SYSTEMS BIOLOGY, SINGLE NUCLEOTIDE POLYMORPHISM (SNP) AND GENE EXPRESSION ANALYSIS IN CORONARY ATHEROSCLEROSIS

[1] DR. SHOBHA MALVIYA
[1] Research Supervisor, Sri Satya Sai University of Technology and Medical Science, Sehore, Bhopal, Madhya Pradesh

[2] ABHINAV PRAKASH NAIR
[2] Research Scholar, Sri Satya Sai University of Technology and Medical Science, Sehore, Bhopal, Madhya Pradesh

ABSTRACT

We constructed models for over 660 genes, including 18000 ns SNPs, using an automated comparative modelling technique, and then compared and contrasted them with prediction methods based on structure and sequence. In addition to OMIM, which contains a minor allele frequency (MAF) of less than 0.01, another source of ns SNP data was the GWAS database. We have employed systems analyses by analyzing multi-omic studies of Coronary Atherosclerosis (CA) to decipher target pathways and underlying mechanisms. Our study showed deregulation in the homocysteine degradation pathway. Transcriptomics analysis showed differential regulation of genes involved in the utilization of homocysteine. However, integration of metabolomics data showed that Vitamin B_6 and B_{12} which act as cofactors of the crucial enzymes of the homocysteine pathway was deficient in CA patients.

KEYWORDS: SNP, Structural, prediction, mutation, GEO, metabolomics pathway analysis.

INTRODUCTION

*e*ISSN 1303-5150

Recent advances in genome-sequencing technology have shown that there are millions of genetic variances between individuals. Thanks to this breakthrough finding, it is now able to identify the genetic predisposition to and explain the origins of common ailments. Among the many frequent causal variations that will be mapped and found in the future are the 90% of human genetic diversity that is represented by single nucleotide polymorphisms (SNPs). Genotyped in the Hap Map project, which not only verifies SNPs and calculates their allele frequency in the general population, but evaluates the degree of linkage disequilibrium (LD) that occurs among them. Case-control studies can now type hundreds of thousands of SNPs in a large number of persons because to recent developments in SNP genotyping technology. In locations with high LD, bioinformatics estimates of SNPs' functional implications will assist drive functional research and narrow down the best candidate SNPs when more causal variants for common diseases are discovered.

We performed differential gene expression analysis of gene expression microarray dataset of cartilage tissue of 12 CA patients compared to that of 12 healthy control samples that were available in the GEO database (GSE id: GSE74089). Genes with adjusted p. value ≤ 0.05 were considered for further downstream analyses. Pathway enrichment analyses were carried out using the ClueGO plugin of Cytoscape (Bindea et al., 2009; Shannon et al., 2003). Reactome pathway database was queried to obtain significant pathways and to construct pathway subnetworks (Croft et al., 2014). Three of the significant pathways were used to construct subnetworks and understand the mechanism of action in detail. The pathway term ‘metabolism of amino acids and derivatives term p.value 0.013 showed 193 differentially regulated genes. This parent pathway was used to construct two pathway subnetworks namely ‘degradation of cysteine and homocysteine’ and choline metabolism. The next major pathway chosen was the Metabolism of water-soluble vitamins and cofactors with a term p. value of 0.025. This pathway

eISSN 1303-5150

term showed 113 differentially regulated gens. This pathway was used to further construct 3 pathway subnetworks namely folate metabolism, vitamin B_6 metabolism, and vitamin B_{12} metabolism. The phase II conjugation pathway having 70 differentially expressed genes had a p. value of 0.0001. The methylation pathway subnetwork was derived from the Phase II conjugation pathway. All the above pathway terms were categorized under the Metabolism pathway of the Reactome database in a hierarchical layout which helped us to browse and construct the pathway sub-networks. This helped us to further analyze the major pathways in detail by studying the gene expression pattern of the sub-network and deciphering the underlying mechanisms.

LITREATURE REVIEW

RESHMARAJ S, D.N. DAS (2021) Traditional SNP genotyping techniques may be used to analyse a limited number of SNPs. SNPs and economic characteristics or sickness may be linked using these technologies. High throughput SNP genotyping methods are used in genomic selection, MAS, and QTL mapping. To better evaluate an animal's estimated breeding value (EBV), genetic selection may find superior individuals before they reach sexual maturity (Schefers and Weigel, 2012). Allele or DNA sequence sets may be organised into clusters of linked SNPs in Heliotype mapping. An region of the genome with significant linkage disequilibrium is represented by this single-nucleotide polymorphism (SNP). Using tag SNPs may be useful for studies genotyping hundreds of thousands of SNPs over the whole genome in one sample. New genomic selection methods are used to identify SNPs in farm animals. Using genomic selection, breeding programmes may be made more efficient and less expensive by decreasing generation intervals while boosting selection intensity.

NICOLÒ MUSSO (2021) The SARS-CoV-2 virus has played a crucial role in the pandemic of COVID-19. The mutation rate of the spike (S) gene is high, resulting in a rapid spread of new variants. We provide a rapid, simple, configurable, and cost-effective approach for detecting all known

eISSN 1303-5150

differences in the minimum amount of time. With our method, identifying the B.1.160 lineage may be performed with only two conventional PCRs on the S gene, two Sanger sequencing procedures, and potentially another PCR/sequencing test on a N gene constituent. A 1312-bp region of the S gene that included a collection of SNPs was chosen to isolate all of the differences. Because certain mutations (e.g., N501Y is present in Alpha, Beta, and Gamma variants) are shared by several variants, it is possible to identify all variations without looking for all mutations associated to each other (e.g., A57 distinctive to the Alpha variant). Findings: A "weight" may be assigned to each of the probable mutations found in the S gene's designated area. Resilience was confirmed by testing 80 SARS-CoV-2 positive samples. There was no need for whole-genome sequencing in this investigation, and our method was proven to be more accurate than commercial kits.

HONGCHANG DONG (2018) Breast cancer (BRCA) affects around one in eight women, making it one of the most common cancers among women. Gene expression data for GSE103512 was gathered in an effort to discover genes that have a role in the development and occurrence of BRCA. There were 75 samples, 65 of which were malignant and 10 that weren't, in this investigation. We were able to find DEGs that were different between BRCA patients and healthy controls using the R programme. GO and KEGG pathway studies were carried out using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (DAVID). For the visualisation of protein-protein interactions of these DEGs, the search method for retrieving interacting genes (STRING) was also employed. As well as additional genes and medications connected with hub genes, CBioportal projected. A total of 357 DEGs were identified, 77 of which were up-regulated while the other 280 were down-regulated. GO terms and pathways linked with BRCA were discovered by analysing these DEGs. The PPI network and the degree of connectivity of each DEG's PPI network were used to screen for IGF1, EGFR, JUN, and Estrogen Receptor 1

eISSN 1303-5150

(ESR1) in DEGs. BRCA's therapeutic targets were finally narrowed down to IGF1 and ESR1. This bioinformatics study revealed that IGF1 genes may have a role in stomach cancer development. As new biomarkers for diagnosis and in the creation of combination therapy for BRCA, they might be used in the future

SHIJIE XIN (2018) Ulcerative colitis is one chronic inflammatory disease that is affected by the cytokine interleukin-6 (IL-6). Genomic DNA sequencing of Qinghai yellow chickens found single-nucleotide polymorphisms (SNPs) in the region spanning 2200 bp upstream to 500 bp downstream of the IL-6 gene (Gallus Gallus). CpG islands and transcription factors in IL-6's promoter region were predicted by computer programmers. The IL-6 genome has a total of 28 single nucleotide polymorphisms (SNPs). GeneBank does not name four of these 28 SNPs, three of which are in the 5′ regulatory region (357 (G > A), 447 (C > G), and 663 (A > G). Researchers detected 11 single nucleotide polymorphisms (SNPs), altering possible transcription factor binding locations, via bioinformatics research. In the 5′ regulatory region of the IL-6 gene, SNPs in the C-939G promoter mutation may influence CpG island number and methylation status. According to genetic diversity analyses, the newly discovered A-663G site deviated greatly from the Hardy-Weinberg equilibrium. Additional research into the function of the IL-6 gene promoter and the links between these SNPs and the resistance of chickens to intestinal inflammation may now be built on these results.

TUNE H. PERS, PASCAL TIMSHEL (2015) The determination of whether disease- or trait-associated SNPs enrich for specific biological annotations is a critical computational step after GWAS. Gene density and high linkage disequilibrium (LD) locations might have a negative impact on SNP-based enrichment analysis. This is because GWAS results tend to be clustered in these areas. SNP-based enrichment analyses may be calibrated using the matched sets of SNPs provided on the SNP snap Web site. In addition, SNP snap can rapidly and correctly detect which random loci are

eISSN 1303-5150

genetically related to those in a given query set based on allele frequency, the number of SNPs in the LD, the distance to the nearest gene, and gene density.

METHODOLOGY

It is common for databases and literature to include functional information, such as which residues catalyse processes, bind legends, or interact with other proteins. Almost 15,000 proteins are covered by the Catalytic Site Atlas, which is based on more than 700 published papers. Other computational methods have been developed for determining which residues are critical for function. In order to anticipate functional residues based on evolutionary conservation of sequences, the "evolutionary trace" technique is a commonly used computer tool. This method ranks residues according to their evolutionary importance using a phylogenetic analysis based on homologous sequences. This conservation may then be mapped using a representative structure. Connoisseurship in clusters-

Table 1 Co-factor dependent protein genes involved in CA

Uniprot	HGNC symbol	HGCN Description	Cofactor
Q8N5Z0	AADAT	aminoadipate aminotransferase [Source:HGNC Symbol;Acc:HGNC:17929]	Vit B6 (PP)
P21549	AGXT	alanine-glyoxylate aminotransferase [Source:HGNC Symbol;Acc:HGNC:341]	Vit B6 (PP)
Q9BYV1	AGXT2	alanine--glyoxylate aminotransferase 2 [Source:HGNC Symbol;Acc:HGNC:14412]	Vit B6 (PP)
P13196	ALAS1	5'-aminolevulinate synthase 1 [Source:HGNC Symbol;Acc:HGNC:396]	Vit B6 (PP)
P22557	ALAS2	5'-aminolevulinate synthase 2 [Source:HGNC Symbol;Acc:HGNC:397]	Vit B6 (PP)
Q96A70	AZIN2	antizyme inhibitor 2 [Source:HGNC Symbol;Acc:HGNC:29957]	Vit B6 (PP)
P54687	BCAT1	branched chain amino-acid transaminase 1, cytosolic [Source:HGNC Symbol;Acc:HGNC:976]	Vit B6 (PP)
O15382	BCAT2	branched chain amino-acid transaminase 2, mitochondrial [Source:HGNC Symbol;Acc:HGNC:977]	Vit B6 (PP)
P35520	CBS	cystathionine-beta-synthase [Source:HGNC Symbol;Acc:HGNC:1550]	Vit B6 (PP)
Q16773	CCBL1	cysteine conjugate-beta lyase, cytoplasmic [Source:HGNC Symbol;Acc:HGNC:1564]	Vit B6 (PP)
Q6YP21	CCBL2	cysteine conjugate-beta lyase 2 [Source:HGNC Symbol;Acc:HGNC:33238]	Vit B6 (PP)
Q9Y600	CSAD	cysteine sulfinic acid decarboxylase [Source:HGNC Symbol;Acc:HGNC:18966]	Vit B6 (PP)
P32929	CTH	cystathionine gamma-lyase [Source:HGNC Symbol;Acc:HGNC:2501]	Vit B6 (PP)
P20711	DDC	dopa decarboxylase (aromatic L-amino acid decarboxylase) [Source:HGNC Symbol;Acc:HGNC:2719]	Vit B6 (PP)
Q8TBG4	ETNPPL	ethanolamine-phosphate phospho-lyase [Source:HGNC Symbol;Acc:HGNC:14404]	Vit B6 (PP)
O95954	FTCD	formimidoyltransferase cyclodeaminase [Source:HGNC Symbol;Acc:HGNC:3974]	Vit B6 (PP)
Q99259	GAD1	glutamate decarboxylase 1 (brain, 67kDa) [Source:HGNC Symbol;Acc:HGNC:4092]	Vit B6 (PP)
Q05329	GAD2	glutamate decarboxylase 2 (pancreatic islets and brain, 65kDa) [Source:HGNC Symbol;Acc:HGNC:4093]	Vit B6 (PP)
Q6ZQY3	GADL1	glutamate decarboxylase-like 1 [Source:HGNC Symbol;Acc:HGNC:27949]	Vit B6 (PP)
O75600	GCAT	glycine C-acetyltransferase [Source:HGNC Symbol;Acc:HGNC:4188]	Vit B6 (PP)
P17174	GOT1	glutamic-oxaloacetic transaminase 1, soluble [Source:HGNC Symbol;Acc:HGNC:4432]	Vit B6 (PP)
Q8NHS2	GOT1L1	glutamic-oxaloacetic transaminase 1-like 1 [Source:HGNC Symbol;Acc:HGNC:28487]	Vit B6 (PP)
P00505	GOT2	glutamic-oxaloacetic transaminase 2, mitochondrial [Source:HGNC Symbol;Acc:HGNC:4433]	Vit B6 (PP)
P24298	GPT	glutamic-pyruvate transaminase (alanine aminotransferase) [Source:HGNC Symbol;Acc:HGNC:4552]	Vit B6 (PP)
Q8TD30	GPT2	glutamic pyruvate transaminase (alanine aminotransferase) 2 [Source:HGNC Symbol;Acc:HGNC:18062]	Vit B6 (PP)
P19113	HDC	histidine decarboxylase [Source:HGNC Symbol;Acc:HGNC:4855]	Vit B6 (PP)
Q16719	KYNU	kynureninase [Source:HGNC Symbol;Acc:HGNC:6469]	Vit B6 (PP)
Q96EN8	MOCOS	molybdenum cofactor sulfurase [Source:HGNC Symbol;Acc:HGNC:18234]	Vit B6 (PP)
P04181	OAT	ornithine aminotransferase [Source:HGNC Symbol;Acc:HGNC:8091]	Vit B6 (PP)
P11926	ODC1	ornithine decarboxylase 1 [Source:HGNC Symbol;Acc:HGNC:8109]	Vit B6 (PP)

eISSN 1303-5150

Q6P996	PDXDC1	pyridoxal-dependent decarboxylase domain containing 1 [Source:HGNC Symbol;Acc:HGNC:28995]	Vit B6(PP)
Q6P474	PDXDC2P	NA	Vit B6(PP)
Q8IUZ5	PHYKPL	5-phosphohydroxy-L-lysine phospho-lyase [Source:HGNC Symbol;Acc:HGNC:28249]	Vit B6(PP)
Q9UG56	PISD	phosphatidylserine decarboxylase [Source:HGNC Symbol;Acc:HGNC:8999]	Vit B6(PP)
O94903	PROSC	proline synthetase co-transcribed homolog (bacterial) [Source:HGNC Symbol;Acc:HGNC:9457]	Vit B6(PP)
Q9Y617	PSAT1	phosphoserine aminotransferase 1 [Source:HGNC Symbol;Acc:HGNC:19129]	Vit B6(PP)
P11216	PYGB	phosphorylase, glycogen; brain [Source:HGNC Symbol;Acc:HGNC:9723]	Vit B6(PP)
P06737	PYGL	phosphorylase, glycogen, liver [Source:HGNC Symbol;Acc:HGNC:9725]	Vit B6(PP)
P11217	PYGM	phosphorylase, glycogen, muscle [Source:HGNC Symbol;Acc:HGNC:9726]	Vit B6(PP)
P09455	RBP1	retinol binding protein 1, cellular [Source:HGNC Symbol;Acc:HGNC:9919]	Vit B6(PP)
P50120	RBP2	retinol binding protein 2, cellular [Source:HGNC Symbol;Acc:HGNC:9920]	Vit B6(PP)
Q96I15	SCLY	selenocysteine lyase [Source:HGNC Symbol;Acc:HGNC:18161]	Vit B6(PP)
P20132	SDS	serine dehydratase [Source:HGNC Symbol;Acc:HGNC:10691]	Vit B6(PP)
Q96GA7	SDSL	serine dehydratase-like [Source:HGNC Symbol;Acc:HGNC:30404]	Vit B6(PP)
Q9HD40	SEPSECS	Sep (O-phosphoserine) tRNA:Sec (selenocysteine) tRNA synthase [Source:HGNC Symbol;Acc:HGNC:30605]	Vit B6(PP)
O95470	SGPL1	sphingosine-1-phosphate lyase 1 [Source:HGNC Symbol;Acc:HGNC:10817]	Vit B6(PP)
P34896	SHMT1	serine hydroxymethyltransferase 1 (soluble) [Source:HGNC Symbol;Acc:HGNC:10850]	Vit B6(PP)
P34897	SHMT2	serine hydroxymethyltransferase 2 (mitochondrial) [Source:HGNC Symbol;Acc:HGNC:10852]	Vit B6(PP)
O15269	SPTLC1	serine palmitoyltransferase, long chain base subunit 1 [Source:HGNC Symbol;Acc:HGNC:11277]	Vit B6(PP)
O15270	SPTLC2	serine palmitoyltransferase, long chain base subunit 2 [Source:HGNC Symbol;Acc:HGNC:11278]	Vit B6(PP)
Q9NUV7	SPTLC3	serine palmitoyltransferase, long chain base subunit 3 [Source:HGNC Symbol;Acc:HGNC:16253]	Vit B6(PP)
Q9GZT4	SRR	serine racemase [Source:HGNC Symbol;Acc:HGNC:14398]	Vit B6(PP)
P17735	TAT	tyrosine aminotransferase [Source:HGNC Symbol;Acc:HGNC:11573]	Vit B6(PP)
Q8IYQ7	THNSL1	threonine synthase-like 1 (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:26160]	Vit B6(PP)
Q86YJ6	THNSL2	threonine synthase-like 2 (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:25602]	Vit B6(PP)
P15104	GLUL	glutamate-ammonia ligase [Source:HGNC Symbol;Acc:HGNC:4341]	Vit B6(PP)
P31939	ATIC	IMP cyclohydrolase [Source:HGNC Symbol;Acc:HGNC:794]	Vit B9(THF)
Q16526	CRY1	cryptochrome circadian clock 1 [Source:HGNC Symbol;Acc:HGNC:2384]	Vit B9(THF)
Q49AN0	CRY2	cryptochrome circadian clock 2 [Source:HGNC Symbol;Acc:HGNC:2385]	Vit B9(THF)
P15328	FOLR1	folate receptor 1 (adult) [Source:HGNC Symbol;Acc:HGNC:3791]	Vit B9(THF)
P14207	FOLR2	folate receptor 2 (fetal) [Source:HGNC Symbol;Acc:HGNC:3793]	Vit B9(THF)
P41439	FOLR3	folate receptor 3 (gamma) [Source:HGNC Symbol;Acc:HGNC:3795]	Vit B9(THF)
P22102	GART	phosphoribosylglycinamide formyltransferase,[Source:HGNC Symbol;Acc:HGNC:4163]	Vit B9(THF)
P27352	GIF	gastric intrinsic factor (vitamin B synthesis) [Source:HGNC Symbol;Acc:HGNC:4268]	Vit B12(AB12)
P53582	METAP1	methionyl aminopeptidase 1 [Source:HGNC Symbol;Acc:HGNC:15789]	Vit B12(AB12)
Q6JBZ8	METAP1D	methionyl aminopeptidase type 1D (mitochondrial) [Source:HGNC Symbol;Acc:HGNC:32583]	Vit B12(AB12)
P50579	METAP2	methionyl aminopeptidase 2 [Source:HGNC Symbol;Acc:HGNC:16672]	Vit B12(AB12)
Q99707	MTR	5-methyltetrahydrofolate-homocysteine methyltransferase [Source:HGNC Symbol;Acc:HGNC:7468]	Vit B12(AB12)
Q99707	MTR	5-methyltetrahydrofolate-homocysteine methyltransferase [Source:HGNC Symbol;Acc:HGNC:7468]	Vit B12(MeB12)
Q99707	MTR	5-methyltetrahydrofolate-homocysteine methyltransferase [Source:HGNC Symbol;Acc:HGNC:7468]	Vit B9(THF)
P22033	MUT	methylmalonyl CoA mutase [Source:HGNC Symbol;Acc:HGNC:7526]	Vit B12(AB12)
P20061	TCN1	transcobalamin I (vitamin B12 binding protein, R binder family) [Source:HGNC Symbol;Acc:HGNC:11652]	Vit B12(AB12)
P20062	TCN2	transcobalamin II [Source:HGNC Symbol;Acc:HGNC:11653]	Vit B12(AB12)
P80404	ABAT	4-aminobutyrate aminotransferase [Source:HGNC Symbol;Acc:HGNC:23]	Vit B6(PP)
Q9Y697	NFS1	NFS1 cysteine desulfurase [Source:HGNC Symbol;Acc:HGNC:15910]	Vit B6(PP)
P23378	GLDC	glycine dehydrogenase (decarboxylating) [Source:HGNC Symbol;Acc:HGNC:4313]	Vit B6(PP)

Serving residues in three dimensions reveals functional surfaces, such as those involved in protein-protein interactions, which may be seen. Clusters of residues with better sequence conservation than expected by their overall sequence identification were identified using the 3DCA method. In terms of amino acid residue conservation, however, it has been discovered to be highly dependent on the protein structure in which they are present. Methods that use conservation of sequence cannot tell between evolutionary constraints resulting from the need to preserve function from those arising from the conservation of the structural context. These techniques tend to concentrate on the protein's surface rather than the protein's core because of structural issues. This method is particularly problematic for enzyme catalytic residues, which are often inaccessible to solution.

Gene Expression Omnibus (GEO) is a public repository of high throughput gene expression data like microarray, RNA sequencing, Chromatin Immuno Precipitation – sequencing or (ChIP-seq) (Edgar et al., 2002). Researchers perform high-throughput experiments such as gene expression analyses for mRNA, miRNA, circRNA, etc., This can be used to determine differential gene expression and post-transcriptional regulations by miRNAs (Kaucsár et al., 2010). Next-generation

eISSN 1303-5150

sequencing techniques such as RNA sequencing are performed, which can be used for differential gene expression analyses, alternate splicing analyses, differential transposable analysis, and so on (Bird, 2012). ChIP – seq analysis is used to understand how proteins interact with DNA. This finds application in determining transcription factors regulating gene expression (Nakato & Sakata, 2020). ChIP – seq can also be used in studying epigenetics by immunoprecipitation of methylated or acetylated histones and sequencing the bound DNA fragment (Nakato & Sakata, 2020). Such high throughput datasets can find various applications in different downstream analyses and are hence valuable for the research community

DATA ANALYSIS

Genome-wide prediction of protein structure

The databases included in Table 1 give information on the variations in the human genome sequence. OMIM included 2,249 ns SNPs in 500 genes, whereas dbSNP125 contained 21,471 ns SNPs in 5,500 different genes. To put it simply, mutations in the buried residues of a protein are more likely to have a detrimental effect on its structure, which may lead to disease. When two or more proteins come into contact, the residues on the protein's surface are more likely to come into play. It is difficult to make general generalisations regarding gene structure since only a small number of experimental crystal structures have been found. To better understand the structure of these genes, we used homology modelling to build an automated, large-scale, structure prediction approach. This technique, which requires a prior prediction of the protein's fold, is sometimes referred to as "homology" or "fold recognition."

All of the genes were homologously recognised using FUGUE. The list of genes has an average of 400 residues, with more than 95% of the sequences having a length of less than 1,000 amino acids. Nearly 8,800 amino acids make up the SYNE1 gene (nuclear envelope spectrum repeat protein). Fewer than a quarter of the genes had no significant FUGUE predictions, with an average of

eISSN 1303-5150

40 HOM-STRAD [40] families predicted per gene by FUGUE (Z-score 7). Genes belonging to a single HOMSTRAD family were expected to account for almost half of the total number of genes. Because several HOMSTRAD families represent many structural domains, this does not reflect a single structural domain inside the genes. Genes with two or more HOMSTRAD families and hence many domains were predicted to contain more than 1,100 genes whereas genes with three families were found to have just 551. Most sequence alignments produced by FUGUE are less than 500 residues long.

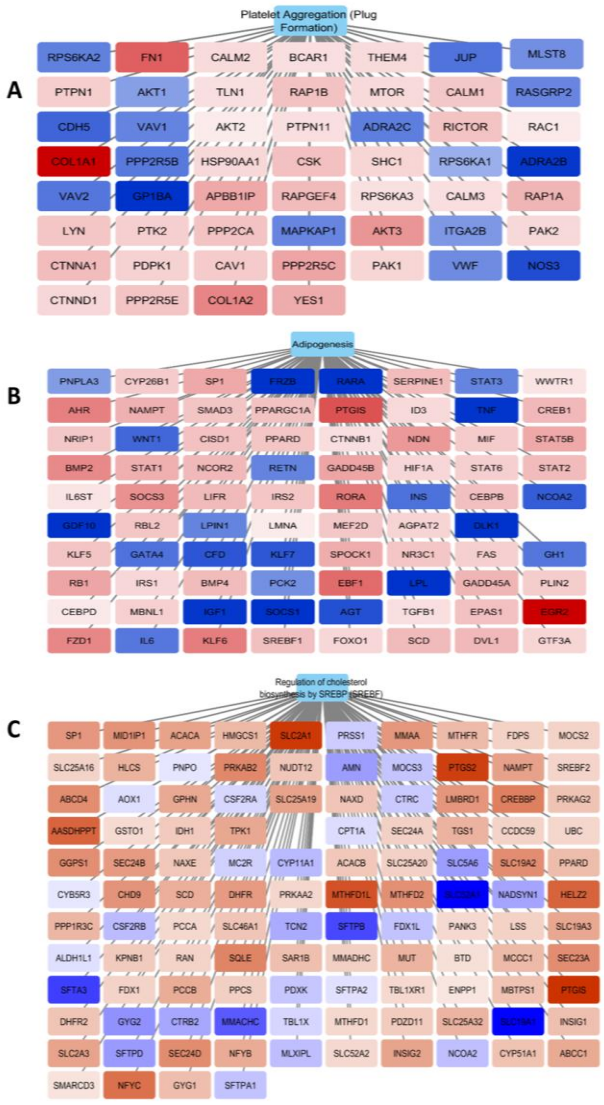

Fig1: CA Pathway enrichment subnetworks showing differentially regulated genes belonging to A) Platelet aggregation B) Adipogenesis C) Regulation of cholesterol biosynthesis (P.Value ≤0.05) (Red– Upregulated; Blue-Downregulated)

Homology recognition of protein sequences

*e*ISSN 1303-5150

For 4,700 genes, the Model-ler tool created more than 16,000 models with high FUGUE scores. Of the genes, 37% had multiple FUGUE scores, suggesting that the gene contains many structural domains. A gene model for each structural domain was generated, but no attempt was made to merge them into a single model. Structure predictability and the amount of accommodated sequence variations varied among datasets. Of the mutations that were shown to be associated with OMIM, more than 62% of the ns SNPs (375 genes) occurred in an area of the gene that had a predicted structure (11009 ns SNPs in 3,956 genes).

The cluepedia plugin of Cytoscape was used to construct subnetworks of these target pathways. The subnetwork showed the pathway term node connected to the participating genes enriched in the pathway. Our transcriptomics analyses also established the roles of cofactors - Vitamin B_6, B_{12}, and heme in CA. Therefore, we integrated the cofactor-dependent – gene network with our target pathways to study the roles of these cofactors in our target pathways.

The individual subnetworks were combined using the merge tool of Cytoscape. This gave us the integrated network model that could represent the pathophysiologic derangement of CA. To determine the key genes in the network, we performed the network centrality analysis using the Cytoscape app CytoNCA (Tang et al., 2015). Using this tool, we obtained the degree of each node present in the network. Next, we set the node size corresponding to its degree. Nodes higher than degree 2 i.e., nodes that belong to at least two subnetworks were colored red and had a larger size.

The integrated network model showed the interdependence of the pathway subnetworks and the cofactor gene networks. The Cholesterol metabolism pathway was found to be dependent on vitamin B_6 and B_{12} as cofactors. The heme cofactor network had interactions with cholesterol metabolism, steroid metabolism, and drug metabolism. This showed the role of heme cofactors in these pathways. However, the heme metabolism pathway itself had interactions with the B_6 cofactor network and was found to be dependent

eISSN 1303-5150

on it. The first and the rate-limiting step in heme biosynthesis involves the ALAS2 gene which requires vitamin B_6 as a cofactor (Ajioka et al., 2006). Further, the network centrality analysis showed that two of the nodes had the highest degree among all nodes. CYP3A4 and CYP11A1 were the two top hub nodes with a degree of 3 each. The CYP3A4 hub node was found to be involved in the very first step of activation of vitamin D to 1α25(OH)2D3 (Z. Wang, Schuetz, et al., 2013, p. 4). CYP3A4 was found to be downregulated in CA patients. The CYP11A1 gene is involved in the first step of conversion of cholesterol during steroid hormone biosynthesis (Slominski et al., 2015, p. 1). CYP11A1 gene was also found to be downregulated in CA. In summary, the integrated network model illustrated the underlying interdependence of the target pathways and the cofactor–gene network. This demonstrated that the critical pathways deranged in CA could be regulated by the efficient functioning of the cofactors. Moreover, the integrated network model also provided the critical hub nodes that could be potential targets in the management of CA.

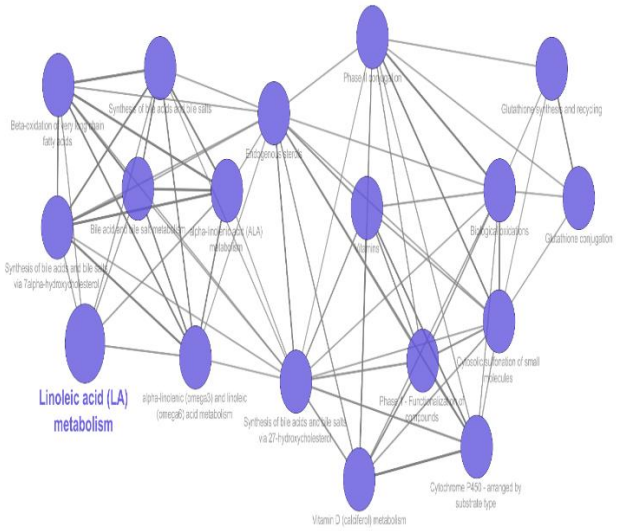

Fig 2: The REACTOME cluster showing CYP450 enzymes that are involved in vit d and cholesterol metabolism

Prediction of deleterious mutations

There are a variety of methods for estimating the functional relevance of ns SNPs, but the sheer number of ns SNPs makes them impractical for genome-wide studies. Only a very small percentage of the projections that have been made public line up with what we have seen in our data. In light of the fact that many of our datasets can be downloaded with pre-calculated predictions, we've only looked at

eISSN 1303-5150

methods we can apply ourselves. Table 2 summarises the various approaches and the total number of ns SNPs used in each (see below).

For 52% of our db SNP dataset, we discovered LS-SNP predictions by cross-referencing with the pre-computed LS-SNP predictions Ten percent of all genes are expected to be altered, either in structure or function (2,291 ns SNPs in 1,511 genes). Only 21% of db SNPs were polyphonic, according to the predictions (4,459 ns SNPs in 2,237 genes). Only 5% of the 1,141 ns SNPs in 845 genes analysed were predicted to be detrimental, or roughly a fourth of the total. LS-SNP and Polyphone SNPs in 367 genes account for less than 3% of the total (549 ns SNPs). Only 1% of the db SNPs projected to be hazardous by both approaches are shown to be dangerous (225 ns SNPs in 203 genes). The OMIM database does not provide any information on the LS-SNP or Polyphen sequence variations.

Environment substitution scores

Numerous methods have been proposed to assess the chance of discovering a certain amino acid alteration in a specific structural context. There are several ways to describe the structure of an amino acid, including the geometry of its backbone (helix, strand, coil), the presence of hydrogen bonds with other amino acids, and even the presence of a tale. Tables calculating the log-odds chance of an amino acid substitution in a certain structural environment have already been produced for a large number of structural contexts. Data on the likelihood that a sequence mutation is identified in other homologous sequences is provided in these tables. If we have a model of the protein structure, we may calculate the environmental score (envy score) for each mutation in the datasets. In this case, the mutation presented cannot be tolerated without major structural and environmental changes to the protein, and hence may be considered harmful. Negative scores show that the environment not only tolerates but actively encourages the mutation.

eISSN 1303-5150

Vitamin B_6, B_{12}, and heme cofactor dependent genes network and pathway analysis:

The Vitamin B_6, B_{12}, and heme cofactors were chosen for our downstream analyses based on our Transcriptomics and pathway clustering data. Respective genes associated with the cofactors were obtained from a published study that discussed the role of cofactors in the management of diseases (Scott-Boyer et al., 2016). A subset of cofactors was taken from this study. Next, we performed pathway enrichment analysis using ClueGO. Interestingly the Reactome analysis also showed that the enriched pathways were clustered into the Vitamin D metabolism pathway **(**Error! Reference source not found.**)**. T he Vitamin D metabolic pathway involved enrichment of Cytochrome P450 genes. Our analysis showed that in addition to the xenobiotic metabolism where the external drugs and supplements are assimilated and cleared from the system, the cytochrome p450 enzymes are essential for the endobiotic metabolism of essential hormones in the system.

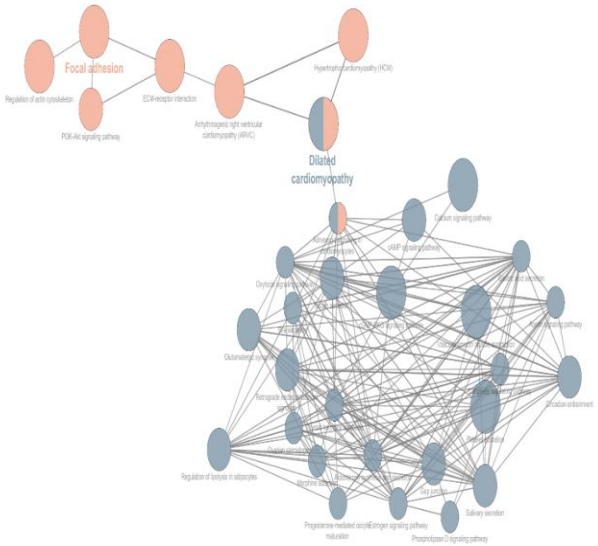

Fig 3: KEGG pathway cluster analysis showing various cross-interaction in Dilated Cardiomyopathy

Prediction of structural stability

A protein structure's stability may be measured utilising two methods in this research. Since a protein structural model is required for these techniques, they are seldom used in genome-wide ns SNP investigations. According to SDM, the majority of OMIM variations that cause illness have G scores of less than 3. In comparison to the db SNP mutations that cause illness (11 percent), this is nearly twice as many (11 percent) (6 percent). The Xanthine dehydrogenates (XDH) gene's R149C sequence alteration in the OMIM has

*e*ISSN 1303-5150

the lowest projected SDM energy score (-12.4). With I mutant's forecast being the gloomiest, SIFT and envy score (-9) all anticipate this to be bad (-1.05). This mutation is linked to Xanthinuria type I due to XDH deficiency. The P1150R gene mutation has an Envy score (-3) and an SDM, both of which indicate that it is structurally harmful (-5.9).

Prediction of functional residues

Examples of functional residues include protein-protein or protein ligand interactions and enzyme catalysis. These interactions may be deduced if the structures of homologous complexes are accessible. LS-predictions SNPs contain these annotations, as well. Since this information is often sparse, it might be difficult to apply it to distantly related homologues. Because of their simplicity, this study relies on monomer structures to build its models. It is possible that certain amino acids will be involved in interfaces or binding to ligands or other molecular entities in dimmer or higher order oligomers. The amino acid sequences involved in protein interfaces will be examined in a second study using homologous protein complexes (Bickerton, unpublished data). Crescendo algorithm was utilised in this work to discover residues that may have a role in the function or binding of other molecules based on their sequence conservation score.

CONCLUSION

A rise in disease or distinctive susceptibility due to some of the potentially hazardous ns SNPs indicated by the database may not be sufficient to cause sickness. A number of other mutagenic or environmental factors must be present before a disease is detectable. An in-depth analysis of the impacts and interactions of protein structural sequence variations is needed to understand the significance of sequence variants in multi-factorial diseases such as cancer, type 1 diabetes, and heart disease. The approaches discussed here may be easily applied to new variants since they have been connected to common disorders. Overall the work carried out in the present thesis helps to understand CA at the systems level. It helps to elucidate the underlying molecular signatures of

eISSN 1303-5150

the disease, potential biomarkers, and therapeutic targets. Our study has addressed various pathophysiologic implications such as elevated homocysteine, defective steroid metabolism, and iron metabolism in CA. Through our study, we discovered that the vitamin and cofactors metabolism was commonly involved in the different chapters of our study. hence the vitamin metabolism could act as a potential therapeutic target spot. our study showed that CBS, MUT, CYP3A4, CYP11A1 could be potential therapeutic targets, from our network model studies. In addition to these genes, the transporters of vitamin B12 namely AMN, TCN2, and the vitamin B6 activating enzyme PNPO have to be overexpressed. CYP 450 enzymes which have both endobiotic and xenobiotic implications in CA can be rectified by improving vitamin B6 mediated heme biosynthesis.

ACKNOWLEDGEMENTS:

We acknowledge the administration and Department of Biotechnology, Sri Satya Sai university of Technology and Medical Sciences, Sehore, Bhopal for the support and infrastructure they have rendered in this project.

CONFLICT OF INTEREST STATEMENT: The authors declare no conflict of interest.

AUTHORSHIP CONTRIBUTION STATEMENT:

Abhinav Nair: Investigation, sample collection, data interpretation, writing original draft.

*Sri Satya Sai University of Technology and Medical Sciences, Sehore, Bhopal

Shobha Malvia: Conceptualization, methodology, supervision, data interpretation

*Sri Satya Sai University of Technology and Medical Sciences, Sehore, Bhopal

REFERENCE

1. ReshmaRaj S , D.N. Das (2021),” Molecular markers and its application in animal breeding,” Advances in Animal Genomics, 2021

*e*ISSN 1303-5150

2. Nicolò Musso (2021),” Discriminatory Weight of SNPs in Spike SARS-CoV-2 Variants: A Technically Rapid, Unambiguous, and Bioinformatically Validated Laboratory Approach,” Viruses 2022, 14, 123. https://doi.org/10.3390/v14010123

3. Hongchang Dong(2018) Bioinformatic analysis of differential expression and core GENEs in breast cancer,” March 2018International Journal of Clinical and Experimental Pathology

4. Shijie Xin (2018),” Bioinformatics Analysis of SNPs in IL-6 Gene Promoter of Jinghai Yellow Chickens,” Published: 6 September 2018

5. Tune H. Pers, Pascal Timshel (2015),” SNPsnap: a Web-based tool for identification and annotation of matched SNPs,” Bioinformatics, Volume 31, Issue 3, 1 February 2015,

6. Abby et al., 2016 S.S. Abby, J. Cury, J. Guglielmini, B. Néron, M. Touchon, E.P. Rocha Identification of protein secretion systems in bacterial genomes Sci. Rep., 6 (2016), p. 23080

7. Adams et al., 2016 M.D. Adams, B. Bishop, M.S. Wright Quantitative assessment of insertion sequence impact on bacterial genome architecture Microbial. Genomics, 2 (2016), Article e000062

8. Lind C, Kilian A, Benzie J. Development of diversity arrays technology markers as a tool for rapid genomic assessment in Nile tilapia. Anim Genet. 2017;48(3):362–4.

9. Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A. Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for

*e*ISSN 1303-5150

molecular breeding of Eucalyptus. BMC Proc. 2011;5(Suppl 7):54–P54.

10. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, et al. Diversity arrays technology: a generic genome profiling technology on open platforms. Methods Mol Biol. 2012;888:67–89.

11. Orsini L, Mergeay J, Vanoverbeke J, De Meester L. The role of selection in driving landscape genomic structure of the waterflea Daphnia magna. Mol Ecol. 2013;22(3):583–601.

12. Tsumura Y, Uchiyama K, Moriguchi Y, Ueno S, Ihara-Ujino T. Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, Cryptomeria japonica. Heredity. 2012;109(6):349–60.

13. Carda, C., Silvestrini, G., de Ferraris, M. G., Peydro, A., & Bonucci, E. (2005). Osteoprotegerin (OPG) and RANKL expression and distribution in developing human craniomandibular joint. *Tissue and Cell*, *37*(3), 247–255.

14. Carmel, R., Lau, K.-H. W., Baylink, D. J., Saxena, S., & Singer, F. R. (1988). Cobalamin and osteoblast-specific proteins [Article]. *New England Journal of Medicine*, *319*(2), 70–75.

15. Catto, M. (1976). Pathology of aseptic bone necrosis. *Aseptic Necrosis of Bone*, 3–100.

16. Catto, M. (1977). Ischaemia of bone. *Journal of Clinical Pathology. Supplement (Royal College of Pathologists).*, *11*, 78.

17. Ceelie, H., Spaargaren-van Riel, C. C., Bertina, R. M., & Vos, H. L. (2004). G20210A is a functional mutation in the prothrombin gene; effect on protein levels and 3′-end formation. *Journal of Thrombosis and Haemostasis*, *2*(1), 119–127.

18. Chai, W., Zhang, Z., Ni, M., Geng, P., Lian, Z., Zhang, G.,

*e*ISSN 1303-5150

Shi, L. L., & Chen, J. (2015). Genetic association between methylenetetrahydrofolate reductase gene polymorphism and risk of osteonecrosis of the femoral head. *BioMed Research International*, *2015*.

19. Chang, C., Greenspan, A., & Gershwin, M. E. (2020). The pathogenesis, diagnosis and clinical manifestations of steroid-induced osteonecrosis. *Journal of Autoimmunity*, *110*, 102460. https://doi.org/10.1016/j.jaut.2020.102460

***e*ISSN** 1303-5150