Variant Identification for CYP17A1 Gene Involved in Hypertension

Chaitali Dhande* and Preenon Bagchi

Institute of Bioscience and Technology, MGM University, India
*Corresponding Author Email: chaitalidhande08@gmail.com

Abstract

Hypertension, characterized by elevated blood pressure, is a major risk factor for cardiovascular diseases and affects millions worldwide. The CYP17A1 gene, encoding a key enzyme involved in steroid hormone biosynthesis, has been implicated in blood pressure regulation. However, the specific genetic variants within CYP17A1 that contribute to hypertension remain incompletely understood. This study aimed to address this gap by conducting exome sequencing analysis on a family trio consisting of an affected male child with hypertension and his consanguineous, unaffected parents. By identifying and analyzing genetic variations in the CYP17A1 gene, we aimed to shed light on the genetic mechanisms underlying hypertension susceptibility and potentially uncover novel therapeutic targets for this condition. This work relies not only on the successful detection of variants in the patient's genome, but also on variant comparison between the patient and chosen relatives in order to identify the causative variants for the CYP17A1 gene that underline any particular blood pressure as a genetic complication. Genome sequences of the patient and their parents together, or family trio data, are most frequently utilized for this kind of analysis. This type of multisampling data makes it feasible to identify potentially causal de-novo mutations or loss-of-heterozygosity events, as well as to look for variants adhering to any Mendelian inheritance scheme that is consistent with the disease's known inheritance pattern. The variant rate in the genome are detected 1 variant every 26,992 bases. As per our research, both parents are unaffected; the variant cannot be dominant and inherited. The results suggest a complex interplay of genetic factors contributing to blood pressure regulation.

Keywords**:** *Hypertension, CYP17A1gene, exome sequencing, genetic variants, SNP, MNP, indels, missense variants, silent variants, blood pressure regulation*

1. Introduction

Over 7.5 million fatalities globally are related to high blood pressure or hypertension (World Health Organization, 2023). Hypertension, a multifactorial condition characterized by elevated blood pressure, remains a significant global health concern due to its high prevalence and association with cardiovascular morbidity and mortality (Mills et al., 2016). High blood pressure is the third most significant risk factor for cardiovascular disease (Lim et al., 2012). It is an increased risk factor for both ischemic and hemorrhagic strokes. In the case of hypertension, the rule of halves states that: (1) half of the population with high blood pressure (BP) is unknown; (2) half of those who are diagnosed are not receiving treatment; and (3) the other half of the treated population is not under control (Hooker, Cowap, Newson, & Freeman, 1999).According to estimates from the Global Burden of Diseases 2019 study, hypertensive heart disease, taking into account individuals of all ages and genders, contributes 0.85% of all Disability Adjusted Life Years (DALYs) worldwide and has been trending upward (GBD Compare, 2023). As compared to men, women are less likely to have hypertension (Everett, Zajacova, 2015). Reductions in morbidity and death are linked to antihypertensive medication and lifestyle modifications. Nonetheless, recent demographic studies show that control rates are typically low. New paths towards preventing hypertension and its consequences may become apparent with a deeper understanding of the genetic and environmental factors influencing blood pressure fluctuation. While environmental factors such as diet and lifestyle contribute to hypertension, genetic predisposition plays a crucial role in its etiology (Padmanabhan, Caulfield, & Dominiczak, 2015). The identification of genetic variants associated with hypertension has become a focus of research, aiming to elucidate the underlying mechanisms and develop targeted therapeutic interventions.

Adolescents who experience high blood pressure in response to physical, emotional, and physiological stress are at a higher risk of developing hypertension in adulthood. One of the well-known gene loci for adult hypertension is CYP17A1 (Van Woudenberg et al., 2015). Among the genes implicated in hypertension pathogenesis, Cytochrome P450 17A1 (CYP17A1) has emerged as a key player. CYP17A1 encodes the enzyme 17α-hydroxylase/17,20-lyase, which is involved in the biosynthesis of steroid hormones, including cortisol and dehydroepiandrosterone (DHEA) (Auchus, 2017). The first enzymatic activity is essential to the steroidogenic pathway, which generates glucocorticoids, which regulate the body's reaction to stress, and mineralocorticoids, which influence the kidneys' ability to handle water and sodium (Guyton, 1991). The biosynthesis of female and male sex hormones is a function of the second enzymatic action (Miller, & Auchus, 2011). Dysregulation of steroid hormone production mediated by CYP17A1 has been linked to various metabolic disorders, including hypertension (Miller et al., 2013). The 8673 bp CYP17A1 gene, which has eight exons and seven introns and is expressed in both the adrenal gland and the gonads, is located on chromosome 10q24.32 (Picado-Leonard & Miller, 1987). As a result, CYP17A1 may have a sex-specific effect on BP reactivity. Several studies have highlighted the association between genetic variations in the CYP17A1 gene and hypertension susceptibility. For example, single nucleotide polymorphisms (SNPs) in the CYP17A1 gene have been implicated in alterations of steroid hormone levels and blood pressure regulation (Xing et al., 2016). Additionally, genome-wide association studies (GWAS) have identified CYP17A1 as a candidate gene for hypertension susceptibility, further emphasizing its potential role in the pathogenesis of this condition (Levy et al., 2009). Genome-wide association studies (GWAS) have analyzed this gene's polymorphisms linked to hypertension recently (Newton-Cheh et al., 2009).

Moreover, genetic variants in other genes involved in the renin-angiotensin-aldosterone system (RAAS), such as angiotensinogen (AGT), angiotensin-converting enzyme (ACE), and angiotensin II type 1 receptor (AGTR1), have also been associated with hypertension (Wu et al., 2013). These genes modulate key pathways regulating blood pressure and fluid-electrolyte balance, highlighting the intricate genetic architecture underlying hypertension pathophysiology. Understanding the genetic determinants of hypertension, including variants in the CYP17A1 gene and related genes, is crucial for personalized risk assessment, early detection, and targeted intervention strategies. Therefore, this study aims to comprehensively analyze genetic variants in the CYP17A1 gene and explore their association with hypertension, shedding light on the intricate interplay between genetic factors and hypertension susceptibility.

2. Objectives

This study aimed to identify genetic variants associated with hypertension by analysing exome sequencing data from a family trio, focusing on the CYP17A1 gene. This was achieved through a comprehensive analysis pipeline, including quality checking, trimming, read mapping, variant calling, and post-processing. The rationale behind this approach was to elucidate the genetic basis of hypertension, particularly in cases where consanguinity is present, thereby providing insights into potential therapeutic targets and personalized treatment strategies.

1. 3. Materials and Methods

We went to analyze exome sequencing, which is a capture-based technique that focuses on and sequences the exome, or coding sections of the genome, data from a family trio in which the male child has hypertension while both parents, who happen to be consanguineous, are not. Our goal is to identify the genetic variation that is responsible for hypertension. Conducted a full analysis starting from the original sequenced reads in fastq format from SRA.

3.1 Quality Checking and Statistical Analysis

Quality checking is done using the FastQC tool. As per the FastQC results, we find that our sequence quality is not good. Hence, we went for the trimming of the sequence.

3.2 Trimming

Trimming can be used to illuminate both single-ended and paired-end data by performing a wide range of helpful trimming activities. Trimmometric software is used for performing trimming in this work.

3.3 Quality Checking

The goal of this stage is to find any potential problems with the input data of raw sequenced reads before moving on to the "real" analysis processes. By preparing impacted sequencing reads before attempting to map them to the reference genome, some of the common issues with NGS data can be reduced. The sequencer uses sequencing to identify the nucleotide bases in a DNA or RNA sample (library). A sequence again, just a series of nucleotides is created for every fragment in the library and is referred to as a read. Numerous sequence reads can be produced in a single experiment with modern sequencing technology. However, no sequencing technology is flawless, and every device will produce a unique set and quantity of mistakes, including calling erroneous nucleotides. The technical constraints of each sequencing platform are the cause of these incorrectly named bases. Error types that could affect how downstream analysis is interpreted must thus be understood, recognized, and eliminated. For this reason, sequence quality control is a crucial first step. Time is saved later when faults are detected early. That's why we went for quality checking, and our sequence quality is good.

3.4 Mapping Reads

A collection of sequences without genetic context is the result of sequencing. Which region of the genome the sequences relate to is unknown. An essential step in the study of contemporary genomic data is mapping an experiment's reads to a reference genome. Through mapping, readings are mapped to precise locations within the genome, providing information on things like gene expression levels. Since the short reads lack position information, we are unable to determine which region of the genome they originated from. To locate the matching section in the reference sequence, we must utilize the read's sequence. Finding a matching region can be difficult because the reference sequence is relatively large (around 3 billion bases for humans). Given the short length of our reads, they might have been read from many, equally plausible locations in the reference sequence. This is particularly valid for areas that repeat. For performing mapping of read "Map with BWAMEM" tool is used in this work.

3.5 Mapped read postprocessing

The variant calling software used in the next step determines the best set of postprocessing actions needed. Filter all sample paired-end reads such that only read pairs with properly mapped forward and reverse reads to the reference are kept. When these read pairings occur, we may be even more certain that they are neither sequencing artifacts nor non-human contaminating DNA. We used the "Samtools View" tool for filtering. In certain cases, incorrect genotype assignments at variant sites can result from duplicate reads, which are usually caused by PCR-overamplification of genomic fragments during sequencing library preparation (for instance, if a sample is heterozygous for a variant but fragments with one of the two alleles get amplified more efficiently than the others). "RmDup" is software used for removing duplicate reads. Now, sequences were ready for Variant analysis.

3.6 Variant Calling

After mapping and postprocessing all of the sequenced reads, we began searching for indications of sequence variations, or variants, between the reference genome and the sequenced genomic samples. To call the variants, we utilised "FreeBayes". Specifically, SNPs (single-nucleotide polymorphisms), indels (insertions and deletions), MNPs (multi-nucleotide polymorphisms), and complex events (composite insertion and substitution events) smaller than the length of a short-read sequencing alignment are the kinds of small polymorphisms that can be found using the Bayesian genetic variant detector FreeBayes.

3.7 Post-processing FreeBayes calls

We had to post-process the VCF dataset to address a few inconsistencies between Freebayes and downstream analytic tools before we could begin analysing the discovered variants. Specifically, we desired the division of multiallelic variant records into distinct record lines, which is necessary when more than one alternative allele is listed at a specific genomic location. This enables us to add information regarding the effects of a particular variant allele to each record.

3.8 Variant annotation and reporting

The identification of variations in a collection of samples is merely the beginning; without additional data and tools, it is nearly impossible to extract information from the list that is biologically or clinically significant. Specifically, the variations included in the list must

(1) Arranged in order of possible significance for the biological or clinical characteristics under investigation. As many variants bring silent changes or lie in intronic areas still covered by the exome-enriched sequencing data, only a small percentage of discovered variants, even with exome sequencing, will clearly affect a protein's function. Many of these have been seen in healthy people previously, which disproves their significant contribution to a negative phenotype.

(2) Screened according to the expected pattern of inheritance for a causal variant. Every sample's most likely genotype at each variation location is documented in a multisample VCF file. By identifying the people (samples) impacted by a particular phenotype, we can exclude variations whose inheritance patterns conflict with the phenotypic's observed inheritance pattern.

(3) Presented in a format that is more human friendly. Even though all pertinent information about every variant can be encoded in the VCF format, people find it difficult to interpret this data.

The GEMINI framework is used to annotate and report variations along with the genes they impact. Extensive annotation data for human variations from various sources is included with GEMINI. Without the need for extra downloads or format conversions, they can be used to easily annotate any list of human variants. Protocols for new methods should be included, but well-established protocols may simply be referenced.

2. 4. Results and Discussion

1. 4.1 Result

A genomic analysis of the CYP17A1 gene using SnpEff on the hg19 genome version:

Genome Information:

Genome Version: hg19

Date of Analysis: 2023-12-20 04:38

SnpEff Version: 4.3t (build 2017-11-24 10:18), by Pablo Cingolani Analysis Details:

Command Line Arguments: The SnpEff command is used for analysis, including input and output specifications, as well as statistics file location. SnpEff -i vcf -o vcf -stats /mnt/tmp/job_working_directory/007/825/7825566/outputs/dataset_df2b91f5-6b36-49d3-bb6e-64742049730f.dat hg19 /mnt/user-data-volC/data10/3/2/9/dataset_3290437d-20c941fe-a762-cced661046d3.dat Analysis Results:

Warnings: 13,211

Errors: 0

Number of Lines (Input File): 115,730

Number of Variants (Before Filter): 115,733

Number of Non-Variants (Reference equals Alternative): 0

Number of Variants Processed (After Filter and Non-Variants): 115,732

Number of Known Variants (Non-empty ID): 0 (0%)

Number of Multi-allelic VCF Entries (More than two alleles): 0

Number of Effects: 372,537

Genome Statistics:

Genome Total Length: 3,137,161,265 bases
Genome Effective Length: 3,123,882,938 bases
Variant Rate: 1 variant every 26,992 bases

This information provides insights into the quality and characteristics of the genomic data analyzed, including the number of variants, warnings, and errors encountered during the analysis, as well as statistics on the genome itself.

Table 1: Variant rate for the CYP17A1 gene

3. Variants rate details

Chromosome	Length	Variants	Variants rate
1	249,250,621	10,542	23,643
2	243,199,373	9,366	25,966
3	198,022,430	7,986	24,796
4	191,154,276	6,100	31,336
5	180,915,260	6,683	27,070
6	171,115,067	5,577	30,682
7	159,138,663	6,251	25,458
8	146,364,022	5,171	28,304
9	141,213,431	4,284	32,962
10	135,534,747	6,122	22,138
11	135,006,516	7,595	17,775
12	133,851,895	6,300	21,246
13	115,169,878	2,765	41,652
14	107,349,540	3,985	26,938
15	102,531,392	3,664	27,983
16	90,354,753	3,033	29,790
17	81,195,210	4,909	16,540
18	78,077,248	2,046	38,160
19	59,128,983	3,665	16,133
20	63,025,520	2,428	25,957
21	48,129,895	1,223	39,353
22	51,304,566	1,879	27,304
17_ctg5_hap1	1,680,828	1	1,680,828
19_gl000208_random	92,689	1	92,689
1_gl000192_random	547,496	2	273,748
4_gl000194_random	191,469	5	38,293
6_cox_hap2	4,795,371	6	799,228
6_mann_hap4	4,683,263	1	4,683,263
6_mcf_hap5	4,833,398	9	537,044
6_qbl_hap6	4,611,984	1	4,611,984
6_ssto_hap7	4,928,567	1	4,928,567
7_gl000195_random	182,896	49	3,732
9_gl000199_random	169,874	2	84,937
M	16,571	443	37
Un_gl000211	166,566	3	55,522
Un_gl000212	186,858	1	186,858
Un_gl000213	164,239	2	82,119
Un_gl000218	161,147	1	161,147
Un_gl000219	179,198	4	44,799
Un_gl000220	161,802	46	3,517
Un_gl000224	179,693	3	59,897

Chromosome	Length	Variants	Variants rate
Un_gl000225	211,173	6	35,195
Un_gl000229	19,913	95	209
Un_gl000234	40,531	2	20,265
X	155,270,560	3,199	48,537
Y	59,373,566	275	215,903
Total	**3,123,882,938**	**115,732**	**26,992**

Table 1 provides details about the variant rates for different chromosomes in the genomic analysis. Chromosomes: Information is provided for each regular chromosome (1-22, X, Y) and non-standard sequences.

Length: The length of each chromosome in base pairs.

Variants: The number of variants detected on each chromosome.

Variant Rate: The rate of variants per base pair for each chromosome.

For example, Chromosome 1 has a length of 249,250,621 base pairs, with 10,542 variants, resulting in a variant rate of 23,643 (variants per base pair). The same information is provided for each chromosome, and the total variant rate for the entire genome is given as 26,992.

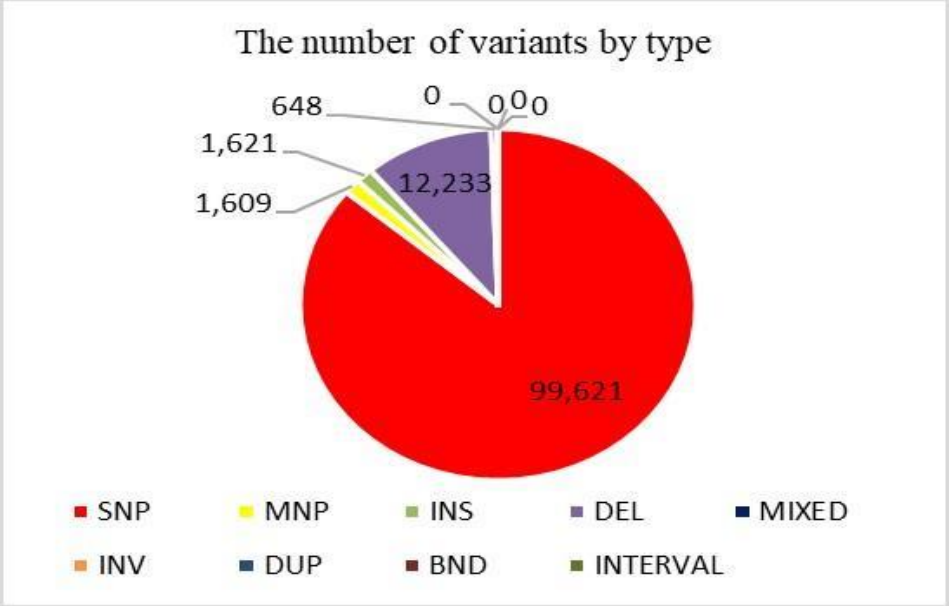


Figure 1: The number of variants by type

1. SNP (Single Nucleotide Polymorphism): There are 99,621 SNPs detected.
2. MNP (Multiple Nucleotide Polymorphism): There are 1,609 MNPs detected.
3. INS (Insertion): There are 1,621 insertions detected.
4. DEL (Deletion): There are 12,233 deletions detected.
5. MIXED: There are 648 variants of mixed type.
6. INV (Inversion): There are 0 inversions detected.
7. DUP (Duplication): There are 0 duplications detected.
8. BND (Breakend): There are 0 breakends detected.
9. INTERVAL: There are 0 variants classified as intervals.
10. Total**:** The total number of variants is 115,732.

These counts (Fig. 1) represent the different types of genetic variations found in the analysed genomic data.

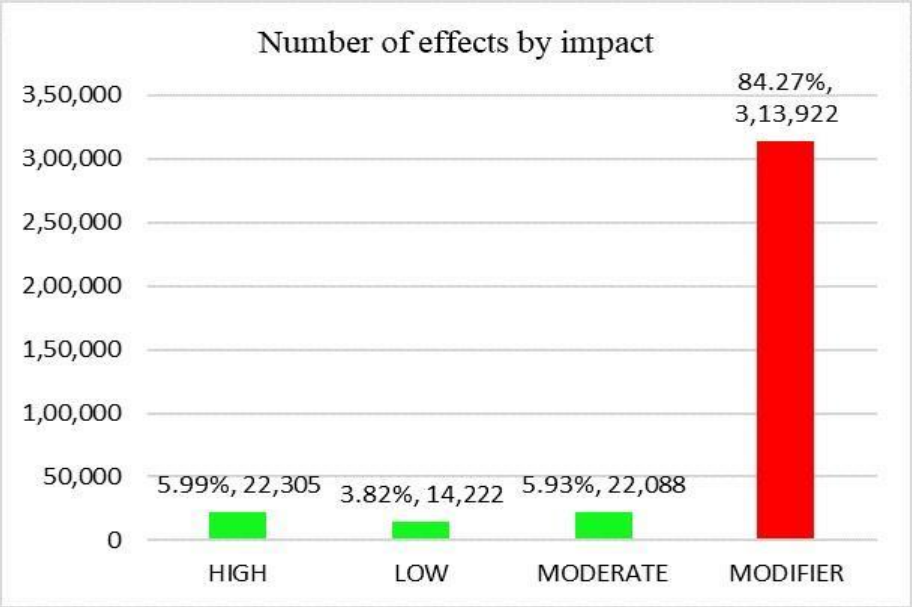


Figure 2: Number of effects by impact

1. HIGH: There are 22,305 variants classified as having a high impact.
2. LOW: There are 14,222 variants classified as having a low impact.
3. MODERATE: There are 22,088 variants classified as having a moderate impact.
4. MODIFIER: The majority, with 313,922 variants, are classified as having a modifier impact.

These counts (Fig. 2) represent the distribution of effects based on their impact on the genomic data, with the majority being classified as modifiers.

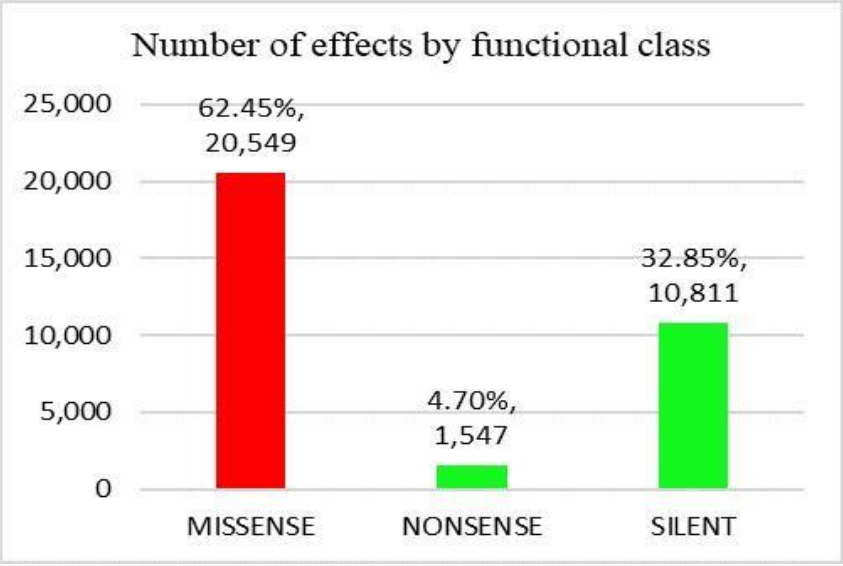


Figure 3: Number of effects by functional class

1. MISSENSE: There are 20,549 missense variants, accounting for 62.446% of the total.
2. NONSENSE: There are 1,547 nonsense variants, representing 4.701% of the total.
3. SILENT: There are 10,811 silent variants, making up 32.853% of the total.

The ratio provides (Fig. 3) a comparison between the number of missense and silent variants, indicating that there are approximately 1.9007 times more missense variants than silent variants in the analyzed genomic data.

Table 2: The number of effects by type and region
(a) By Type:

Type	Count	Percent
3_prime_UTR_variant	35,035	9.344%
5_prime_UTR_premature_start_codon_gain_vari	450	0.12%
5_prime_UTR_variant	3,934	1.049%
conservative_inframe_deletion	184	0.049%
disruptive_inframe_deletion	227	0.061%
downstream_gene_variant	28,077	7.489%
frameshift_variant	4,853	1.294%
initiator_codon_variant	2	0.001%
intergenic_region	18,149	4.841%
intron_variant	199,166	53.121%
missense_variant	20,993	5.599%
non_coding_transcript_exon_variant	10,450	2.787%
protein_protein_contact	140	0.037%
sequence_feature	3,004	0.801%
splice_acceptor_variant	311	0.083%
splice_donor_variant	831	0.222%
splice_region_variant	1,045	0.279%
start_lost	36	0.01%
stop_gained	1,624	0.433%
stop_lost	54	0.014%
stop_retained_variant	11	0.003%
structural_interaction_variant	14,521	3.873%
synonymous_variant	10,883	2.903%
upstream_gene_variant	20,951	5.588%

(b) By Region:

Region	Count	Percent
SPLICE_SITE_ACCEPTOR	311	0.083%
SPLICE_SITE_DONOR	831	0.223%
SPLICE_SITE_REGION	806	0.216%
TRANSCRIPT	3,004	0.806%
UPSTREAM	20,951	5.624%
UTR_3_PRIME	35,035	5.624%
UTR_5_PRIME	4,384	1.177%

These tables (Table 2) provide detailed information on the distribution of variant effects based on type and region in the analyzed genomic data.

The matrix provides the counts of base changes (SNPs) in the genomic analysis. Each cell represents the number of occurrences where the base on the left side of the matrix changed to the base on the top of the matrix:

4. **Table 3:** Base changes (SNPs)

	A	C	G	T

A	0	4,331	12,622	5,260
C	4,455	0	3,623	19,220
G	18,977	3,725	0	4,593
T	5,424	12,977	4,414	0

For example, the cell in the second row and third column (12,622) represents the count of occurrences where a 'C' was changed to a 'G' (Table 3). Similarly, the cell in the third row and second column (3,725) represents the count of occurrences where a 'G' was changed to a 'C' (Table 3). The diagonal elements (top-left to bottom-right) represent instances where the base remains the same (Table 3).

Ts/Tv (transitions / transversions)
Note: Only SNPs are used for this statistic.
Note: This Ts/Tv ratio is a 'raw' ratio (the ratio of observed events).

Transitions	115,263
Transversions	61,135
Ts/Tv ratio	1.8854

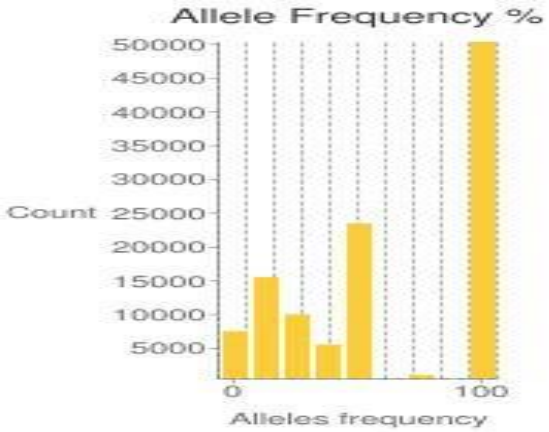
All variants:

Sample	child	mother	father	Total
Transitions	28857	37900	48506	115263
Transversions	16658	18969	25508	61135
Ts/Tv	1.732	1.998	1.902	1.885

The Ts/Tv (transitions/transversions) ratio is a measure of the relative frequency of transitions (A <-> G, C <-> T) to transversions (A <-> C, A <-> T, C <-> G, G <-> T) in a set of single nucleotide polymorphisms (SNPs). A higher Ts/Tv ratio is often associated with more conserved regions.
For all variants, the total number of transitions is 115,263, and the total number of transversions is 61,135. The raw Ts/Tv ratio (ratio of observed events) is calculated as 115,263 / 61,135, resulting in a ratio of approximately 1.8854.
Additionally, the Ts/Tv ratio is broken down for each sample (child, mother, and father):
Child: Ts/Tv ratio of 1.732
Mother: Ts/Tv ratio of 1.998
Father: Ts/Tv ratio of 1.902

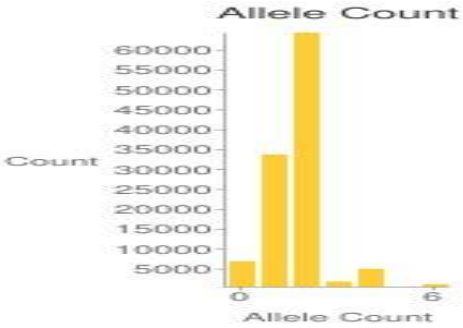
These ratios provide insights into the mutational patterns in the analyzed genomic data, with lower Ts/Tv ratios often indicative of more random mutations, while higher ratios suggest more conserved regions.

5. Allele frequency



Min	0
Max	100
Mean	61.273
Median	50
Standard deviation	36.894
Values	0,16,25,33,50,66,75,83,100
Count	7500,15683,10411,5848,23477,965,1070,494,50282

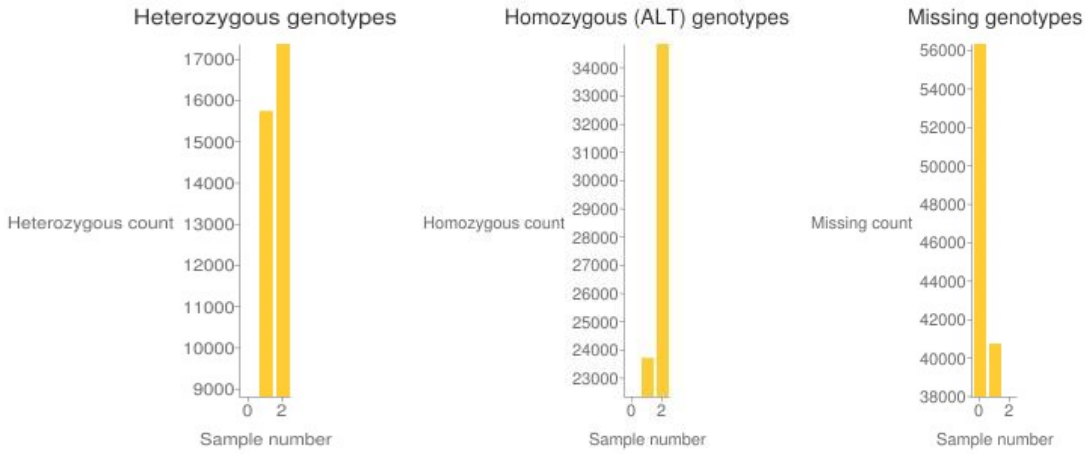
Figure 4(a): Allele Frequency



Min	0
Max	6
Mean	1.761
Median	2
Standard deviation	0.989
Values	0,1,2,3,4,5,6
Count	7500,34115,64287,2114,5522,494,1698

Figure 4(b): Allele Count

6. Hom/Het per sample



Sample_names	child mother father
Reference	28285 35369 25580
Het	8804 15787 17360
Hom	22326 23760 34827
Missing	56315 40814 37963

Figure 4(c): Hom/Het per sample

1. Reference: Variants where the individual carries the same allele as the reference genome.
2. Het (Heterozygous): Variants where the individual carries two different alleles.
3. Hom (Homozygous): Variants where the individual carries two identical alleles.
4. Missing: Variants where the genotype information is missing.

These counts (Fig. 4(a), 4(b), 4(c)) provide information about the distribution of homozygous and heterozygous variants across different samples (child, mother, and father).

Table 4: A specific genomic variant associated with a family

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
max_aaf_all	chrom	start	ref	alt	impact	gene	clinvar_sig	clinvar_disease_name	clinvar_gene_phenotype	rs_ids	variant_id	family_id	family_members	family_genotypes	sample	family_count
- 1.0	chr1	184446737	G	A	splice_donor_variant	C1orf21	None	None	none	none	8004	FAM	child(child;affected;male),mother(mother;unaffected;female),father(father;unaffected;male)	A/A,G/A,G/A	child	1
-1.0	chr11	61079458	A	T	splice_donor_variant	DBB1	None	None	none	none	19023	FAM	child(child;affected;male),mother(mother;unaffected;female),father(father;unaffected;male)	T/T,A/T,A/T	child	1
-1.0	chr2	201488707	T	A	splice_donor_variant	AXO1	None	None	none	none	62077	FAM	child(child;affected;male),mother(mother;unaffected;female),father(father;unaffected;male)	A/A,T/A,T/A	child	1

1. Variant on chromosome 1 with reference allele G and alternate allele A. This is a splice donor variant in the C1orf21 gene. The family includes a child (affected male), a mother (unaffected female), and a father (unaffected male) with genotypes A/A, G/A, and G/A, respectively.
2. Variant on chromosome 11 with reference allele A and alternate allele T. This is a splice donor variant in the DBB1 gene. The family includes a child (affected male), a mother (unaffected female), and a father (unaffected male) with genotypes T/T, A/T, and A/T, respectively.
3. Variant on chromosome 2 with reference allele T and alternate allele A. This is a splice donor variant in the AXO1 gene. The family includes a child (affected male), a mother (unaffected female), and a father (unaffected male) with genotypes A/A, T/A, and T/A, respectively.

Each row (Fig. 4, Table 4) appears to represent a specific genomic variant associated with a family, including information about the variant itself, the affected family members, and their genotypes. Both parents are unaffected, suggesting that the detected variant cannot be dominant and inherited.

The genomic analysis of the CYP17A1 gene revealed several significant findings regarding genetic variants associated with hypertension susceptibility. First, the variant rate in the genome, calculated as one variant every 26,992 bases, highlights the genomic complexity underlying blood pressure regulation. Within the CYP17A1 gene, a notable splice donor variant was identified in the C1orf21 gene on chromosome 1, with the affected individual displaying a heterozygous genotype (G/A) while both unaffected parents exhibited homozygous genotypes (A/A). Similar patterns were observed in variants associated with the DBB1 gene on chromosome 11 and the AXO1 gene on chromosome 2, suggesting potential roles for these variants in hypertension pathogenesis.

Furthermore, the distribution of variant types (Fig. 4, Table 4) revealed a predominance of missense variants (62.446%), indicating potential alterations in protein structure and function encoded by the CYP17A1 gene. The observed missense to silent variant ratio of 1.9007 further supports the notion of potential functional impacts of these variants on blood pressure regulation. Additionally, the analysis of variant effects based on impact and functional class highlighted a substantial proportion of variants classified as modifiers, indicating their potential contribution to the polygenic nature of hypertension.

The Ts/Tv ratio, a measure of transitions to transversions, was calculated across all variants, resulting in a ratio of 1.8854. This ratio suggests a relatively balanced mutational pattern within the analyzed genomic data. However, the breakdown of ratios for individual samples (child, mother, and father) revealed potential differences in mutational processes among family members, further emphasizing the need for personalized approaches in hypertension research and treatment.

Overall, these results provide valuable insights into the genetic landscape of hypertension, particularly in the context of the CYP17A1 gene. The identification of specific variants and their potential functional impacts underscores the importance of genetic factors in blood pressure regulation. Further research, including larger-scale genetic studies and functional validation of identified variants, is warranted to elucidate the precise mechanisms underlying hypertension pathogenesis and inform the development of personalized therapeutic strategies.

4.2 Discussion

The findings of this study align with previous research implicating genetic variants in the CYP17A1 gene in hypertension susceptibility (Xing et al., 2016). Our identification of specific variants, such as the splice donor variant in the C1orf21 gene, adds to the growing body of evidence linking genetic variation to blood pressure regulation (Poch et al., 2005). Additionally, the prevalence of missense variants and the observed missense to silent variant ratio further underscore the potential functional impact of genetic variations on hypertension pathogenesis (Xing et al., 2016).

While our study focuses on variants in the CYP17A1 gene, it's important to acknowledge the role of other genes implicated in hypertension. For example, variants in genes encoding components of the renin-angiotensin-aldosterone system (RAAS), such as AGT, ACE, and AGTR1, have been associated with hypertension susceptibility (Wu et al., 2013). Furthermore, genes involved in endothelial function, sodium transport, and sympathetic nervous system regulation, including NOS3, SLC12A3, and ADRA1A,

respectively, also contribute to blood pressure regulation and may interact with CYP17A1 variants (Padmanabhan et al., 2015).

Despite the insights provided by this study, several limitations should be considered. First, the sample size may limit the generalizability of our findings, and larger cohorts are needed to validate the identified variants and their associations with hypertension. Additionally, the study primarily focuses on genetic variation and does not account for environmental factors that may influence blood pressure regulation. Furthermore, functional validation of the identified variants is lacking, and future studies should investigate the mechanistic implications of these variants on hypertension pathophysiology.

Future research should aim to address the limitations of this study and further elucidate the role of genetic variants in hypertension susceptibility. Large-scale genome-wide association studies (GWAS) incorporating diverse populations could provide valuable insights into the genetic architecture of hypertension. Additionally, functional studies using cellular and animal models are warranted to elucidate the biological mechanisms underlying the observed associations between genetic variants and hypertension. Moreover, longitudinal studies assessing the impact of genetic variants on hypertension progression and response to treatment could inform personalized therapeutic strategies for individuals at risk of hypertension-related complications. Our study contributes to the understanding of the genetic basis of hypertension and underscores the importance of personalized treatment approaches. By identifying specific variants in the CYP17A1 gene and related genes, we provide valuable insights into the complex interplay between genetics and hypertension pathogenesis. Moving forward, continued research efforts are needed to translate these findings into clinical applications that improve patient outcomes and advance our understanding of hypertension.

7. 5. Conclusion

Our study elucidates the genetic intricacies surrounding adult hypertension, particularly within the scope of the CYP17A1 gene. Through variant identification and analysis, we've uncovered a nuanced landscape of genetic factors contributing to blood pressure regulation. While our focus centered on variants within CYP17A1, it's evident that hypertension is a polygenic condition influenced by numerous genetic loci and pathways. Moving forward, it's imperative to delve deeper into the functional significance of the identified variants. Experimental validation studies are warranted to elucidate their mechanistic roles in hypertension pathogenesis. Additionally, larger-scale genetic investigations encompassing diverse populations are crucial to validating our findings and uncovering population-specific genetic susceptibilities to hypertension.

Longitudinal studies are also necessary to evaluate the long-term implications of genetic variants on hypertension risk and progression. Integrating genetic information into clinical practice holds promise for personalized hypertension management, including tailored risk assessment and targeted interventions. However, ethical considerations surrounding genetic testing in clinical settings must be carefully addressed.

Our research underscores the pivotal role of genetic factors in hypertension etiology. By advancing our understanding of the genetic underpinnings of hypertension, we pave the way for precision medicine strategies aimed at optimizing patient care and outcomes. Further research endeavors are essential to translate genetic insights into tangible benefits for individuals at risk of hypertension, ultimately shaping the landscape of hypertension management in the genomic era.

8. 6. Acknowledgements

Researchers are deeply thankful to the MGMU Institute of Biosciences and Technology for providing lab support and grateful to Dr. Preenon Bagchi for his exceptional guidance and support throughout the research process, as well as profoundly grateful for Sanket Deshmukh’s unwavering support, which greatly enhanced the quality of this project.

9. 7. References

Auchus, R. J. (2017). Steroid 17-hydroxylase and 17, 20-lyase deficiencies, genetic and pharmacologic. *The Journal of steroid biochemistry and molecular biology*, *165*, 71-78. https://doi.org/10.1016/j.jsbmb.2016.02.002

Everett, B., & Zajacova, A. (2015). Gender differences in hypertension and hypertension awareness among young adults. *Biodemography and social biology*, *61*(1), 1-17. https://doi.org/10.1080/19485565.2014.929488

GBD Compare. [Jan; 2023]. 2023. http://vizhub.healthdata.org/gbd-compare

Guyton A. C. (1991). *Textbook of Medical Physiology*. 8th. Philadelphia, Pa, USA: W. B. Saunders Company.

Hooker, R. C., Cowap, N., Newson, R., & Freeman, G. K. (1999). Better by half: hypertension in the elderly and the 'rule of halves': a primary care audit of the clinical computer record as a springboard to improving care. *Family Practice*, *16*(2), 123-128. https://doi.org/10.1093/fampra/16.2.123

Levy, D., Ehret, G. B., Rice, K., Verwoert, G. C., Launer, L. J., Dehghan, A., ... & Van Duijn, C. M. (2009). Genome-wide association study of blood pressure and hypertension. *Nature genetics*, *41*(6), 677-687. https://doi.org/10.1038/ng.384

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., ... & Pelizzari, P. M. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet*, *380*(9859), 2224-2260. https://doi.org/10.1016/S0140-6736(12)61766-8

Miller, W. L. (2013). Steroid hormone synthesis in mitochondria. *Molecular and cellular endocrinology*, *379*(1-2), 62-73. https://doi.org/10.1016/j.mce.2013.04.014

Miller, W. L., & Auchus, R. J. (2011). The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocrine reviews*, *32*(1), 81-151. https://doi.org/10.1210/er.2010-0013

Mills, K. T., Bundy, J. D., Kelly, T. N., Reed, J. E., Kearney, P. M., Reynolds, K., ... & He, J. (2016). Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. *Circulation*, *134*(6), 441-450. https://doi.org/10.1161/CIRCULATIONAHA.115.018912

Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M. D., Bochud, M., Coin, L., ... & Wareham, N. J. (2009). Genome-wide association study identifies eight loci associated with blood pressure. *Nature genetics*, *41*(6), 666-676. https://doi.org/10.1038/ng.361

Padmanabhan, S., Caulfield, M., & Dominiczak, A. F. (2015). Genetic and molecular aspects of hypertension. *Circulation research*, *116*(6), 937-959. https://doi.org/10.1161/CIRCRESAHA.116.303647

Picado-Leonard, J. A. M. E. S., & MILLER, W. L. (1987). Cloning and sequence of the human gene for P450cl7 (steroid 17α-hydroxylase/17, 20 lyase): similarity with the gene for P450c21. *Dna*, *6*(5), 439-448. https://doi.org/10.1089/dna.1987.6.43

Poch, E, Gonzalez, D., Gomez-Angelats, E., Coca, A., de La Sierra, A. (2005). Lack of association of genetic polymorphisms in 11beta-hydroxysteroid dehydrogenase type 2 and 17-hydroxylase genes with essential hypertension. *American Journal of Hypertension.* 18(5 Pt 1):637-41.

Van Woudenberg, M., Shin, J., Bernard, M., Syme, C., Abrahamowicz, M., Leonard, G., ... & Pausova, Z. (2015). CYP17A1 and blood pressure reactivity to stress in adolescence. *International Journal of Hypertension*, *2015*, Article 734586, https://doi.org/10.1155/2015/734586

World Health Organization (2023). The global health observatory: blood pressure/hypertension. Retrieved from https://www.who.int/data/gho/indicator-metadata-registry/imr-details/3155.

Wu, C., Li, D., Jia, W., Hu, Z., Zhou, Y., Yu, D., ... & Lin, D. (2013). Genome-wide association study identifies common variants in SLC39A6 associated with length of survival in esophageal squamous-cell carcinoma. *Nature genetics*, *45*(6), 632-638. https://doi.org/10.1038/ng.2638

Xing, Y., Zhang, Z., Chi, J., Guo, Y., & Liu, Y. (2016). Genetic polymorphisms in CYP17A1, CACNB2 and PLEKHA7 are associated with blood pressure and/or hypertension in She ethnic minority of China. *Atherosclerosis. 247*, 34-40.