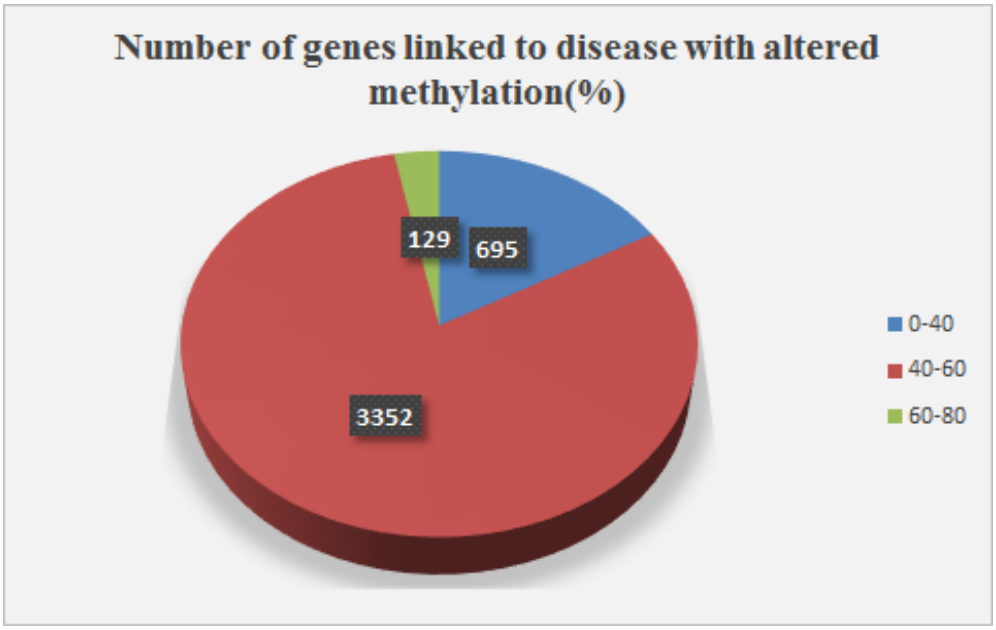


Contact page contains information to contact for any queries.

The results of MethCard has been summarized in the following graphs.




% of genes linked to disease with altered methylation pattern
4%
1%
1%
12%
4%
78%
Dimethylation
Hypermethylation
Hypomethylation
Methylation
Tri-methylation
demethylation


Number of genes linked to disease with altered methylation(%)
129
695
3352
0-40
40-60
60-80

CONCLUSION

MethCard is a first of its type, user-interactive repository of the genes associated with cardiovascular diseases has been developed. This database enables the user to retrieve elaborate methylation information of genes associated with cardiovascular diseases at genome-wide systems level. Mining the database for biologically meaningful data is likely to reveal hitherto unknown facts about the underlying causes of cardiovascular diseases. It is believed that the present database may be used to uncover hidden links between various cardiovascular diseases and methylation pattern providing valuable perspective to physicians, counsellors and biomedical researchers as epigenome can serve as a modifiable target for intervention. Thus can be used by medical community to formulate better prevention and treatment methods.

DISCUSSION AND FUTURE PROSPECTS

The contribution of DNA methylation to the cardiovascular diseases development is an active and exciting yet intricate and complex field of research. DNA methylation analysis and measurement requires careful considerations. Easy access to cheaper and more precise methods for DNA methylation quantification and enhanced ability to perform genome-wide methylation analysis on limited tissue quantities will enable faster and more accurate mapping of DNA methylation pattern in cardiovascular disease patients. This will provide deeper insights into etiology of cardiovascular diseases and open up new avenues of drug development and targeted therapies. Thus to aid efforts in this direction, an integrated research podium for learning the key interplay of DNA methylation and gene expression in cardiovascular diseases has been developed. Future efforts will be continued to update the database to have an improvised data quality and database functionality.

REFERNCES

1) Mozaffarian D., Benjamin E.J., Go A.S., Arnett D.K., Blaha M.J., Cushman M., de Ferranti S., Després J.P., Fullerton H.J., Howard V.J., Huffman M.D., Judd S.E., Kissela B.M., Lackland D.T., Lichtman J.H., Lisabeth L.D., Liu S., Mackey R.H., Matchar D.B., McGuire D.K., Mohler E.R., Moy C.S., Muntner P., Mussolino M.E., Pandey D.K., Reeves M.J., Rodriguez C.J., Sorlie P.D., Stein J., Towfighi A., Turan T.N., Virani S.S., Willey J.Z., Woo D., Yeh R.W. and Turner MB.; Heart Disease and Stroke Statistics – At-a-Glance. Journal of American Heart Association. 1-4; 2015.

2) Rourke K.O., Zanden A.V., Shepard D. and Kemon K.L.; Cardiovascular Disease Worldwide, 1999-2013. Journal of the American Medical Association. 314(18):1905; 2016.

3) Prabhakaran D., Jeemon P. and Roy A.; Cardiovascular Diseases in India Current Epidemiology and Future Directions. Journal of American Heart Association. 133:1605–1620; 2016

4) Gupta S., Gudapati R., Gaurav K. and Bhise M.; Emerging risk factors for cardiovascular diseases: Indian context. Indian Journal of Endocrinology and Metabolism. 17(5): 806–814; 2013.

5) McGill H.C., McMahan C.A. and Gidding S.S.; Preventing Heart Disease in the 21st Century Implications of the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) Study. Journal of American Heart Association. 117:1216-1227; 2008.

6) Zhong J., Agha G. and Baccarelli A.A.; The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies. Circulation Research. 118(1): 119–131; 2016.

7) Apurva S.V. and Marsden P.A.; Epigenetics in cardiovascular disease. Current Opinion in Cardiology. 26(3): 209-215; 2011.

8) Webster A.L.H., Yan M.S.C. and Marsden P.A.; Epigenetics and Cardiovascular Disease. Canadian Journal of Cardiology. 29(1): 46-57; 2013.

9) Moore L.D., Le T. and Fan G.; DNA Methylation and Its Basic Function. Neuropsychopharmacology. 1-6; 2012.

10) Epigenetics: The Science of Change. Environmental Health Perspectives. 114(3): 160-167; 2006.

11) Portela A. and Esteller M.; Epigenetic Modifications and Human Disease. Nature Biotechnology. 28: 1057-1068; 2010.

12) Bird A.; DNA methylation patterns and epigenetic memory. Genes & Development. 16:6–21; 2002.

13) Jin B., Li Y. and Robertson K.D.; DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy. Genes & Cancer. 2(6): 607–617; 2011.

14) Kim J.K., Samaranayake M. and Pradhan S.; Epigenetic mechanisms in mammals. Cellular and Molecular Life Sciences. 66: 596 – 612; 2009.

15) Buttar H.S., Li T. and Ravi N.; Prevention of cardiovascular diseases: Role of exercise, dietary interventions, obesity and smoking cessation. Experimental &Clinical Cardiology. 10(4): 229-249; 2005.

16) Kim M., Long T.I., Arakawa K., Wang R., Yu M.C., Laird P.W.; DNA Methylation as a Biomarker for Cardiovascular Disease Risk. PLoS One. 5(3): 1-8; 2010.

17) Khalil C.A.; The emerging role of epigenetics in cardiovascular disease. Therapeutic Advances in Chronic Disease. 5(4) 178–187; 2014.

18) Liu H., Liu W., Liao Y., Cheng L., Liu Q., Ren X., Shi L., Tu X., Wang Q.K. and An-Guo Y.; CADgene: a comprehensive database for coronary artery disease genes. Nucleic Acids Research. 39: 991-996; 2010.

19) Ero J.P., Rosinach N.R.Q. , Bravo A.L., Pons J.D., Mehren A.B., Baron M., Sanz F. and Furlong L.I.; DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database. 1–17; 2015.

20) Ero J.P., Bravo A.L., Rosinach N.R.Q., Sacrista A.G.R., Pons J.D, Centeno E., Garc G.J., Sanz F. and Furlong L.I.; DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Research. 45: 833–839; 2016.

21) Davis A.P., Grondin C.J., Hopkins K.L., Richards C.S., Sciaky D., King B.L., Wiegers T.C. and Mattingly C.J.; The Comparative Toxicogenomics Database's 10thyear anniversary: update 2015. Nucleic Acids Research. 43: 914-920; 2015.

22) Raies A.B., Mansour H., Incitti R. and Bajic V.B.; DDMGD: the database of text-mined associations between genes methylated in diseases from different species. Nucleic Acids Research. 43: 879–886; 2015.

23) Lv J., Liu H., Su J., Wu X., Liu H., Li B., Xiao X., Wang F., Wu Q. and Zhang Y.; DiseaseMeth: a human disease methylation database. Nucleic Acids Research. 40: 1030–1035; 2012.

24) Wu C.H., Apweiler R., Bairoch A., Natale D.A., Barker W.C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Rodrigo Lopez R., Magrane M., Martin M.J., Mazumder R., Donovan C., Redaschi N. and Suzek B.; The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Research. 34(1): 187-191; 2006.

25) Bruford E.A., Lush M.J., Wright M.W., Sneddon T.P., Povey S. and Birney E.; The HGNC Database in 2008: a resource for the human genome. Nucleic Acids Research. 36: 445–448; 2008.

26) Kuo H.C., Lin P.Y., Chung T.C., Chao C.M., Lai L.C., Tsai M.H. and Chuang E.Y.; DBCAT: database of CpG islands and analytical tools for identifying comprehensive methylation profiles in cancer cells. Journal for Computational Biology. 18(8): 1013-1017; 2011.

27) Hubbard T., Barker D., Birney E., Cameron G., Chen Y., Clark L., Cox T., Cuff J., Curwen V., Down T., Durbin R., Eyras E., Gilbert J., Hammond M., Huminiecki Y., Kasprzyk A., I. Vastrik and M. Clamp. The Ensembl genome database project. Nucleic Acids Research. 30(1): 38-41; 2002.

28) Freitag D.; Information Extraction from HTML: Application of a General Machine Learning Approach. Association for the Advancement of Artificial Intelligence. 1-7; 1998.

29) Jensen S.H., Moller A. and Thiemann P.; Type Analysis for Javascript. International Static Analysis Symposium. 5673: 238-255; 2009.

30) Condo F.; Cascading Style Sheets (CSS). 1-10; 2004.

31) Kanehisa M., Araki M., Goto S., Hattori M., Hirakawa M., Itoh M., Katayama T., Kawashima S., OkudaS., Tokimatsu T. and Yamanishi Y.; KEGG for linking genomes to life and the environment. Nucleic Acids Research. 36: 480–484; 2008.

32) Croft D., Kelly G., Wu G., Haw R., Gillespie M., Matthews L., Caudy M., Garapati P., Gopinath G., Jassal B., Jupe S., Kalatskaya I., Mahajan S., May B., Ndegwa N., Schmidt E., Shamovsky V., Yung C., Birney E., Hermjakob H., Eustachi P. and Stein L.; Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Research. 39: 691-697; 2011.

33) Hill L.; Help Me Understand Genetics Genomic Research. Genetics Home Reference. 1-14; 2017.

34) The International SNP Map Working Group; A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 409: 928-933; 2001.

35) Gene Ontology Consortium; The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research. 32: 258-261; 2004.

36) Leary N.A., Wright M.W., Brister J.R., Ciufo S., Haddad D., McVeigh R., Rajput B., Robbertse B., White B.S., Adjei D.A., Astashyn A., Badretdin A., Bao Y., Blinkova O., Brover V., Chetvernin V., Choi J., Cox E., Farrell C.M., Goldfarb T., Gupta T., Joardar V.S., Kodali V.K., Li W., Maglott D., Masterson P., McGarvey K.M., Murphy M.R., Neill K., Pujar S., Rangwala S.H., Rausch D., Shkeda A., Storz S.S., Sun H., Thibaud-issen F., Tolstoy I., Tully R.E., Vatsan A.R., Wallin C., Webb D., Wu W., Landrum M.J., Kimchi A., Tatusova T., DiCuccio M., Kitts P., Murphy T.D. and Pruitt K.D.; Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research. 44: 733-745; 2016.