AVESTHAGEN BRINGING SCIENCE TO LIFE

PHARMACOGENOMICS REPORT

Submitted by Priyanka Chaudhari

Pharmacogenomics Gene Panel Screening of Zoroastrian-Parsi Family members and *in-silico* identification of Novel Missense variants.

DECLARATION

I, **Priyanka Chaudhari**, would like to declare that the project report entitled "Pharmacogenomics Gene Panel Screening of Zoroastrian-Parsi Family members and *in-silico* identification of Novel Missense Variants" is submitted toward the completion of my internship at Avesthagen Limited. A comprehensive literature review and data analysis was done, and the relevant information was meticulously compiled into this report. The data presented in the report, the data generated through this work, and any extensions of it are proprietary knowledge of Avesthagen Limited. No part of this project has been used by or submitted to any other organization. This project was conducted at Avesthagen Limited from **January 5th, 2023, to April 4th, 2023,** under the direction of Dr. Villoo Zareer Morawala-Patell, with the guidance and supervision of Dr. Kashyap Krishnasamy and Dr. Vasundhara Gadiyaram.

The Avestagenome Project®

The Avestagenome Project® is a systems biology study that aims to understand the molecular basis of longevity in the Zoroastrian-Parsi (ZP) population and identify critical disease associated genomic variants that underscore the health of this community.

The project collected blood samples from members of the Parsi community (cohort population) and Whole Genome Sequencing (WGS) using NEXTGEN Sequencing Technology and genomic analysis of the samples was carried out.

Initial analysis of data indicated greater longevity, fewer cases of head, neck and oesophageal cancers and higher instances of neurodegenerative diseases (especially Parkinson's and Alzheimer's diseases), breast and prostate cancers, cardiac related illnesses, musculoskeletal disorders and male and female infertility.

The Avestagenome Project® aims at identifying the genetic factors within the Zoroastrian populations that predisposes certain individuals to diseases and determine the genetic basis for longevity. It also aims to develop drug targets and molecular biomarkers for Predictive, Preventive and Personalized Healthcare. This would help devise strategies and advances in healthcare systems to prevent the decline of the population.

ACKNOWLEDGEMENTS

I express my sincere appreciation and immense gratitude to **Dr. Villoo Zareer Morawala-Patell** (Founder, Chairman, and Managing Director at Avesthagen Limited) for giving me this exciting oppurtunity of being a Project Intern and for allowing me to work on this report by analyzing the Pharmacogenomic data of Zoroastrian-Parsi members.

I thank **Dr. Renuka Jain** (Senior Vice President, Avesthagen Limited) for her constant support, encouragement, and expert guidance throughout the duration of my internship.

I also thank my supervisors **Dr. Kashyap Krishnasamy** (Senior Scientist, Avesthagen Limited) and **Dr. Vasundhara Gadiyaram** (Head of Bioinfomatics, Avesthagen Limited) for their expertise and valuable inputs.

I would also like to express my gratitude to everyone else at Avesthagen Limited for providing an enabling environment to learn and grow. Their work ethic and professionalism spurred me to give my best.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Precision or personalized medicine is all about the right drug being administered in the right dose to the right patient. Pharmacogenetic-pharmacogenomic tests can help the medical practitioner assess the risks and benefits accurately before prescribing a drug for chronic conditions. Pharmacogenetics is defined as the study of genetic causes that lead to variation in individual response to drugs. Pharmacogenomics is defined as the study of the whole genome and all possible mutations that can alter this individual drug response. These terms are often used interchangeably.

The interindividual genetic variants can either be inherited (germ-line variations) or acquired (somatic mutations). Pharmacogenomics is paving the way for personalized medicine as it allows the assessment of these specific interindividual genetic differences. Such genome-wide association studies (GWAS) can help identify important traits and genomic variants that are specific to certain population groups and can measure the risk associated with a particular disease. Whole genomes and exomes of large numbers of individuals can be mapped for millions of genomic variants to check for associations with a particular disease or condition. Early detection/prediction, prevention, diagnosis and treatment of a disease, drug safety and toxicity can be analyzed with the help of pharmacogenomics.

The Parsis are an ethnoreligious group of the Indian subcontinent who practice Zoroastrianism. Around 7th century AD, the Parsis fled persecution in Persia and arrived on the Western coast of India and blended into Indian culture. To retain the purity of their lineage and heritage they opted for endogamous marriages. Present-day Iranian Zoroastrians stopped mixing with other groups shortly after the Arab conquest of Persia in 633-654 CE. Zoroastrian-Parsis in Iran and India have higher genetic homogeneity than other sampled groups in their respective countries. This is consistent with their current endogamy practices. It is now well-established that genetic isolation and endogamous practices are linked to higher rates of disease prevalence. After inbreeding for many generations there are several diseases associated genetic mutations accumulated within this community that alter an individual's response to drugs or predispose these individuals to life-long chronic conditions like certain cancers, cardiac conditions, neurodegenerative diseases, and infertility.

The study entailed performance and review of variant calling on the Pharmacogenomic gene panel of 22 members of Zoroastrian-Parsi (ZP) lineage using Whole Genome Sequence (WGS) data analysis. This was done with the help of Congenica® software and in-silico biological function (Polyphen and SIFT) predictions were carried out on proteins encoded by genes in which novel missense mutations occur.

Mutations that result in amino acid substitutions are known to affect the stability of proteins and their ability to bind to biomolecules and can thus be linked to diseases. Identifying genetic factors in Zoroastrian populations could be useful in understanding the prevalence of disease or distinct phenotypic traits in the community, as well as in devising strategies and advances in healthcare systems to halt the population decline.

DISCUSSION

Pedigree Chart of the Zoroastrian-Parsi Family Members

Each individual with a WGS was given Sample IDs by the Congenica® software. These sample IDs have been replaced with individual numbers for ease of identification.

Sample ID	Individual No.	Sample ID	Individual No.
SO_8917_145644454_30X	1	14810707	12
SO_8917_145644457_30X	2	14810709	13
SO_8917_145644466_30X	3	14810712	14
14810642_30X	4	14810715	15
14810643_30X	5	14810717	16
14810644_452	6	14810718	17
14810648_30X	7	14810719	18
14810649_30X	8	14810720	19
14810662	9	14810727	20
14810694	10	14811043	21
14810704	11	14811055	22

Table No. 1 – Sample IDs replaced with Individual numb

Pharmacogenomic Gene Panel

Following WGS and study of published literature and databases, a gene panel was created on the Congenica® software. This panel is comprised of 11 Pharmacogenomic genes: ABCG2, CYP2C19, CYP2D6, DPYD, HLA-B, NUDT15, RARG, SLCO1B1, TPMT, UGT1A1, and VKORC1. Variants were found only across 6 genes for the 22 ZP Family members: ABCG2, CYP2C19, CYP2D6, DPYD, HLA-B, and SLCO1B1. Hence, the following descriptions, tables and figures in the report will only include these 6 genes.

WGS results data of all samples were obtained from the Congenica® software and the raw data was compiled into excel spreadsheets for data interpretation and analysis. A variant calling was performed on the WGS data of the family using the Congenica® software and all the NCBI databases and the variants were given Polyphen and SIFT prediction scores. A comprehensive analysis of the raw data acquired from the Congenica® software and NCBI databases was done and the following results were obtained:

Type of variants	No. of variants
Missense Variant	64
Frameshift Variant	10
Synonymous Variant	9
Intron Variant	8
Upstream Gene Variant	1
Inframe Deletion	1
Total	**93**

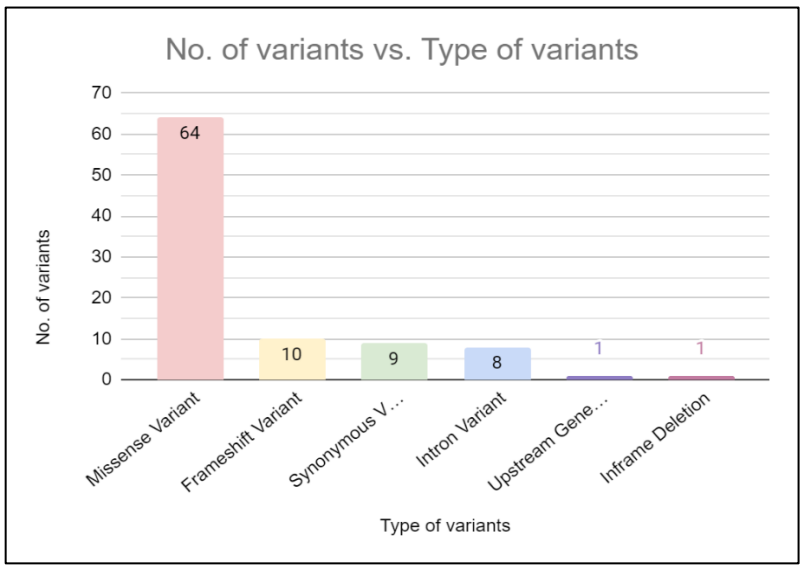

Table 2 and Figure 1 – Table and Column chart showing the different type of variants observed.

There are a total of **93 distinct variations** distributed across the 6 Pharmacogenomic genes of 22 ZP individuals. There are 64 (68.8%) Missense, 10 (10.8%) Frameshift, 9 (9.7%) Synonymous, and 8 (8.6%) Intronic variations. 1 (1.05%) Upstream Gene Variant and 1 (1.05%) Inframe Deletion are also observed.

Types of variants

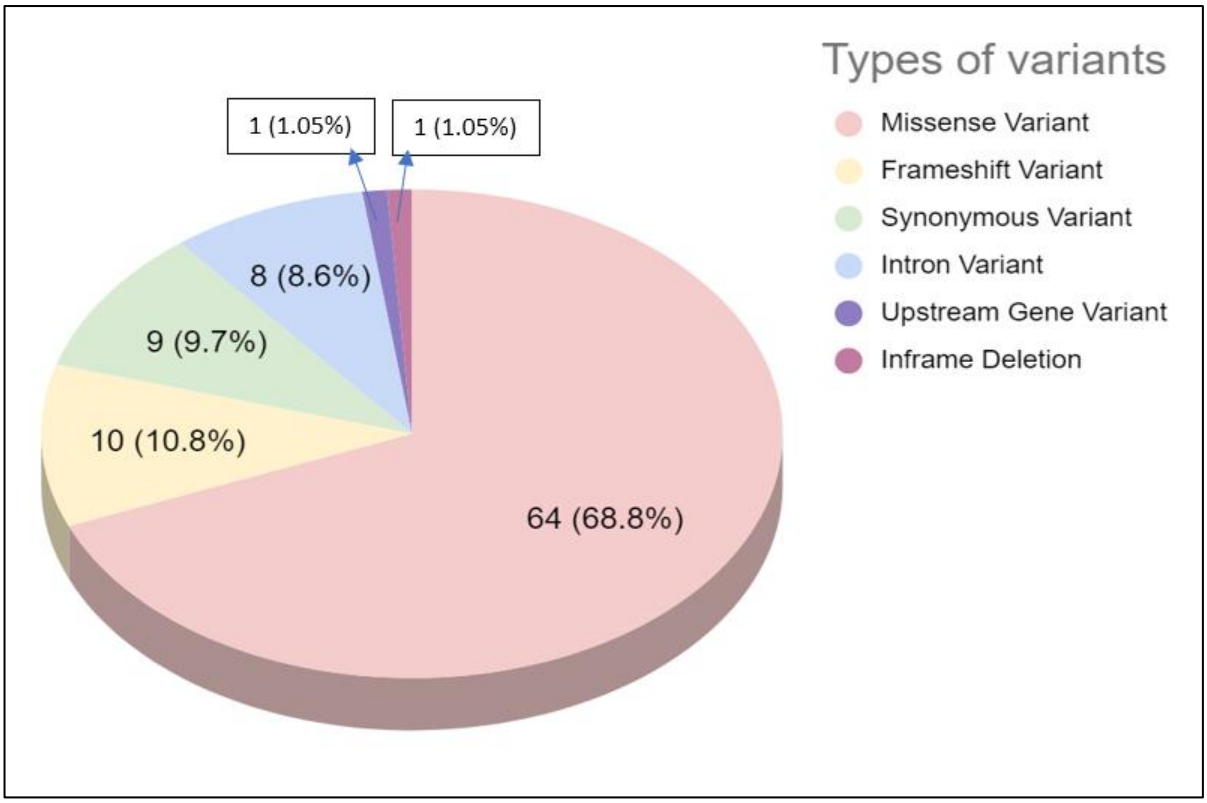

Figure 2 – Pie chart showing the distribution of different type of variants observed in 6 Pharmacogenomic genes across a sample size of 22 individuals.

As represented below, HLA-B gene has the most number of distinct variations and is known to be a highly polymorphic gene. A total number of 61 variations are seen in HLA-B gene, 11 in CYP2D6, 7 in CYP2C19 and DPYD, 4 in ABCG2 and 3 variation in SLCO1B1 gene.

Gene	Missense Variant	Frameshift Variant	Synonymous Variant	Intron Variant	Upstream Gene Variant	Inframe Deletion	Total
ABCG2	3				1		**4**
CYP2C19	6		1				**7**
CYP2D6	9		1			1	**11**
DPYD	5		2				**7**
HLA-B	38	10	5	8			**61**
SLCO1B1	3						**3**
Total	**93 variants across all genes**						

Table 3 – Number and type of variants found per gene.

No. of variants per gene

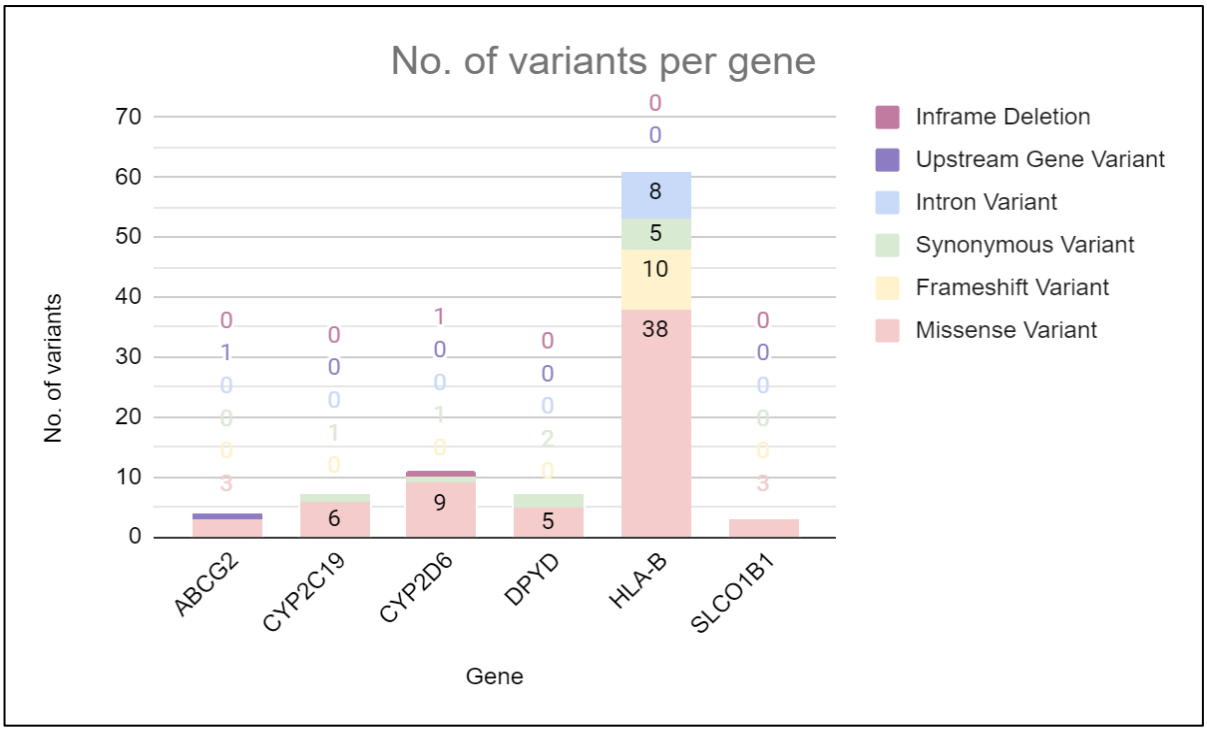

Figure 3 – Column chart showing types of variants across 6 Pharmacogenomic genes.

ABCG2 gene has 3 Missense and 1 Upstream Gene variant. CYP2C19 gene has 6 Missense and 1 Synonymous variant present. Similarly, CYP2D6 gene has 9 Missense, 1 Synonymous and 1 Inframe Deletion observed whereas, DPYD gene only has 5 Missense variants and 2 Synonymous variants present. The gene with the most number and type of variations observed is HLA-B gene. It has 38 Missense, 10 Frameshift, 8 Intronic and 5 Synonymous variations. And lastly, the gene with the least number of variants is SLCO1B1 as it has only 3 Missense variants across all 22 individuals.

The most number of variants is observed in 14810715 (Individual No. 15) with 7 Missense variants, 1 Synonymous and 1 Upstream Gene Variant. 7 Missense Variants are also present in 14810643_30X (Individual No. 5). Only 1 variant is found in both SO_8917_145644466_30X (Individual No. 3) and 14810662 (Individual No. 9).

No. of variants per individual

Sr. No.	Sample ID	No. of variants	Sr. No.	Sample ID	No. of variants
1	SO_8917_145644454_30X	3	12	14810707	4
2	SO_8917_145644457_30X	3	13	14810709	6
3	SO_8917_145644466_30X	1	14	14810712	2
4	14810642_30X	6	15	14810715	**9**
5	14810643_30X	7	16	14810717	6
6	14810644_452	5	17	14810718	2
7	14810648_30X	2	18	14810719	7
8	14810649_30X	6	19	14810720	6
9	14810662	1	20	14810727	2
10	14810694	3	21	14811043	3
11	14810704	4	22	14811055	5
Total			**93**		

Table 4 – No. of variants per individual.

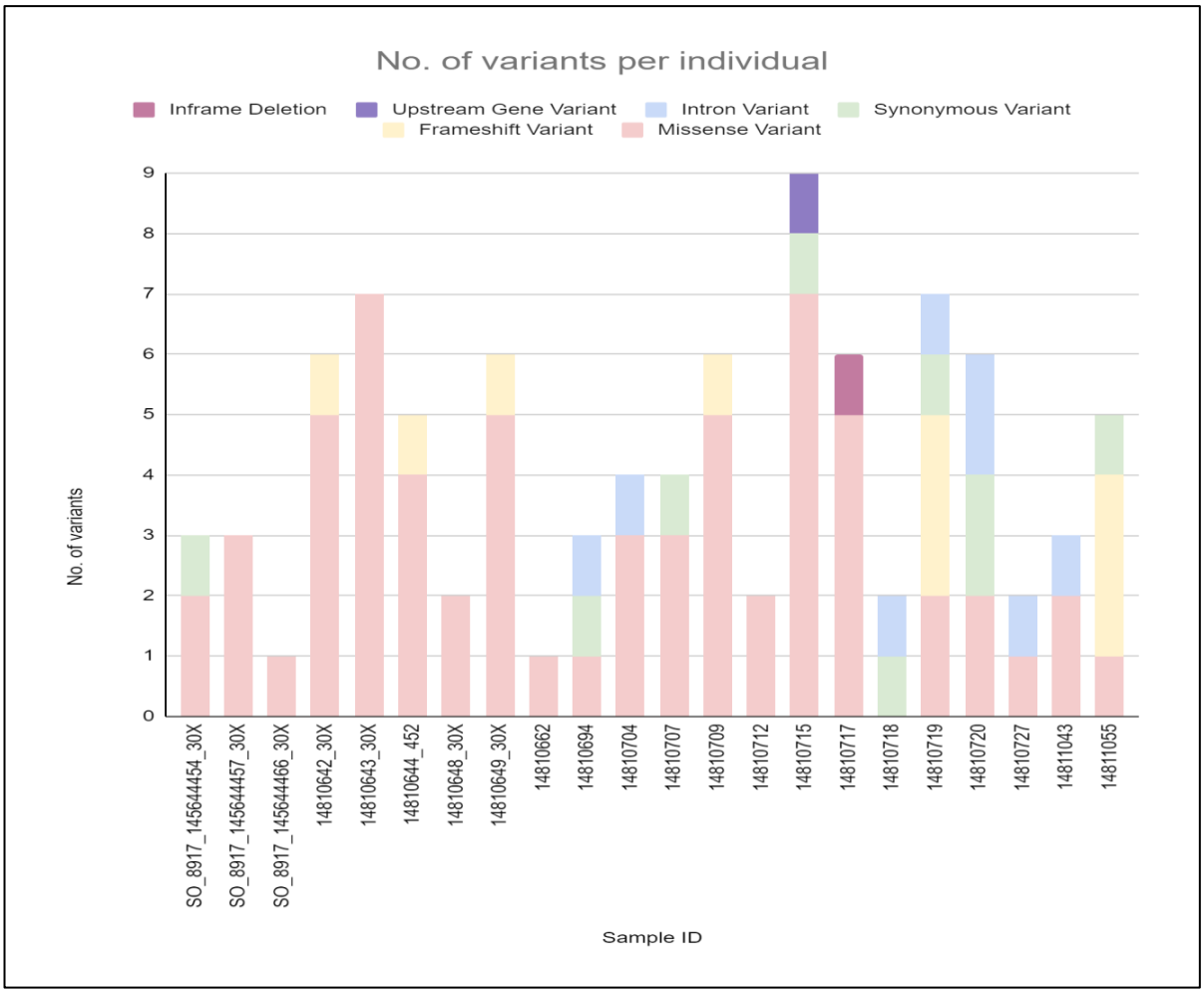

Figure 4 – Column chart showing the total no. of variants observed in each individual.

Total number and type of variants found per individual

Sr. No.	Sample ID	Missense Variant	Frameshift Variant	Synonymous Variant	Intron Variant	Upstream Gene Variant	Inframe Deletion
1	SO_8917_145644454_30X	2		1			
2	SO_8917_145644457_30X	3					
3	SO_8917_145644466_30X	1					
4	14810642_30X	5	1				
5	14810643_30X	7					
6	14810644_452	4	1				
7	14810648_30X	2					
8	14810649_30X	5	1				
9	14810662	1					
10	14810694	1		1	1		
11	14810704	3			1		
12	14810707	3		1			
13	14810709	5	1				
14	14810712	2					
15	14810715	7		1		1	
16	14810717	5					1
17	14810718			1	1		
18	14810719	2	3	1	1		
19	14810720	2		2	2		
20	14810727	1			1		
21	14811043	2			1		
22	14811055	1	3	1			
Total		**64**	**10**	**9**	**8**	**1**	**1**

Table 5 – The total number and type of variants found per individual.

Distribution of variants per individual per gene

Sr. No.	SAMPLE ID	ABCG2	CYP2C19	CYP2D6	DPYD	HLA-B	SLCO1B1
1	SO_8917_145644454_30X				1	1	1
2	SO_8917_145644457_30X	1				2	
3	SO_8917_145644466_30X	1					
4	14810642_30X		2		1	2 1	
5	14810643_30X					7	
6	14810644_452		1		1	2 1	
7	14810648_30X	1			1		
8	14810649_30X		1		1	3 1	
9	14810662						1
10	14810694			1		1 1	
11	14810704			2		1	1
12	14810707					3 1	
13	14810709			4		1 1	
14	14810712		1	1			
15	14810715	1			1	7	
16	14810717			1		5	
17	14810718		1			1	
18	14810719					3 2 1 1	
19	14810720			1		2 1 2	
20	14810727				1	1	
21	14811043		1	1		1	
22	14811055					3 1 1	

Color	Type of variants	No. of variants
	Missense Variant	64
	Frameshift Variant	10
	Synonymous Variant	9
	Intron Variant	8
	Upstream Gene Variant	1
	Inframe Deletion	1
Total no. of variants		**93**

Table 6 – Distribution of variants per individual per gene

Individual No. 1 – SO_8917_145644454_30X

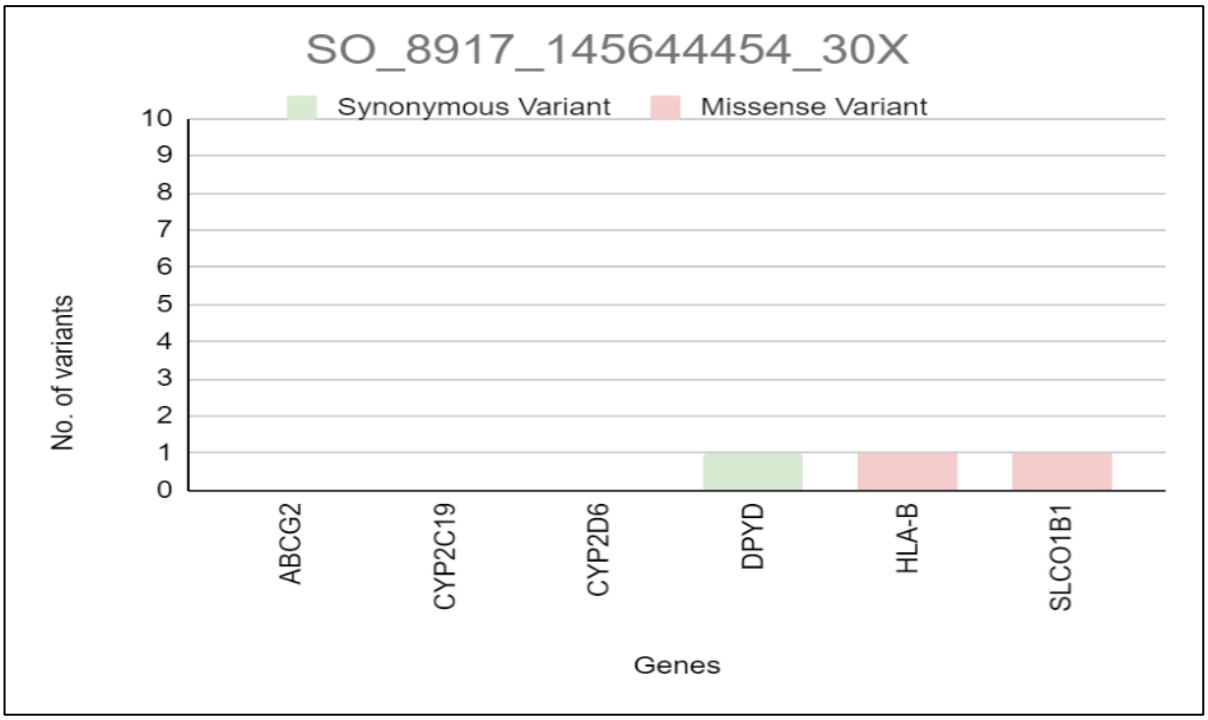

Figure 5 – Column chart showing variants observed in SO_8917_145644454_30X

Individual No. 1 has the following variants present:

c.1236G>A p.(=) is a DPYD gene synonymous variant. It has a heterozygous genotype with a gene score of HI 0.272. Variant is C > T and the amino acid change is at E – Glutamic acid. The maximum allelic frequency of this variant is 0.0239.

c.97T>G p.Tyr33Asp is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.005 and SIFT score of tolerated_low_confidence 0.21. Variant is A > C and the amino acid change is Y33D – Aspartic acid (D) replaces Tyrosine (Y) at codon 33. The maximum allelic frequency of this variant is 0.0754.

c.1929A>C p.Leu643Phe is a SLCO1B1 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.055. Polyphen score is benign 0.005 and SIFT score is tolerated 1. Variant is A > C and the amino acid change is L643F – Phenylalanine (F) replaces Leucine (L) at codon 643. The maximum allelic frequency of this variant is 0.0654.

Individual No. 2 – SO_8917_145644457_30X

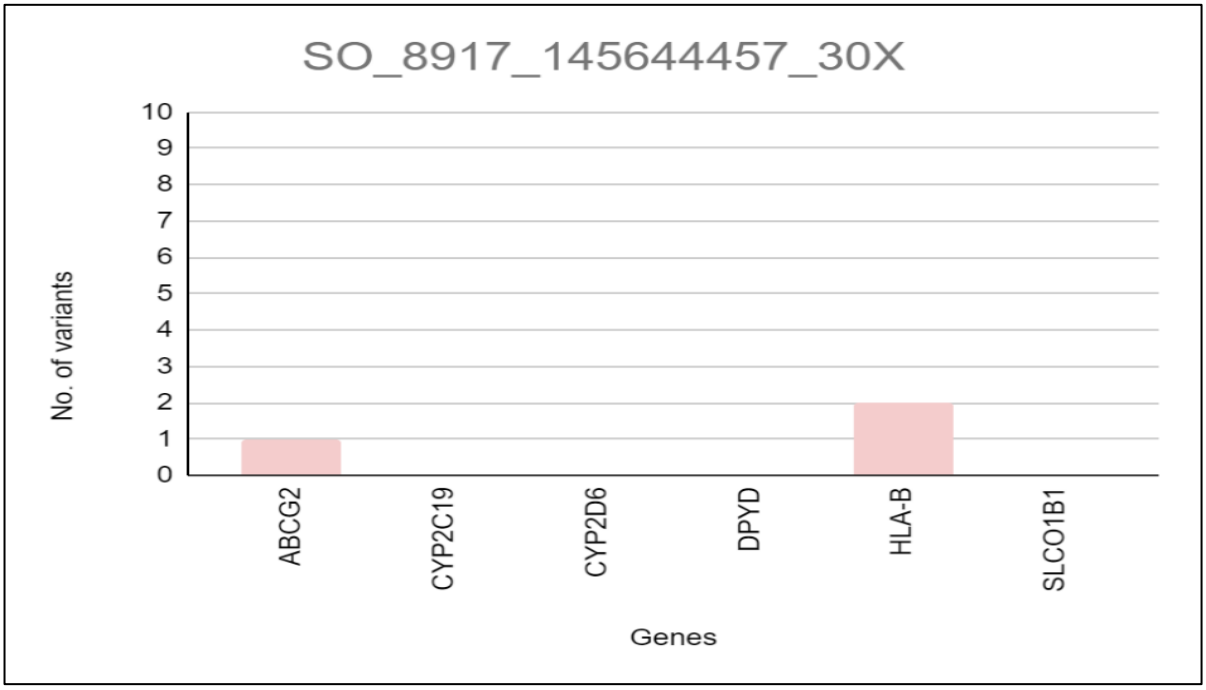

Figure 6 – Column chart showing variants observed in SO_8917_145644457_30X

Individual No. 2 has the following variants present:

c.1147C>T p.Arg383Cys is a ABCG2 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.061. Polyphen score is probably_damaging 0.999 and SIFT score is deleterious 0. Variant is G > A and the amino acid change is R383C – Cysteine (C) replaces Arginine (R) at codon 383. The maximum allelic frequency of this variant is 0.001.

c.368A>G p.Tyr123Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of probably_damaging 0.924 and SIFT score of deleterious_low_confidence 0.05. Variant is T > C and the amino acid change is Y123C – Cysteine (C) replaces Tyrosine (Y) at codon 123. The maximum allelic frequency of this variant is < 0.0001.

c.361A>T p.Ser121Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of possibly_damaging 0.876 and SIFT score of tolerated_low_confidence 0.17. Variant is T > A and the amino acid change is S121C – Cysteine (C) replaces Serine (S) at codon 121. The maximum allelic frequency of this variant is 0.0227.

Individual No. 3 – SO_8917_145644466_30X

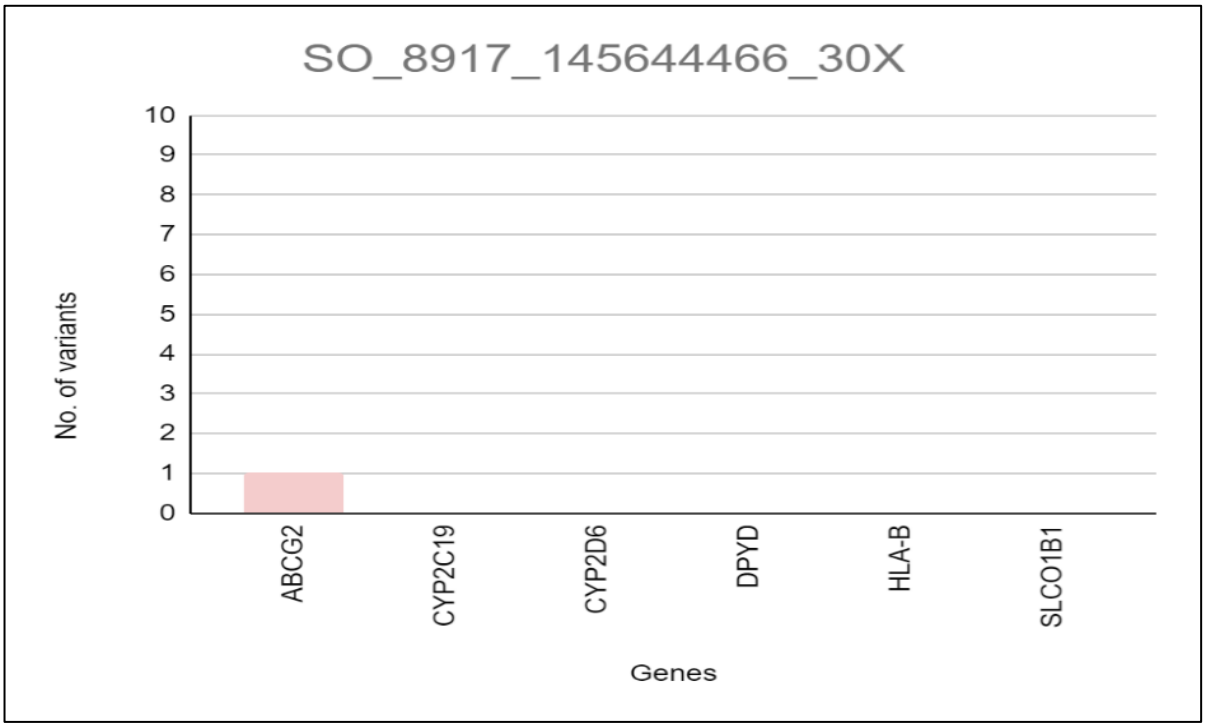

Figure 7 – Column chart showing variants observed in SO_8917_145644466_30X

Individual No. 3 has the following variant present:

c.1147C>T p.Arg383Cys is a ABCG2 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.061. Polyphen score is probably_damaging 0.999 and SIFT score is deleterious 0. Variant is G > A and the amino acid change is R383C – Cysteine (C) replaces Arginine (R) at codon 383. The maximum allelic frequency of this variant is 0.001.

Individual No. 4 – 14810642_30X

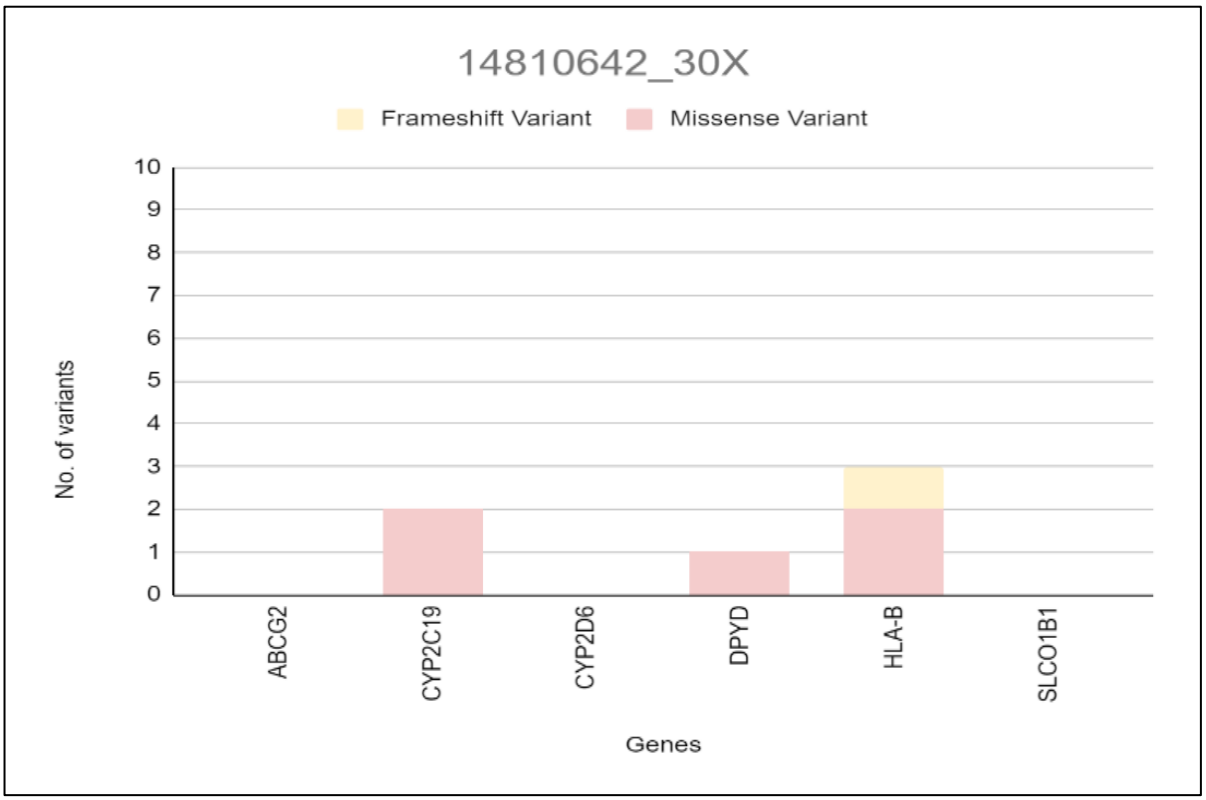

Figure 8 – Column chart showing variants observed in 14810642_30X

Individual No. 4 has the following variants present:

c.556C>T p.Arg186Cys is a CYP2C19 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.044. Polyphen score is possibly_damaging 0.553 and SIFT score is deleterious 0.02. Variant is C > T and the amino acid change is R186C – Cysteine (C) replaces Arginine (R) at codon 186. The maximum allelic frequency of this variant is 0.0014.

c.985C>T p.Arg329Cys is a CYP2C19 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.044. Polyphen score is benign 0.034 and SIFT score is tolerated 0.09. Variant is C > T and the amino acid change is R329C – Cysteine (C) replaces Arginine (R) at codon 329. The maximum allelic frequency of this variant is 0.00166.

c.2194G>A p.Val732Ile is a DPYD gene missense variant. It has a heterozygous genotype with a gene score of HI 0.272. Polyphen score is possibly_damaging 0.901 and SIFT score is tolerated 0.1. Variant is C > T and the amino acid change is V732I – Isoleucine (I) replaces Valine (V) at codon 732. The maximum allelic frequency of this variant is 0.0975.

c.368A>G p.Tyr123Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of probably_damaging 0.924 and SIFT score of deleterious_low_confidence 0.05. Variant is T > C and the amino acid change is Y123C – Cysteine (C) replaces Tyrosine (Y) at codon 123. The maximum allelic frequency of this variant is < 0.0001.

c.361A>T p.Ser121Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of possibly_damaging 0.876 and SIFT score of tolerated_low_confidence 0.17. Variant is T > A and the amino acid change is S121C – Cysteine (C) replaces Serine (S) at codon 121. The maximum allelic frequency of this variant is 0.0227.

c.354_355delCC p.Leu119ProfsTer19 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is AGG > A and the amino acid change is L119X – Unknown amino acid (X) replaces Leucine (L) at codon 119. The maximum allelic frequency of this variant is 0.

Individual No. 5 – 14810643_30X

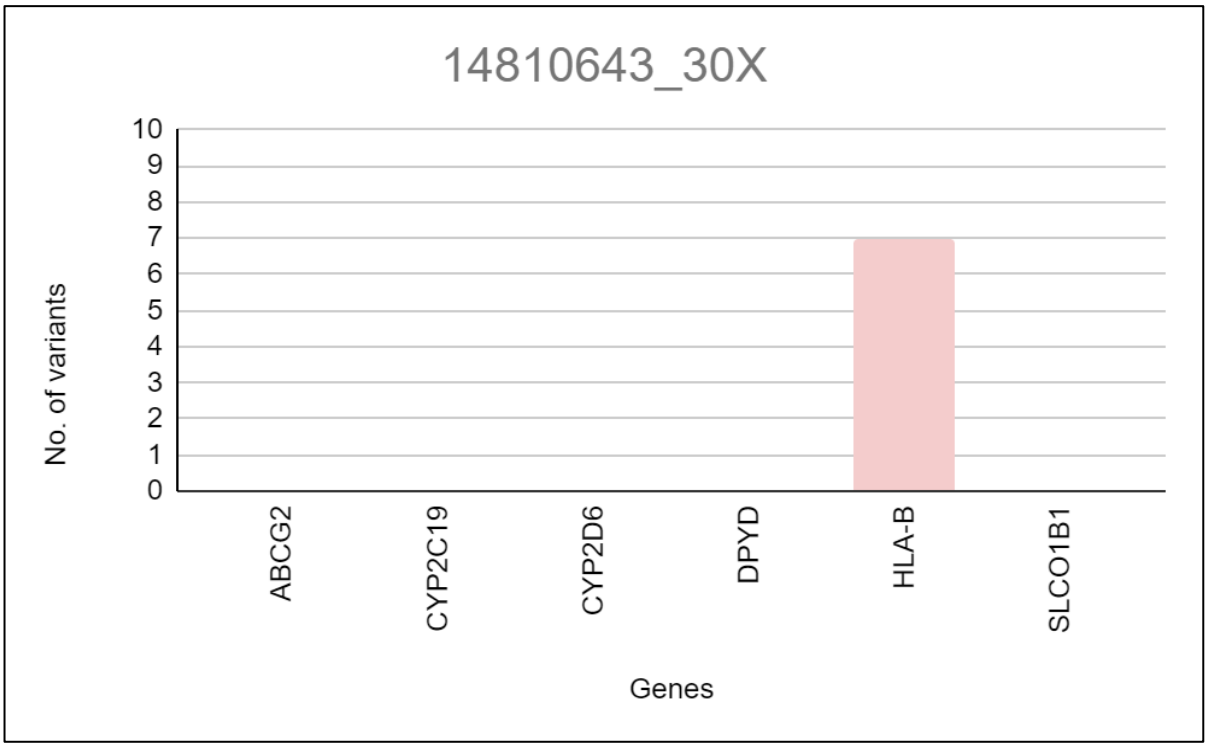

Figure 9 – Column chart showing variants observed in 14810643_30X

Individual No. 5 has the following variants present:

c.239G>A p.Gly80Glu is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of probably_damaging 0.948 and SIFT score of tolerated_low_confidence 0.08. Variant is C > T and the amino acid change is G87E – Glutamic acid (E) replaces Glycine (G) at codon 80. The maximum allelic frequency of this variant is 0.

c.259A>C p.Asn87His is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of possibly_damaging 0.524 and SIFT score of deleterious_low_confidence 0.01. Variant is T > G and the amino acid change is N87H – Histidine (H) replaces Asparagine (N) at codon 87. The maximum allelic frequency of this variant is 0.

c.265C>G p.Gln89Glu is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.183 and SIFT score of deleterious_low_confidence 0.03. Variant is G > C and the amino acid change is Q89E – Glutamic acid (E) replaces Glutamine (Q) at codon 89. The maximum allelic frequency of this variant is 0.

c.290C>T p.Thr97Ile is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.343 and SIFT score of tolerated_low_confidence 0.17. Variant is G > A and the amino acid change is T97I – Isoleucine (I) replaces Threonine (T) at codon 97. The maximum allelic frequency of this variant is 0.

c.296G>T p.Arg99Leu is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.439 and SIFT score of deleterious_low_confidence 0.02. Variant is C > A and the amino acid change is R99L – Leucine (L) replaces Arginine (R) at codon 99. The maximum allelic frequency of this variant is 0.

c.341C>A p.Ala114Asp is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.001 and SIFT score of tolerated_low_confidence 0.25. Variant is G > T and the amino acid change is A114D – Aspartic acid (D) replaces Alanine (A) at codon 114. The maximum allelic frequency of this variant is 0.0015.

c.397C>T p.Leu133Phe is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.02 and SIFT score of tolerated_low_confidence 0.18. Variant is G > A and the amino acid change is L133F – Phenylalanine (F) replaces Leucine (L) at codon 133. The maximum allelic frequency of this variant is 0.0144.

Individual No. 6 – 14810644_452

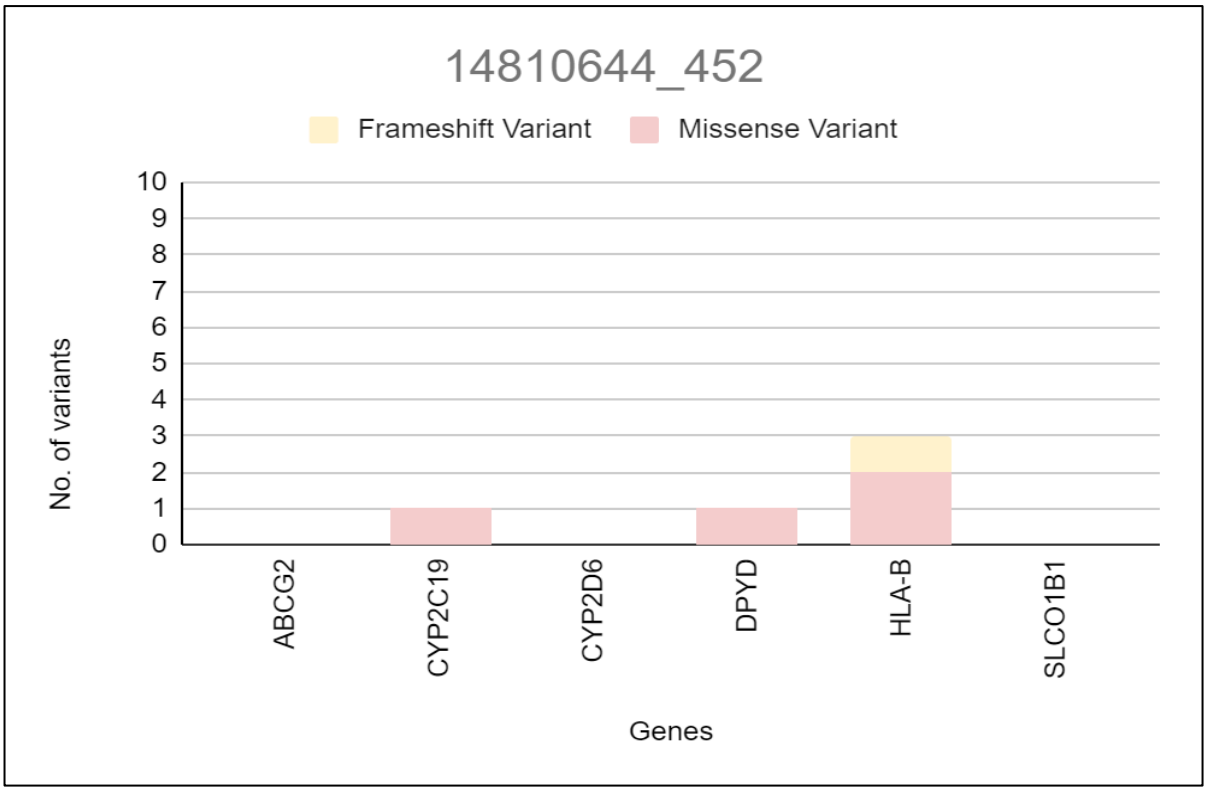

Figure 10 – Column chart showing variants observed in 14810644_452

Individual No. 6 has the following variants present:

c.556C>T p.Arg186Cys is a CYP2C19 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.044. Polyphen score is possibly_damaging 0.553 and SIFT score is deleterious 0.02. Variant is C > T and the amino acid change is R186C – Cysteine (C) replaces Arginine (R) at codon 186. The maximum allelic frequency of this variant is 0.0014.

c.2194G>A p.Val732Ile is a DPYD gene missense variant. It has a heterozygous genotype with a gene score of HI 0.272. Polyphen score is possibly_damaging 0.901 and SIFT score is tolerated 0.1. Variant is C > T and the amino acid change is V732I – Isoleucine (I) replaces Valine (V) at codon 732. The maximum allelic frequency of this variant is 0.0975.

c.368A>G p.Tyr123Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of probably_damaging 0.924 and SIFT score of deleterious_low_confidence 0.05. Variant is T > C and the amino acid change is Y123C – Cysteine (C) replaces Tyrosine (Y) at codon 123. The maximum allelic frequency of this variant is < 0.0001.

c.361A>T p.Ser121Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of possibly_damaging 0.876 and SIFT score of tolerated_low_confidence 0.17. Variant is T > A and the amino acid change is S121C – Cysteine (C) replaces Serine (S) at codon 121. The maximum allelic frequency of this variant is 0.0227.

c.354_355delCC p.Leu119ProfsTer19 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is AGG > A and the amino acid change is L119X – Unknown amino acid (X) replaces Leucine (L) at codon 119. The maximum allelic frequency of this variant is 0.

Individual No. 7 – 14810648_30X

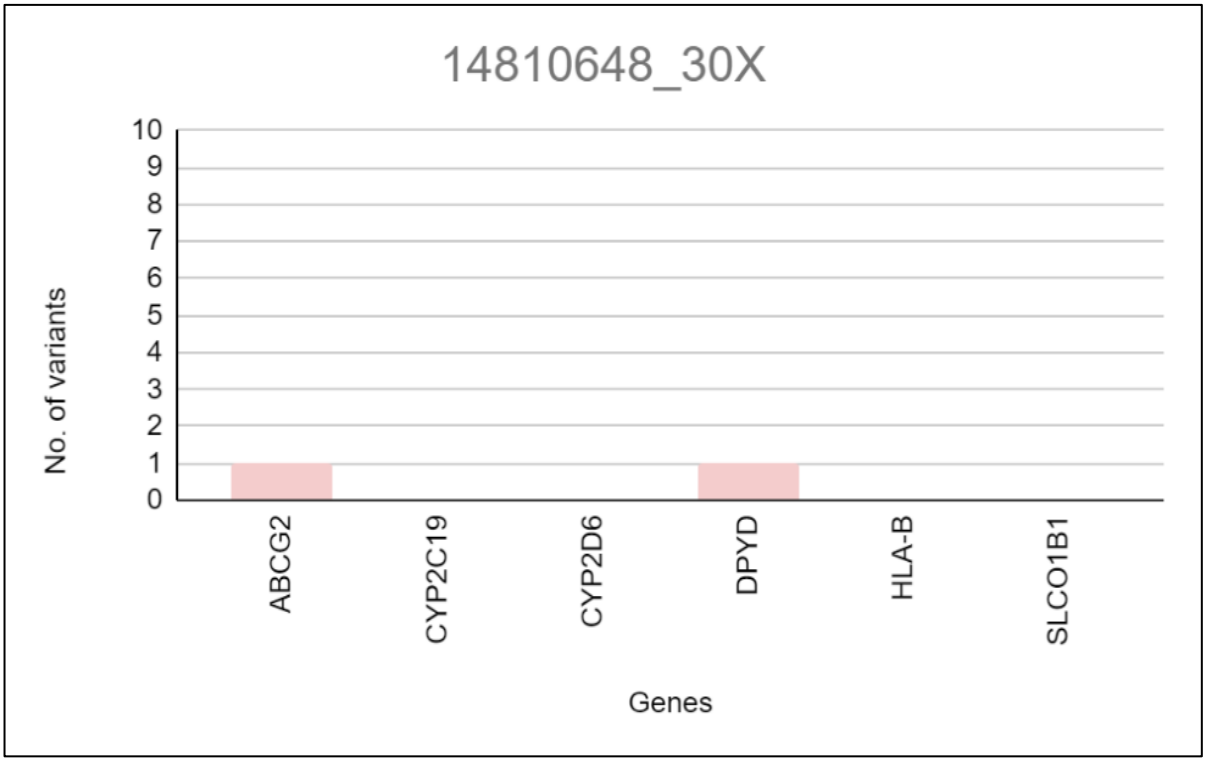

Figure 11 – Column chart showing variants observed in 14810648_30X

Individual No. 7 has the following variants present:

c.1147C>T p.Arg383Cys is a ABCG2 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.061. Polyphen score is probably_damaging 0.999 and SIFT score is deleterious 0. Variant is G > A and the amino acid change is R383C – Cysteine (C) replaces Arginine (R) at codon 383. The maximum allelic frequency of this variant is 0.001.

c.2194G>A p.Val732Ile is a DPYD gene missense variant. It has a heterozygous genotype with a gene score of HI 0.272. Polyphen score is possibly_damaging 0.901 and SIFT score is tolerated 0.1. Variant is C > T and the amino acid change is V732I – Isoleucine (I) replaces Valine (V) at codon 732. The maximum allelic frequency of this variant is 0.0975.

Individual No. 8 – 14810649_30X

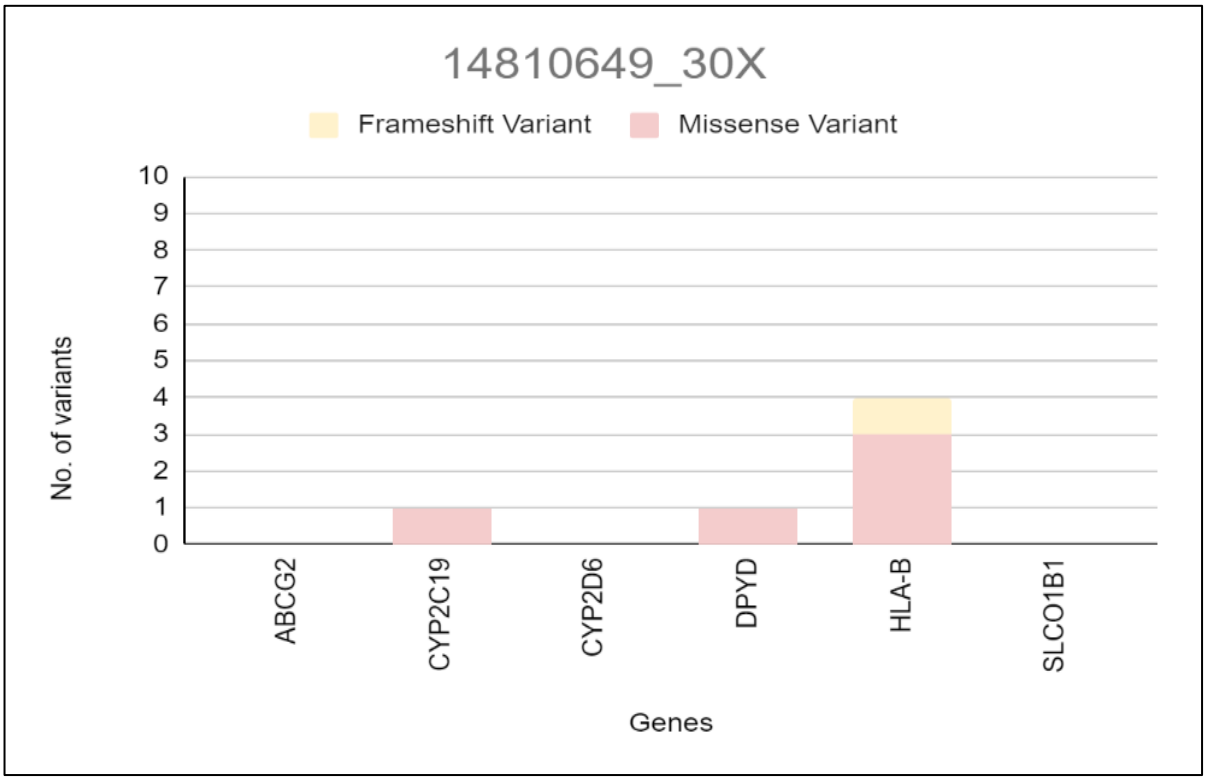

Figure 12 – Column chart showing variants observed in 14810649_30X

Individual No. 8 has the following variants present:

c.985C>T p.Arg329Cys is a CYP2C19 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.044. Polyphen score is benign 0.034 and SIFT score is tolerated 0.09. Variant is C > T and the amino acid change is R329C – Cysteine (C) replaces Arginine (R) at codon 329. The maximum allelic frequency of this variant is 0.00166.

c.2194G>A p.Val732Ile is a DPYD gene missense variant. It has a heterozygous genotype with a gene score of HI 0.272. Polyphen score is possibly_damaging 0.901 and SIFT score is tolerated 0.1. Variant is C > T and the amino acid change is V732I – Isoleucine (I) replaces Valine (V) at codon 732. The maximum allelic frequency of this variant is 0.0975.

c.397C>T p.Leu133Phe is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.02 and SIFT score of tolerated_low_confidence 0.18. Variant is G > A and the amino acid change is L133F – Phenylalanine (F) replaces Leucine (L) at codon 133. The maximum allelic frequency of this variant is 0.0144.

c.368A>G p.Tyr123Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of probably_damaging 0.924 and SIFT score of deleterious_low_confidence 0.05. Variant is T > C and the amino acid change is Y123C – Cysteine (C) replaces Tyrosine (Y) at codon 123. The maximum allelic frequency of this variant is < 0.0001.

c.361A>T p.Ser121Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of possibly_damaging 0.876 and SIFT score of tolerated_low_confidence 0.17. Variant is T > A and the amino acid change is S121C – Cysteine (C) replaces Serine (S) at codon 121. The maximum allelic frequency of this variant is 0.0227.

c.354_355delCC p.Leu119ProfsTer19 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is AGG > A and the amino acid change is L119X – Unknown amino acid (X) replaces Leucine (L) at codon 119. The maximum allelic frequency of this variant is 0.

Individual No. 9 – 14810662

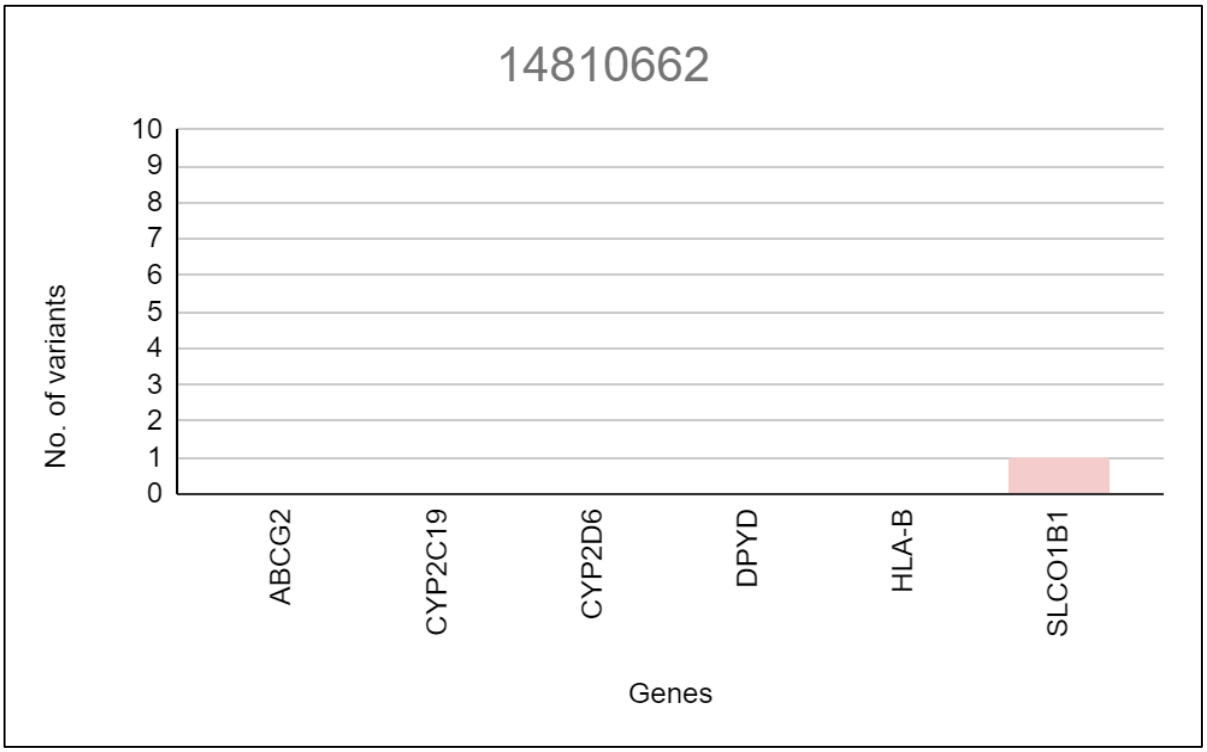

Figure 13 – Column chart showing variants observed in 14810662

Individual No. 9 has the following variant present:

c.1929A>C p.Leu643Phe is a SLCO1B1 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.055. Polyphen score is benign 0.005 and SIFT score is tolerated 1. Variant is A > C and the amino acid change is L643F – Phenylalanine (F) replaces Leucine (L) at codon 643. The maximum allelic frequency of this variant is 0.0654.

Individual No. 10 – 14810694

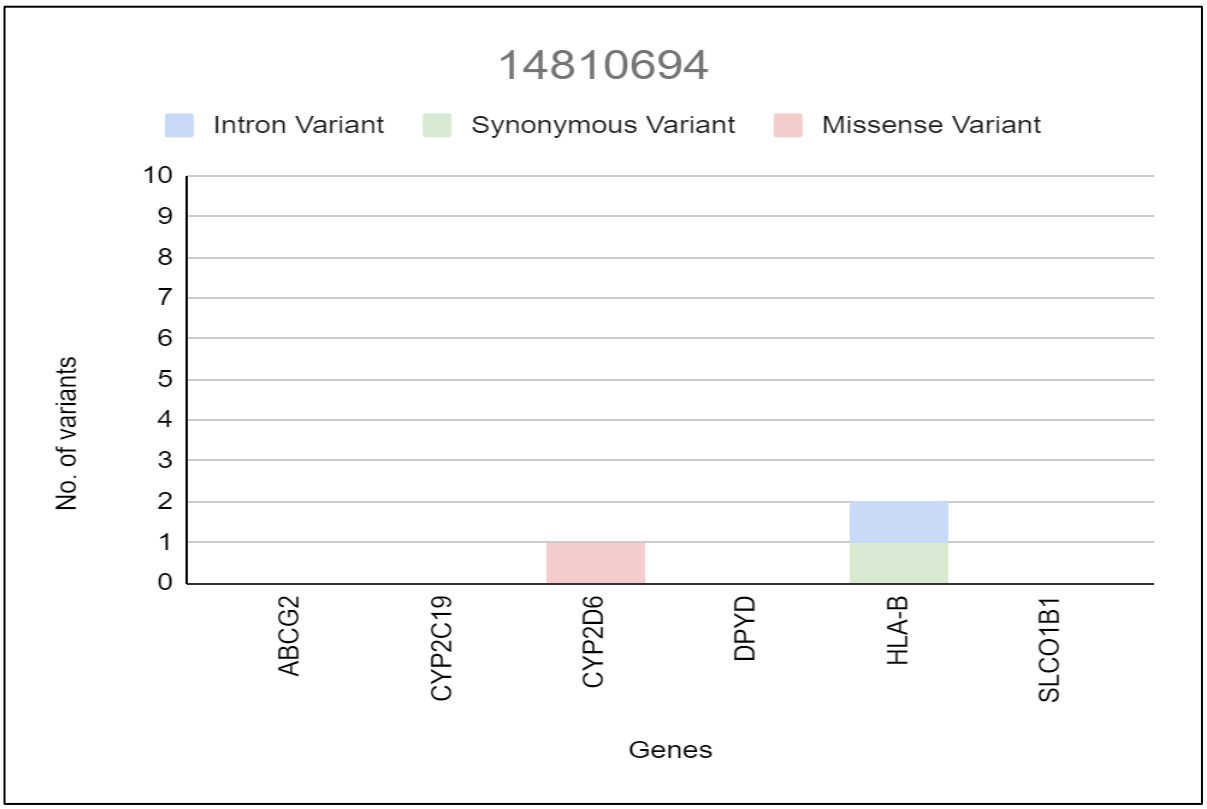

Figure 14 – Column chart showing variants observed in 14810694

Individual No. 10 has the following variants present:

c.31G>A p.Val11Met is a CYP2D6 gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.024 and SIFT score of tolerated 0.14. Variant is C > T and the amino acid change is V11M – Methionine (M) replaces Valine (V) at codon 11. The maximum allelic frequency of this variant is 0.0563.

c.141C>T p.(Ile47=) is a HLA-B gene synonymous variant. It has a heterozygous genotype. Variant is G > A and the amino acid change is at I – Isoleucine at codon 47. The maximum allelic frequency of this variant is 0.0978.

c.74-22C>T is a HLA-B gene intron variant. It has a heterozygous genotype. Variant is G > A and the maximum allelic frequency of this variant is 0.0605.

Individual No. 11 – 14810704

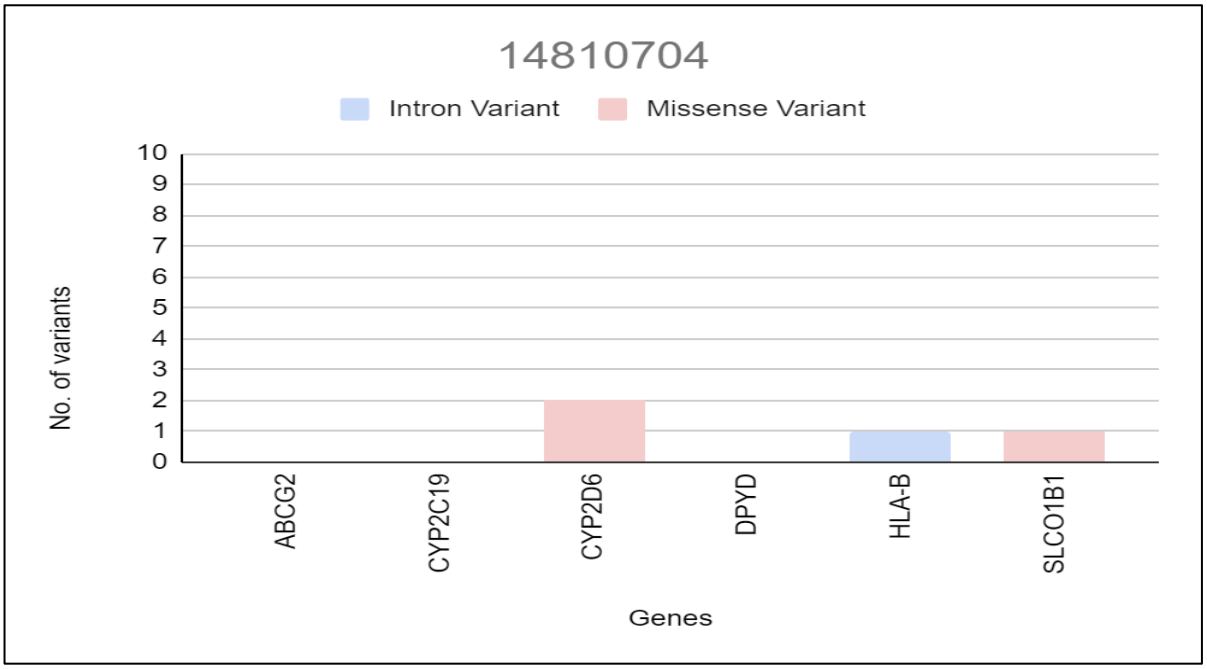

Figure 15 – Column chart showing variants observed in 14810704

Individual No. 11 has the following variants present:

c.1108G>A p.Val370Ile is a CYP2D6 gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.001and SIFT score of tolerated 0.87. Variant is C > T and the amino acid change is V370I – Isoleucine (I) replaces Valine (V) at codon 370. The maximum allelic frequency of this variant is 0.00229.

c.1117G>A p.Gly373Ser is a CYP2D6 gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.074 and SIFT score of tolerated 0.23. Variant is C > T and the amino acid change is G373S – Serine (S) replaces Glycine (G) at codon 373. The maximum allelic frequency of this variant is 0.00417.

c.74-22C>T is a HLA-B gene intron variant. It has a heterozygous genotype. Variant is G > A and the maximum allelic frequency of this variant is 0.0605.

c.1929A>C p.Leu643Phe is a SLCO1B1 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.055. Polyphen score is benign 0.005 and SIFT score is tolerated 1. Variant is A > C and the amino acid change is L643F – Phenylalanine (F) replaces Leucine (L) at codon 643. The maximum allelic frequency of this variant is 0.0654.

Individual No. 12 – 14810707

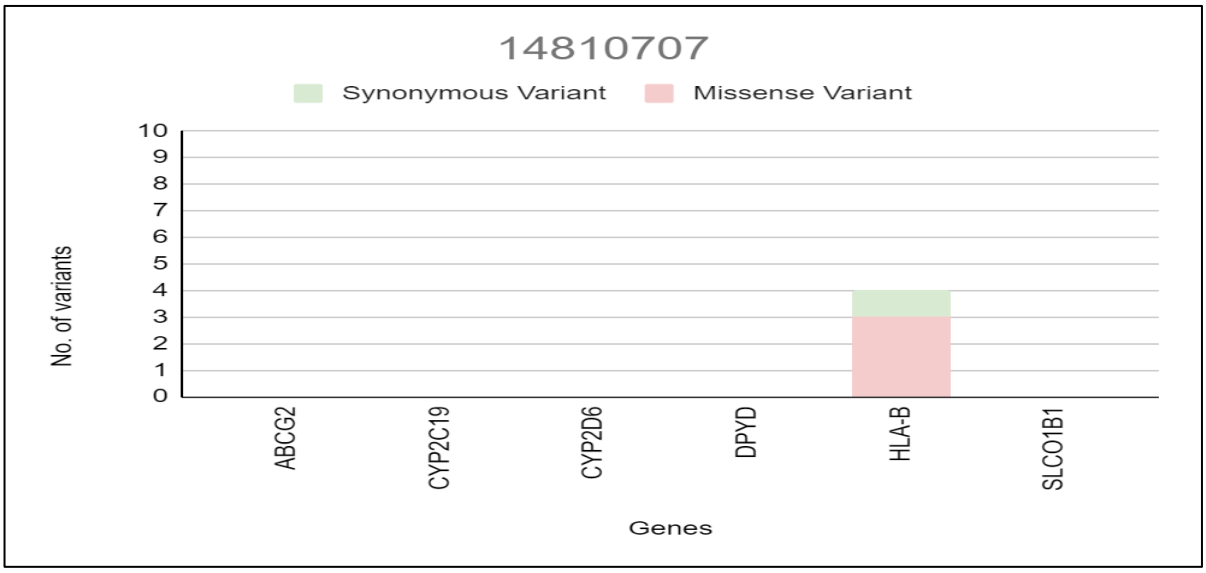

Figure 16 – Column chart showing variants observed in 14810707

Individual No. 12 has the following variants present:

c.233C>T p.Ala78Val is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.046 and SIFT score of deleterious 0.029. Variant is G > A and the amino acid change is A78V – Valine (V) replaces Alanine (A) at codon 78. The maximum allelic frequency of this variant is 0.0665.

c.412G>C p.Asp138His is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.025 and SIFT score of tolerated_low_confidence 0.569. Variant is C > G and the amino acid change is D138H – Histidine (H) replaces Aspartic acid (D) at codon 138. The maximum allelic frequency of this variant is 0.0547.

c.991A>G p.Met331Val is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0 and SIFT score of tolerated_low_confidence 0.09. Variant is T > C and the amino acid change is M331V – Valine (V) replaces Methionine (M) at codon 331. The maximum allelic frequency of this variant is 0.0043.

c.411T>C p.(His137=) is a HLA-B gene synonymous variant. It has a homozygous genotype. Variant is A > G and the amino acid change is at H – Histidine at codon 137. The maximum allelic frequency of this variant is 0.0547.

Individual No. 13 – 14810709

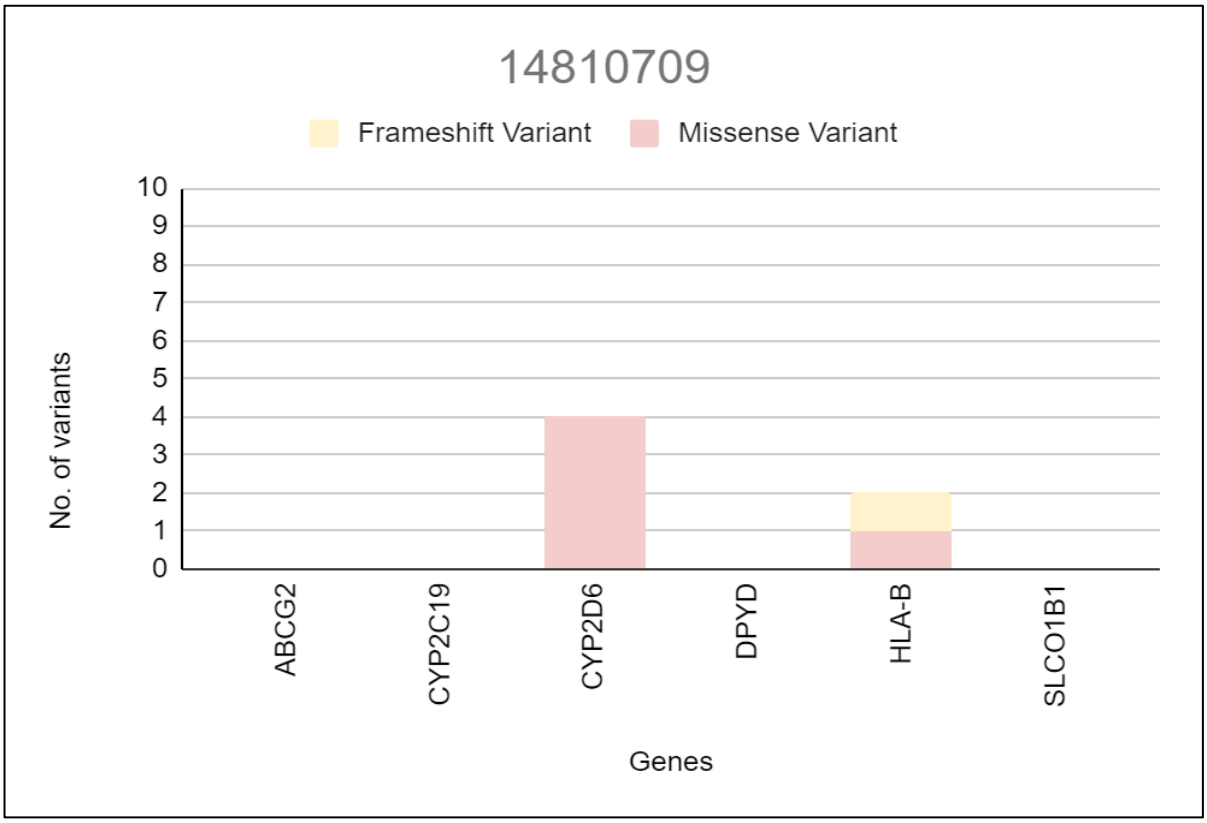

Figure 17 – Column chart showing variants observed in 14810709

Individual No. 13 has the following variants present:

c.19G>A p.Val7Met is a CYP2D6 gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.001 and SIFT score of tolerated 0.22. Variant is C > T and the amino acid change is V7M – Methionine (M) replaces Valine (V) at codon 7. The maximum allelic frequency of this variant is 0.00542.

c.451C>G p.Gln151Glu is a CYP2D6 gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.003 and SIFT score of tolerated 0.88. Variant is G > C and the amino acid change is Q151E – Glutamic acid (E) replaces Glutamine (Q) at codon 151. The maximum allelic frequency of this variant is 0.00526.

c.1108G>A p.Val370Ile is a CYP2D6 gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.001 and SIFT score of tolerated 0.87. Variant is C > T and the amino acid change is V370I – Isoleucine (I) replaces Valine (V) at codon 370. The maximum allelic frequency of this variant is 0.00244.

c.1117G>A p.Gly373Ser is a CYP2D6 gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.074 and SIFT score of tolerated 0.23. Variant is C > T and the amino acid change is G373S – Serine (S) replaces Glycine (G) at codon 373. The maximum allelic frequency of this variant is 0.00402.

c.397C>T p.Leu133Phe is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.02 and SIFT score of tolerated_low_confidence 0.18. Variant is G > A and the amino acid change is L133F – Phenylalanine (F) replaces Leucine (L) at codon 133. The maximum allelic frequency of this variant is 0.0144.

c.354_355delCC p.Leu119ProfsTer19 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is AGG > A and the amino acid change is L119X – Unknown amino acid (X) replaces Leucine (L) at codon 119. The maximum allelic frequency of this variant is 0.

Individual No. 14 – 14810712

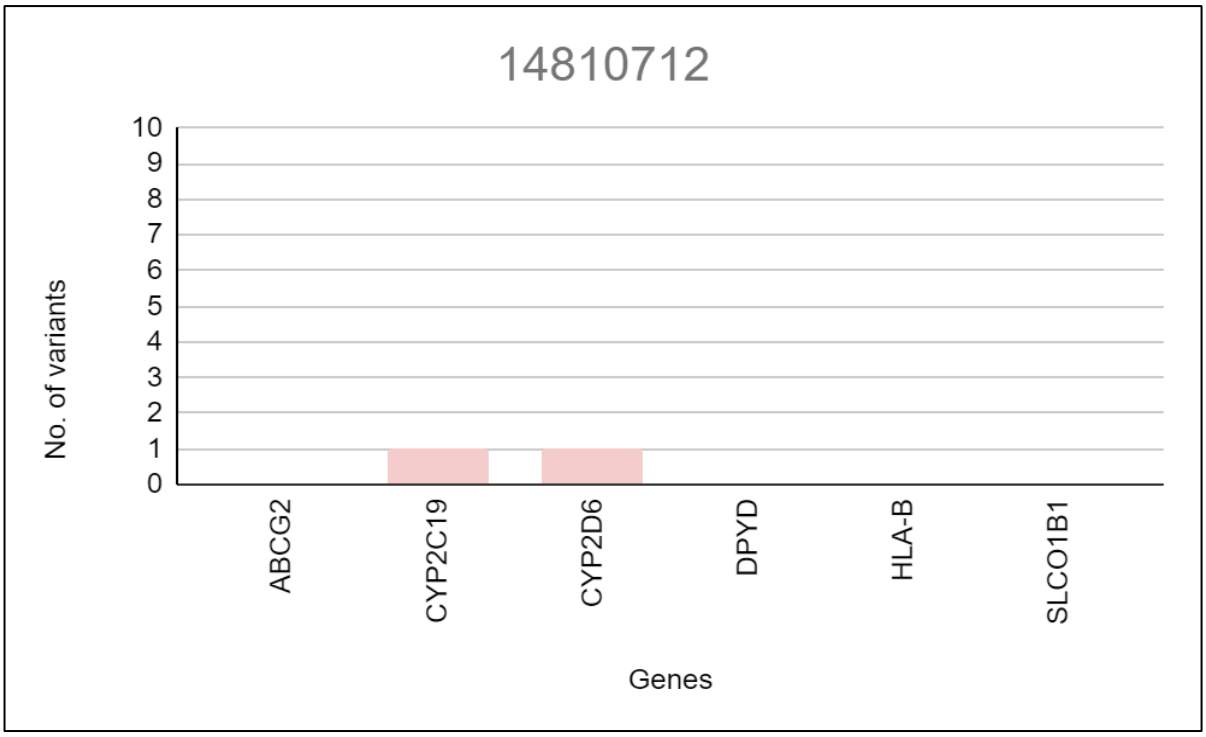

Figure 18 – Column chart showing variants observed in 14810712

Individual No. 14 has the following variants present:

c.518C>T p.Ala173Val is a CYP2C19 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.044. Polyphen score is possibly_damaging 0.518 and SIFT score is tolerated 0.1. Variant is C > T and the amino acid change is A173V – Valine (V) replaces Alanine (A) at codon 173. The maximum allelic frequency of this variant is 0.0429.

c.1463C>T p.Ser488Phe is a CYP2D6 gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.0649 and SIFT score of tolerated 0.69. Variant is G > A and the amino acid change is S488F – Valine (V) replaces Alanine (A) at codon 173. The maximum allelic frequency of this variant is 0.00314.

Individual No. 15 – 14810715

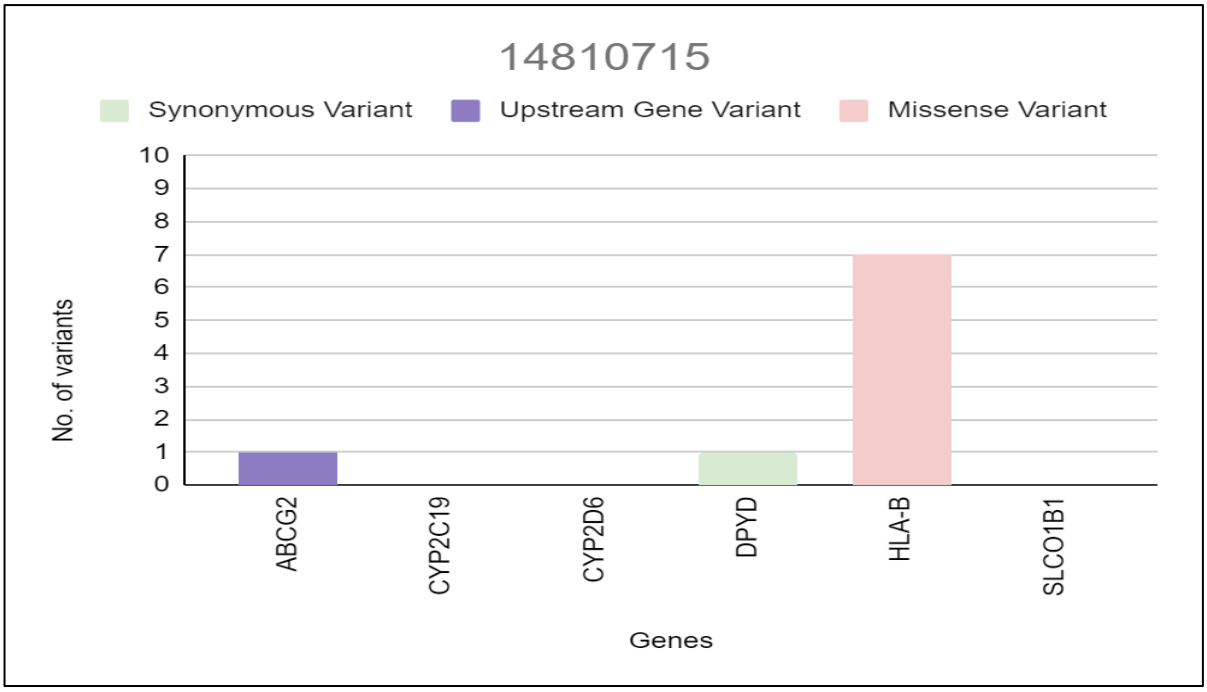

Figure 19 – Column chart showing variants observed in 14810715

Individual No. 15 has the following variants present:

An **Upstream Gene Variant** is present in the ABCG2 gene of this individual. It has a heterozygous genotype with a gene score of HI 0.061. The variant is C > G and the maximum allelic frequency is 0.0414.

c.672T>C p.(=) is a DPYD gene synonymous variant. It has a heterozygous genotype with a gene score of HI 0.272. Variant is A > G and the amino acid change is at G – Glycine. The maximum allelic frequency of this variant is < 0.0001.

c.228A>G p.Ile76Met is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.039 and SIFT score of tolerated_low_confidence 0.97. Variant is T > C and the amino acid change is I76M – Methionine (M) replaces Isoleucine (I) at codon 76. The maximum allelic frequency of this variant is 0.068.

c.239G>A p.Gly80Glu is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of probably_damaging 0.948 and SIFT score of tolerated_low_confidence 0.08. Variant is C > T and the amino acid change is G87E – Glutamic acid (E) replaces Glycine (G) at codon 80. The maximum allelic frequency of this variant is 0.

c.259A>C p.Asn87His is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of possibly_damaging 0.524 and SIFT score of deleterious_low_confidence 0.01. Variant is T > G and the amino acid change is N87H – Histidine (H) replaces Asparagine (N) at codon 87. The maximum allelic frequency of this variant is 0.

c.265C>G p.Gln89Glu is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.183 and SIFT score of deleterious_low_confidence 0.03. Variant is G > C and the amino acid change is Q89E – Glutamic acid (E) replaces Glutamine (Q) at codon 89. The maximum allelic frequency of this variant is 0.

c.290C>T p.Thr97Ile is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.343 and SIFT score of tolerated_low_confidence 0.17. Variant is G > A and the amino acid change is T97I – Isoleucine (I) replaces Threonine (T) at codon 97. The maximum allelic frequency of this variant is 0.

c.296G>T p.Arg99Leu is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.439 and SIFT score of deleterious_low_confidence 0.02. Variant is C > A and the amino acid change is R99L – Leucine (L) replaces Arginine (R) at codon 99. The maximum allelic frequency of this variant is 0.

c.341C>A p.Ala114Asp is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.001 and SIFT score of tolerated_low_confidence 0.25. Variant is G > T and the amino acid change is A114D – Aspartic acid (D) replaces Alanine (A) at codon 114. The maximum allelic frequency of this variant is 0.0015.

Individual No. 16 – 14810717

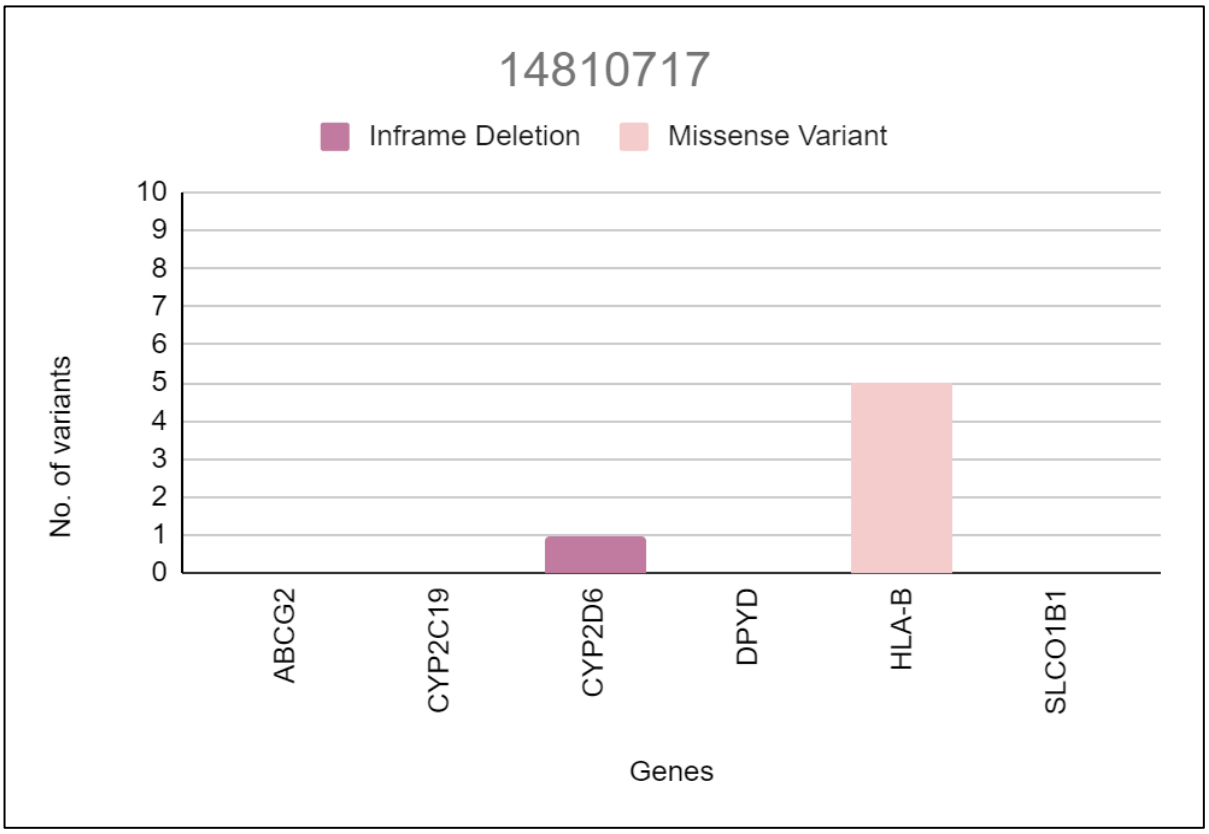

Figure 20 – Column chart showing variants observed in 14810717

Individual No. 16 has the following variants present:

c.841_843delAAG p.Lys281del is a CYP2D6 gene inframe deletion. It has a heterozygous genotype. Variant is CCTT > C and the amino acid change is K281del – Lysine (K) at codon 281. The maximum allelic frequency of this variant is 0.0258.

c.228A>G p.Ile76Met is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.039 and SIFT score of tolerated_low_confidence 0.97. Variant is T > C and the amino acid change is I76M – Methionine (M) replaces Isoleucine (I) at codon 76. The maximum allelic frequency of this variant is 0.068.

c.239G>A p.Gly80Glu is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of probably_damaging 0.948 and SIFT score of tolerated_low_confidence 0.08. Variant is C > T and the amino acid change is G87E – Glutamic acid (E) replaces Glycine (G) at codon 80. The maximum allelic frequency of this variant is 0.

c.265C>G p.Gln89Glu is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.183 and SIFT score of deleterious_low_confidence 0.03. Variant is G > C and the amino acid change is Q89E – Glutamic acid (E) replaces Glutamine (Q) at codon 89. The maximum allelic frequency of this variant is 0.

c.290C>T p.Thr97Ile is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.343 and SIFT score of tolerated_low_confidence 0.17. Variant is G > A and the amino acid change is T97I – Isoleucine (I) replaces Threonine (T) at codon 97. The maximum allelic frequency of this variant is 0.

c.296G>T p.Arg99Leu is a HLA-B gene missense variant. It has a homozygous genotype with a Polyphen score of benign 0.439 and SIFT score of deleterious_low_confidence 0.02. Variant is C > A and the amino acid change is R99L – Leucine (L) replaces Arginine (R) at codon 99. The maximum allelic frequency of this variant is 0.

Individual No. 17 – 14810718

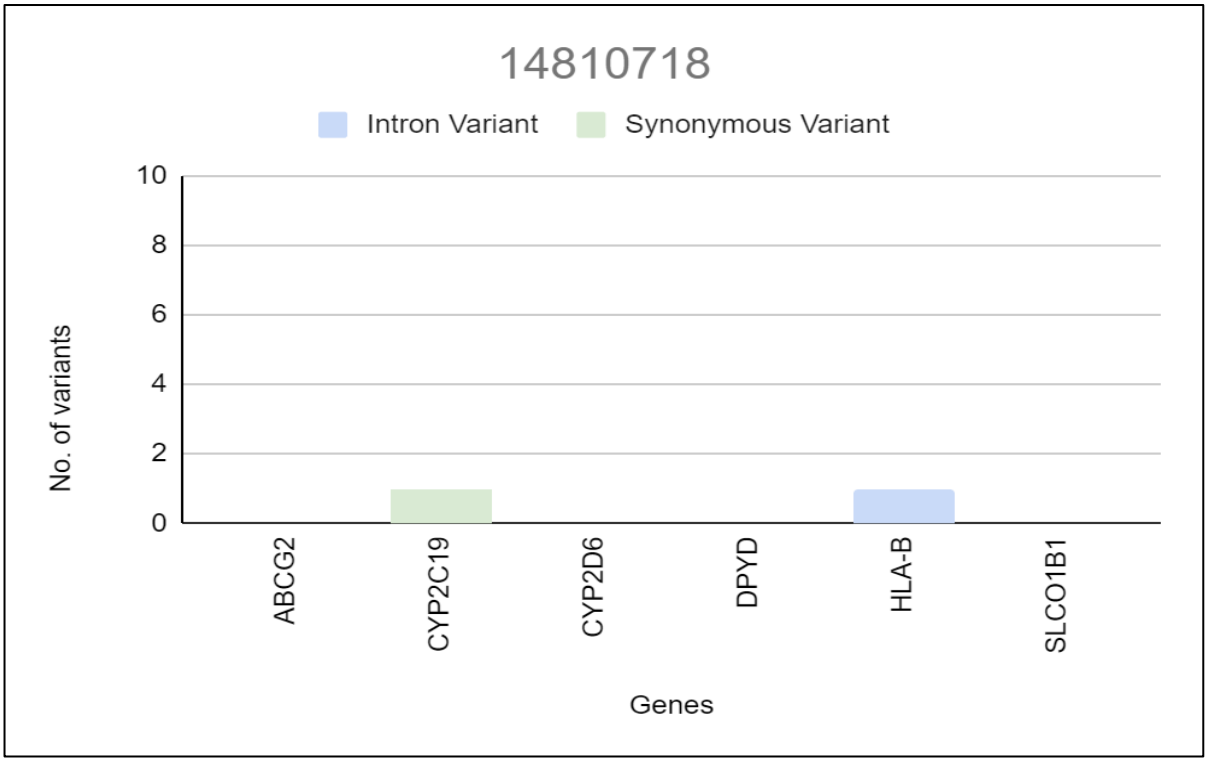

Figure 21 – Column chart showing variants observed in 14810718

Individual No. 17 has the following variants present:

c.1266T>C p.(=) is a CYP2C19 gene synonymous variant. It has a heterozygous genotype with a gene score of HI 0.044. Variant is T > C and the amino acid change is at S – Serine. The maximum allelic frequency of this variant is 0.

c.74-22C>T is a HLA-B gene intron variant. It has a homozygous genotype. Variant is G > A and the maximum allelic frequency of this variant is 0.0605.

Individual No. 18 – 14810719

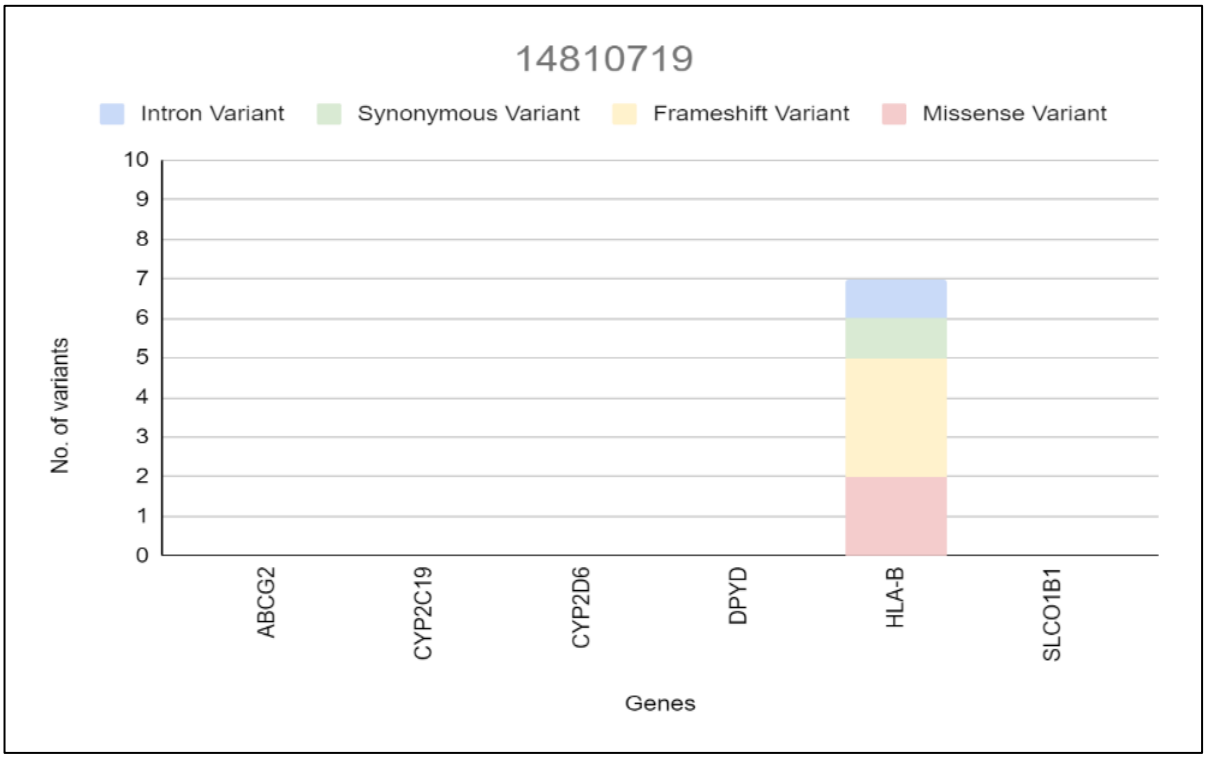

Figure 22 – Column chart showing variants observed in 14810719

Individual No. 18 has the following variants present:

c.74-9>T is a HLA-B gene intron variant. It has a heterozygous genotype. Variant is G > C and the maximum allelic frequency of this variant is 0.0288.

c.97T>G p.Tyr33Asp is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.005 and SIFT score of tolerated_low_confidence 0.21. Variant is A > C and the amino acid change is Y33D – Aspartic acid (D) replaces Tyrosine (Y) at codon 33. The maximum allelic frequency of this variant is 0.0775.

c.309G>C p.(Arg103=) is a HLA-B gene synonymous variant. It has a heterozygous genotype. Variant is C > G and the amino acid change is at R – Arginine at codon 103. The maximum allelic frequency of this variant is 0.0825.

c.311delA p.Asn104ThrfsTer47 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is GT > G and the amino acid change is N104X – Unknown amino acid (X) replaces Asparagine (N) at codon 104. The maximum allelic frequency of this variant is 0.0838.

c.314delT p.Leu105ArgfsTer46 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is CA > C and the amino acid change is L105X – Unknown amino acid (X) replaces Leucine (L) at codon 105. The maximum allelic frequency of this variant is 0.0838.

c.319_320insCC p.Gly107AlafsTer45 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is C > CGG and the amino acid change is G107AX – Alanine and/or Unknown amino acid (X) replace Glycine (G) at codon 107. The maximum allelic frequency of this variant is 0.0646.

c.319G>T p.Gly107Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of possibly_damaging 0.869 and SIFT score of deleterious_low_confidence 0. Variant is C > A and the amino acid change is G107C – Cysteine (C) replaces Glycine (G) at codon 107. The maximum allelic frequency of this variant is 0.0874.

Individual No. 19 – 14810720

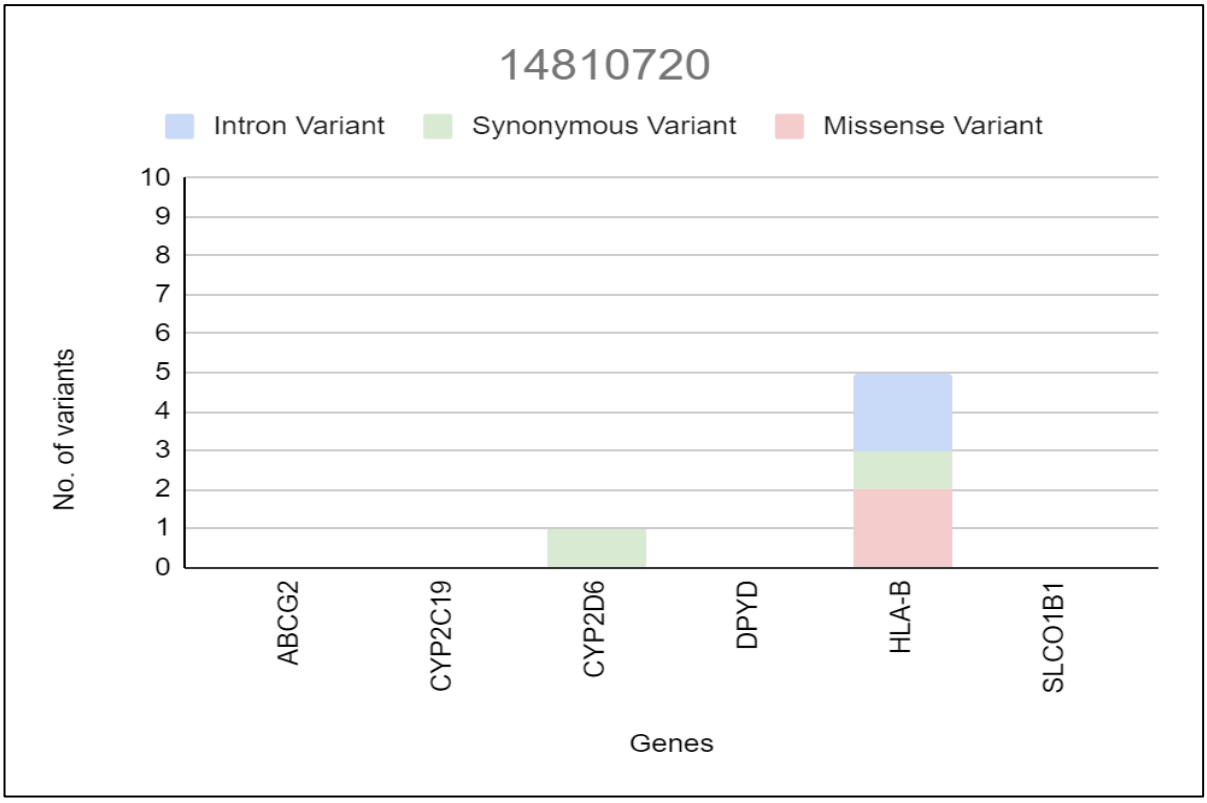

Figure 23 – Column chart showing variants observed in 14810720

Individual No. 19 has the following variants present:

c.1107C>T p.(=) is a CYP2D6 gene synonymous variant. It has a heterozygous genotype. Variant is G > A and the amino acid change is at I – Isoleucine. The maximum allelic frequency of this variant is 0.000194.

c.301A>G p.Ser101Gly is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0 and SIFT score of tolerated_low_confidence 0.23. Variant is T > C and the amino acid change is S101G – Glycine (G) replaces Serine (S) at codon 101. The maximum allelic frequency of this variant is 0.0586.

c.341C>A p.Ala114Asp is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.001 and SIFT score of tolerated_low_confidence 0.25. Variant is G > T and the amino acid change is A114D – Aspartic acid (D) replaces Alanine (A) at codon 114. The maximum allelic frequency of this variant is 0.0015.

c.117C>T p.(=) is a HLA-B gene synonymous variant. It has a heterozygous genotype. Variant is G > A and the amino acid change is at P – Proline. The maximum allelic frequency of this variant is 0.0015.

c.74-22C>T is a HLA-B gene intron variant. It has a heterozygous genotype. Variant is G > A and the maximum allelic frequency of this variant is 0.0605.

c.74-40G>T is a HLA-B gene intron variant. It has a heterozygous genotype. Variant is C > A and the maximum allelic frequency of this variant is 0.0015.

Individual No. 20 – 14810727

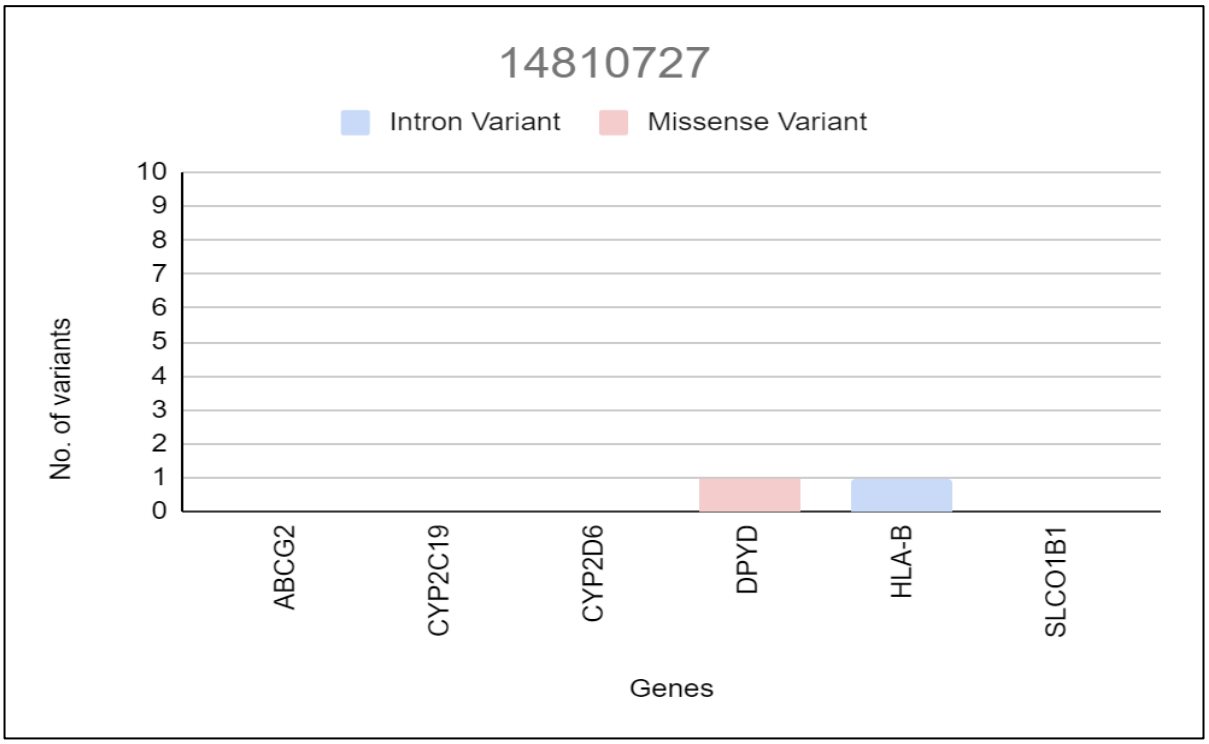

Figure 24 – Column chart showing variants observed in 14810727

Individual No. 20 has the following variants present:

c.2194G>A p.Val732Ile is a DPYD gene missense variant. It has a heterozygous genotype with a gene score of HI 0.272. Polyphen score is possibly_damaging 0.901 and SIFT score is tolerated 0.1. Variant is C > T and the amino acid change is V732I – Isoleucine (I) replaces Valine (V) at codon 732. The maximum allelic frequency of this variant is 0.0975.

c.74-22C>T is a HLA-B gene intron variant. It has a homozygous genotype. Variant is G > A and the maximum allelic frequency of this variant is 0.0605.

Individual No. 21 – 14811043

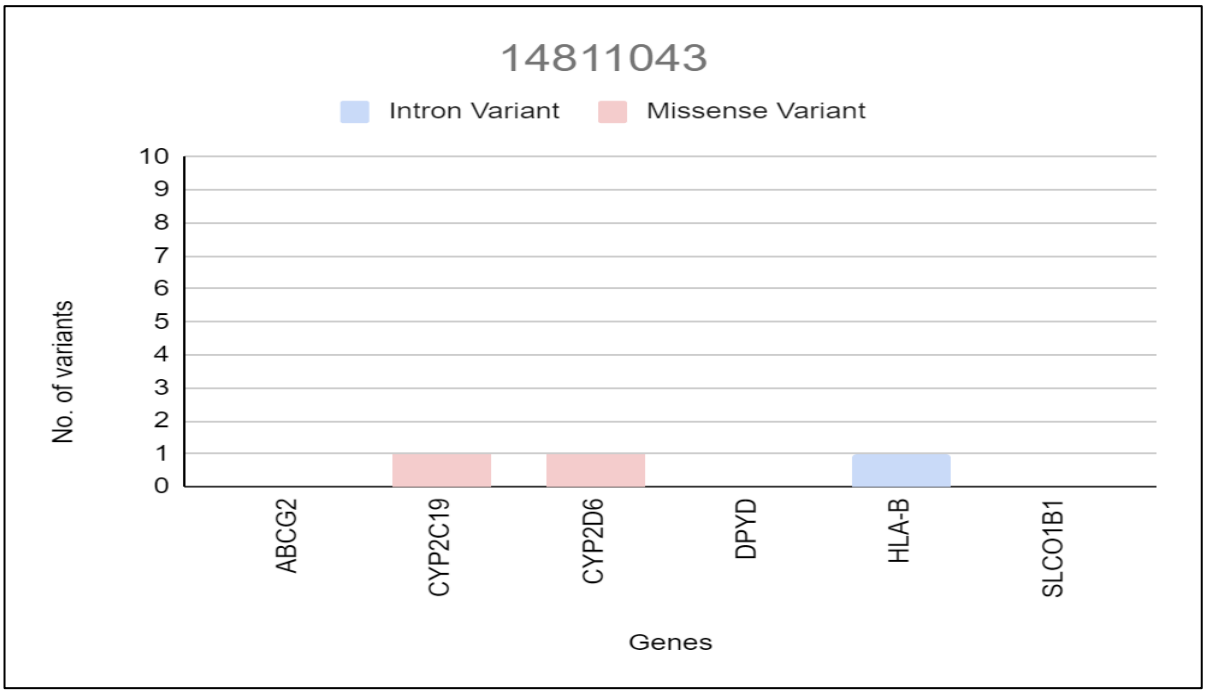

Figure 25 – Column chart showing variants observed in 14811043

Individual No. 21 has the following variants present:

c.518C>T p.Ala173Val is a CYP2C19 gene missense variant. It has a heterozygous genotype with a gene score of HI 0.044. Polyphen score is possibly_damaging 0.518 and SIFT score is tolerated 0.1. Variant is C > T and the amino acid change is A173V – Valine (V) replaces Alanine (A) at codon 173. The maximum allelic frequency of this variant is 0.0429.

c.1463C>T p.Ser488Phe is a CYP2D6 gene missense variant. It has a heterozygous genotype with a Polyphen score of benign 0.0649 and SIFT score of tolerated 0.69. Variant is G > A and the amino acid change is S488F – Valine (V) replaces Alanine (A) at codon 173. The maximum allelic frequency of this variant is 0.00314.

c.74-22C>T is a HLA-B gene intron variant. It has a homozygous genotype. Variant is G > A and the maximum allelic frequency of this variant is 0.0605.

Individual No. 22 – 14811055

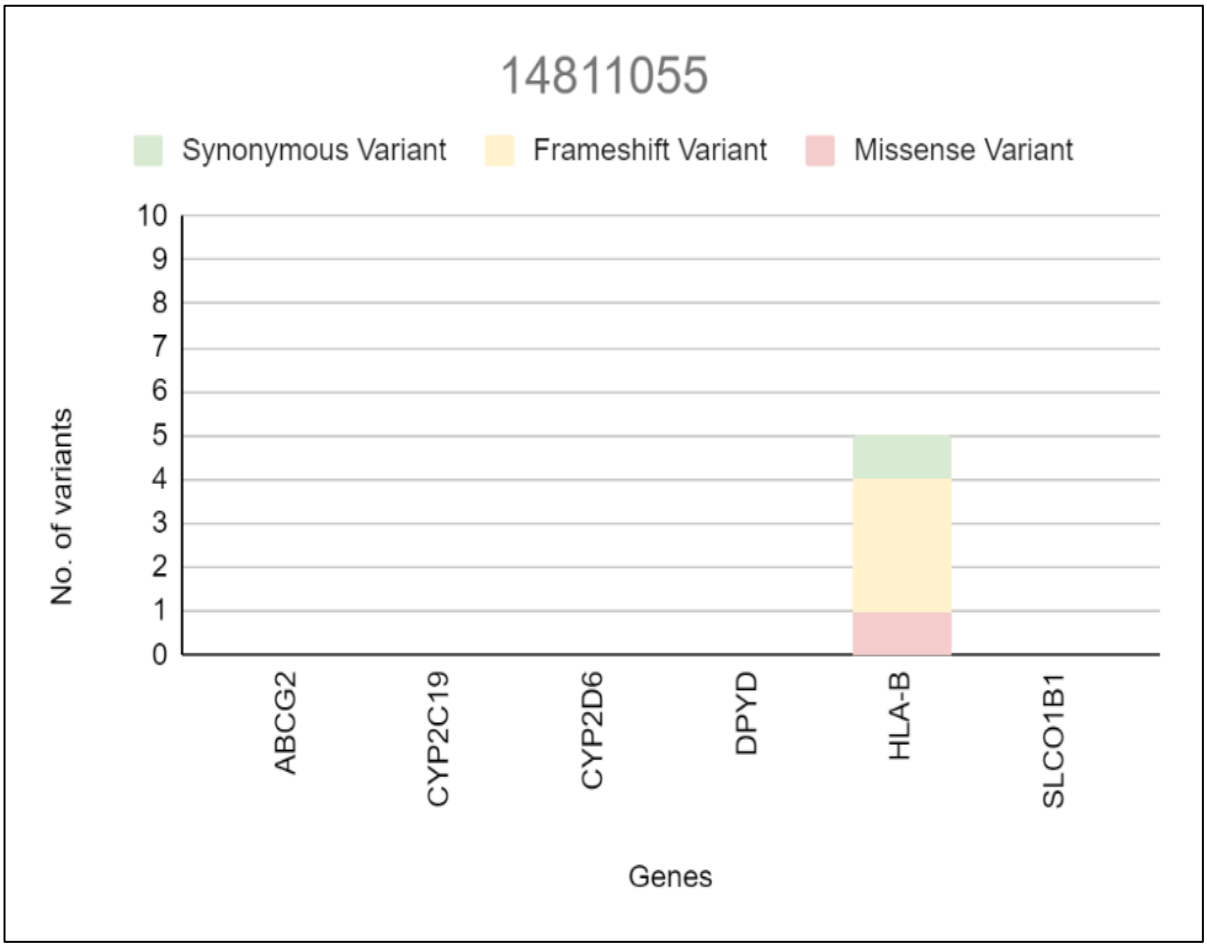

Figure 26 – Column chart showing variants observed in 14811055

Individual No. 22 has the following variants present:

c.309G>C p.(Arg103=) is a HLA-B gene synonymous variant. It has a heterozygous genotype. Variant is C > G and the amino acid change is at R – Arginine at codon 103. The maximum allelic frequency of this variant is 0.0825.

c.311delA p.Asn104ThrfsTer47 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is GT > G and the amino acid change is N104X – Unknown amino acid (X) replaces Asparagine (N) at codon 104. The maximum allelic frequency of this variant is 0.0838.

c.314delT p.Leu105ArgfsTer46 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is CA > C and the amino acid change is L105X – Unknown amino acid (X) replaces Leucine (L) at codon 105. The maximum allelic frequency of this variant is 0.0838.

c.319_320insCC p.Gly107AlafsTer45 is a HLA-B gene frameshift variant and has a heterozygous genotype. Variant is C > CGG and the amino acid change is G107AX – Alanine and/or Unknown amino acid (X) replace Glycine (G) at codon 107. The maximum allelic frequency of this variant is 0.0646.

c.319G>T p.Gly107Cys is a HLA-B gene missense variant. It has a heterozygous genotype with a Polyphen score of possibly_damaging 0.869 and SIFT score of deleterious_low_confidence 0. Variant is C > A and the amino acid change is G107C – Cysteine (C) replaces Glycine (G) at codon 107. The maximum allelic frequency of this variant is 0.0874.

ABCG2 – c.1147C>T p.Arg383Cys (R383)

GENE	ABCG2
HGVSc	ENST00000237612.7:c.1147C>T
HGVSp	ENSP00000237612.3:p.Arg383Cys
VARIANT	G > A
GENOTYPE	Heterozygous
POLYPHEN	probably_damaging 0.999
SIFT	deleterious 0
GENE SCORES	HI 0.061
AMINO ACID	R383C
VEP CONSEQUENCE	Missense Variant
MAX AF	0.001

Table no. 7 – ABCG2 – c.1147C>T

The ABCG2 membrane protein is a major xenobiotic and endobiotic transporter, regulating pharmacological agent absorption and metabolism and inducing multidrug resistance in cancer. ABCG2 is also important in uric acid elimination, and its dysfunction causes gout. ABCG2-R383C has been shown to be a folding mutant. Although not examined as yet in human samples, has been already characterized in detailed in vitro expression studies, and was found to result in impaired protein folding, abrogated glycosylation and membrane surface expression in mammalian cells, as well as a loss of transport activity in an Sf9-baculovirus expression system. (Boglárka Zámbó Z. B., 2018)

Western blotting was used to assess ABCG2 protein expression in HEK293 cells, although the R383C variant showed no detectable expression in this system. This mutant was not identified in the plasma membrane and displayed minimal activity. The R383C mutant was also expressed in Sf9 cells, however it failed to show any ABCG2-ATPase activity. It is well known that Sf9 cells can efficiently synthesize membrane proteins with folding or processing issues. In these tests, the R383C ABCG2 mutant showed neither basal or drug-stimulated activity.

Western blotting studies showed the total protein expression of ABCG2 variants in stable HeLa cell lines. The R383C ABCG2 variant showed no stable protein expression, had low levels of ABCG2 proteins, and Hoechst dye extrusion was not detectable in the R383C variant. No ABCG2-dependent transport activity was detected in stable cell lines expressing the R383C variant and the variant did not reach the plasma membrane.

The high-resolution confocal microscopy images of stable HeLa cells revealed intracellular localization for the R383C variant. The co-localization of ABCG2 (green) and the Endoplasmic Reticulum marker GRP78 (magenta) in R383C indicates that this variant is retained in the ER. The ABCG2 R383C variant lacks fully glycosylated protein expression, whereas the appearance of a major non-glycosylated band when proteasome activity is inhibited indicates that the misfolded, non-glycosylated protein (retained in the ER) is eliminated by the proteasomes. This is clearly consistent with the ER accumulation observed in co-localization studies.

ABCG2 R383C protein variant is seriously damaged and rarely expressed in mammalian cells. The confocal microscopy results show that the small amounts of this ABCG2 variant do not reach the plasma membrane, are primarily found in the ER, and thus have significant folding issues. R383C is localized to the connector region. This region has been shown to be critical in the stabilization and structural rearrangements that occur during the ABCG2 transport cycle. An artificial mutation in the R383 position (R383A) has been reported to impair this variant's function and trafficking. This emphasizes the importance of this region within the ABCG2 protein and predicts that mutations in this region will be damaging.

R383C showed defects in expression and/or protein folding, indicating the mutation could contribute to ABCG2 functional deficiency. Because naturally occurring ABCG2 variants like R383C are detrimental to both expression and function, heterozygous and especially homozygous or compound heterozygous patients carrying such ABCG2 variants may be more prone to gout formation and drug toxicity, while being less likely to have drug resistant tumors. Furthermore, any potential treatment that promotes the trafficking of these variants to the plasma membrane will not correct their defective function. (Boglárka Zámbó O. M., 2020)

CYP2C19 – c.518C>T p.Ala173Val (A173V)

GENE	CYP2C19
HGVSc	NM_000769.2:c.518C>T
HGVSp	NP_000760.1:p.Ala173Val
VARIANT	C > T
GENOTYPE	Heterozygous
POLYPHEN	possibly_damaging 0.518
SIFT	tolerated 0.1
GENE SCORES	HI 0.044
PROTEIN CHANGE	A173V
VEP CONSEQUENCE	Missense variant
MAX AF	0.0429

Table no. 8 – CYP2C19 – c.518C>T

CYP2C19	Alamut visual prediction		This study		Weighted gnomAD		Fisher's exact test
Variant	PolyPhen-2	SIFT	Alleles counts		Alleles counts		
c.518C>T p.(Ala173Val)	possibly damaging	deleterious	alternate allele	reference allele	alternate allele	reference allele	P-value
			3	179	70	20,530	0.026

Table no. 9 – Alamut visual prediction, Weighted gnomAD and Fisher's test data for CYP2C19 – c.518C>T

According to the study conducted by Loke *et al*, rare variants were identified in 46 patients. 54 non-synonymous exonic variants (frequencies <1%) had statistically significantly (P-value ≤0.05) higher frequencies than the corresponding frequencies for East (94.6%) and South Asians (5.4%) in the Weighted Genome Aggregation Database (gnomAD). Polyphen-2 and SIFT predicted that 21 of the 52 missense variants would have a significant effect on the encoded proteins, while another 13 were predicted to be damaging. The following data was specific to the CYP2C19 rare variant c.518C>T p.Ala173Val. (Mun-Fai Loke, 2019)

Protein abundance score of CYP2C19 variants from Mayo Right 10K study (Lingxin Zhang, 2020) –

Variant	Allele Frequency	Wild-type	Heterozygous	Homozygous	Functional study	Abundance score	PharmVar
c.518C>T	0.005818	10080	4	0	Tolerated	0.643519769	No record
c.556C>T	8.242E-05	-	-	-	Severely damaging	0.564313994	No record
c.985C>T	0.0003872	10078	6	0	Severely damaging	0.554502254	No record

Table no. 10 – Protein abundance score of CYP2C19 variants from Mayo Right 10K study.

CYP2C19 – c.556C>T p.Arg186Cys (R186C)

GENE	CYP2C19
HGVSc	NM_000769.2:c.556C>T
HGVSp	NP_000760.1:p.Arg186Cys
VARIANT	C > T
GENOTYPE	Heterozygous
POLYPHEN	possibly_damaging 0.553
SIFT	deleterious 0.02
GENE SCORES	HI 0.044
PROTEIN CHANGE	R186C
VEP CONSEQUENCE	Missense Variant
MAX AF	0.0014

Table no. 11 – CYP2C19 – c.556C>T

CYP2C19 – c.985C>T p.Arg329Cys (R329C)

GENE	CYP2C19
HGVSc	NM_000769.2:c.985C>T
HGVSp	NP_000760.1:p.Arg329Cys
VARIANT	C > T
GENOTYPE	Heterozygous
POLYPHEN	benign 0.034
SIFT	tolerated 0.09
GENE SCORES	HI 0.044
PROTEIN CHANGE	R329C
VEP CONSEQUENCE	Missense Variant
MAX AF	0.00166

Table no. 12 – CYP2C19 – c.985C>T

Measuring metabolizing enzyme activity assay for variants of CYP2C19 –

Variant	ACMG/AMP	SIFT	PolyPhen-2	CADD	1000g allele frequency
c.556C>T [p.Arg186Cys]	variant of uncertain significance	0.02	1.00	24.80	0.0004
c.985C>T [p.Arg329Cys]	variant of uncertain significance	0.06	0.49	13.32	0.0002

Table no. 13 – metabolizing enzyme activity assay for variants of CYP2C19

Specific P450 measurement and metabolizing capability assay for mephenytoin and omeprazole –

Variant	c.556C>T [p.Arg186Cys]
Specific P450 content (pmol/mg protein)	not detected
OH-mephenytoin formation a (pmol/min/pmol P450) (%WT)	not detected
5-Hydroxy omeprazole formation a (pmol/min/pmol P450) (%WT)	not detected
ACMG/AMP	variant of uncertain significance
SIFT	Damaging
PolyPhen-2	Probably damaging
CADD	24.80

Variant	c.985C>T [p.Arg329Cys]
Specific P450 content (pmol/mg protein)	96.5
OH-mephenytoin formation a (pmol/min/pmol P450) (%WT)	0.39 ± 0.05 (165.2%)
5-Hydroxy omeprazole formation a (pmol/min/pmol P450) (%WT)	4.88 ± .22 (115.4%)
ACMG/AMP	variant of uncertain significance
SIFT	Tolerated
PolyPhen-2	Possibly damaging
CADD	13.32

Table no. 14 – P450 measurement and metabolizing capability assay for mephenytoin and omeprazole

The c.985C>T CYP2C19 variant had a metabolic capability of 165.2% to S-mephenytoin (S-mephenytoin 165.2% whereas, it was 115.4% (p = 0.0209) for omeprazole. SIFT and CADD evaluated the c.985C>T variant as benign but PolyPhen-2 evaluated it as possibly damaging. (Myung-Eui Seo, 2023)

CYP2D6 – c.19G>A p.Val7Met (V7M)

GENE	CYP2D6
HGVSc	NM_000106.5:c.19G>A
HGVSp	NP_000097.3:p.Val7Met
VARIANT	C > T
GENOTYPE	Heterozygous
POLYPHEN	benign 0.001
SIFT	tolerated 0.22
GENE SCORES	-
PROTEIN CHANGE	V7M
VEP CONSEQUENCE	Missense Variant
MAX AF	0.00542

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	NO
CADD	2.415
EXAC Frequency	0.328
GNOMAD Frequency	0.2351
DBSNP Frequency	0.3067
No. of Heterozygotes	16
Heterozygote frequency	39.02439024
No. of Homozygotes	1
Homozygote frequency	2.43902439
Allele Frequency	10.97560976

Table no. 15 – CYP2D6 – c.19G>A

CYP2D6 – c.31G>A p.Val11Met (V11M)

GENE	CYP2D6
HGVSc	NM_000106.5:c.31G>A
HGVSp	NP_000097.3:p.Val11Met
VARIANT	C > T
GENOTYPE	Heterozygous

POLYPHEN	benign 0.024
SIFT	tolerated 0.14
GENE SCORES	-
PROTEIN CHANGE	V11M
VEP CONSEQUENCE	Missense Variant
MAX AF	0.0563

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	NO
CADD	5.192
EXAC Frequency	5.3012
GNOMAD Frequency	3.9348
DBSNP Frequency	-
No. of Heterozygotes	22
Heterozygote frequency	53.65853659
No. of Homozygotes	17
Homozygote frequency	41.46341463
Allele Frequency	34.14634146

Table no. 16 – CYP2D6 – c.31G>A

CYP2D6 – c.451C>G p.Gln151Glu (Q151E)

GENE	CYP2D6
HGVSc	NM_000106.5:c.451C>G
HGVSp	NP_000097.3:p.Gln151Glu
VARIANT	G > C
GENOTYPE	Heterozygous
POLYPHEN	benign 0.003
SIFT	tolerated 0.88
GENE SCORES	-
PROTEIN CHANGE	Q151E
VEP CONSEQUENCE	Missense Variant
MAX AF	0.00526

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	NO
CADD	0.002
EXAC Frequency	0.2392
GNOMAD Frequency	0.2333
DBSNP Frequency	99.7663
No. of Heterozygotes	1
Heterozygote frequency	2.43902439
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	0.609756098

Table no. 17 – CYP2D6 – c.451C>G

CYP2D6 – c.1108G>A p.Val370Ile (V370I)

GENE	CYP2D6
HGVSc	NM_000106.5:c.1108G>A
HGVSp	NP_000097.3:p.Val370Ile
VARIANT	C > T
GENOTYPE	Heterozygous
POLYPHEN	benign 0.001
SIFT	tolerated 0.87
GENE SCORES	-
PROTEIN CHANGE	V370I
VEP CONSEQUENCE	Missense Variant
MAX AF	0.00229

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	NO
CADD	2.428
EXAC Frequency	1.3867
GNOMAD Frequency	0.1439

DBSNP Frequency	1.3867
No. of Heterozygotes	16
Heterozygote frequency	39.02439024
No. of Homozygotes	12
Homozygote frequency	29.26829268
Allele Frequency	24.3902439

Table no. 18 – CYP2D6 – c.1108G>A

CYP2D6 – c.1117G>A p.Gly373Ser (G373S)

GENE	CYP2D6
HGVSc	NM_000106.5:c.1117G>A
HGVSp	NP_000097.3:p.Gly373Ser
VARIANT	C > T
GENOTYPE	Heterozygous
POLYPHEN	benign 0.074
SIFT	tolerated 0.23
GENE SCORES	-
PROTEIN CHANGE	G373S
VEP CONSEQUENCE	Missense Variant
MAX AF	0.00417

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	NO
CADD	15.94
EXAC Frequency	1.7085
GNOMAD Frequency	0.3559
DBSNP Frequency	80.7692
No. of Heterozygotes	28
Heterozygote frequency	68.29268293
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	17.07317073

Table no. 19 – CYP2D6 – c.1117G>A

CYP2D6 – c.841_843delAAG p.Lys281del (K281del)

GENE	CYP2D6
HGVSc	NM_000106.5:c.841_843delAAG
HGVSp	NP_000097.3:p.Lys281del
VARIANT	CCTT > C
GENOTYPE	-
POLYPHEN	-
SIFT	-
GENE SCORES	-
PROTEIN CHANGE	K281del
VEP CONSEQUENCE	Inframe deletion
MAX AF	0.0258

Gene Zone	Coding exon
Protein Type	Deletion
Splicing Region	YES
In Dataset	YES
CADD	18.02
EXAC Frequency	1.8972
GNOMAD Frequency	1.5486
DBSNP Frequency	-
No. of Heterozygotes	20
Heterozygote frequency	48.7804878
No. of Homozygotes	20
Homozygote frequency	48.7804878
Allele Frequency	36.58536585

Table no. 20 – CYP2D6 – c.841_843delAAG

(Luis Ramudo-Cela, 2021)

DPYD – c.1236G>A p.(=) (E412E)

GENE	DPYD
HGVSc	NM_000110.3:c.1236G>A
HGVSp	p.(=)
VARIANT	C > T
GENOTYPE	Heterozygous
POLYPHEN	-
SIFT	-
GENE SCORES	HI 0.272
PROTEIN CHANGE	E412E
VEP CONSEQUENCE	Synonymous Variant
MAX AF	0.0239

Table no. 21 – DPYD – c.1236G>A

DPYD – c.2194G>A p.Val732Ile (V732I)

GENE	DPYD
HGVSc	NM_000110.3:c.2194G>A
HGVSp	NP_000101.2:p.Val732Ile
VARIANT	C > T
GENOTYPE	Heterozygous
POLYPHEN	possibly_damaging 0.901
SIFT	tolerated 0.1
GENE SCORES	HI 0.272
PROTEIN CHANGE	V732I
VEP CONSEQUENCE	Missense Variant
MAX AF	0.0975

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	NO
CADD	25.9
EXAC Frequency	4.6473
GNOMAD Frequency	4.5309

DBSNP Frequency	-
No. of Heterozygotes	1
Heterozygote frequency	2.43902439
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	0.609756098

Table no. 22 – DPYD – c.2194G>A

(Luis Ramudo-Cela, 2021)

DPYD gene encodes the dihydropyrimidine dehydrogenase enzyme which is responsible for the catabolism of pyrimidine bases uracil and thymine. It converts uracil to 5,6-dihydrouracil and thymine to 5,6-dihydrothymine. (Kniffin, 2010)

Individuals with dihydropyrimidine dehydrogenase deficiency can develop conditions like microcephaly and eye conditions like abnormal ocular movement, microphthalmia, coloboma, nystagmus, and optic atrophy. It can also lead to neurologic conditions like seizures, motor retardation, mental retardation, delayed speech development, lethargy, hypotonia, hypertonia, tetraplegia, white matter abnormalities, cerebral atrophy and a rare condition like agenesis of the corpus callosum. The deficiency is also known to be associated with behavioral psychiatric conditions such as hyperactivity and autism disorder.

Increased urinary uracil, increased urinary thymine and decreased or absent dihydropyrimidine dehydrogenase activity are the abnormal biomarkers (laboratory abnormalities). Individuals with this DPYD gene variants could either be affected by such conditions during infancy or later in life. These individuals could also be asymptomatic despite having the gene variant. DPYD gene has a highly variable phenotype and heterozygous mutation carriers show severe toxicity to antimetabolite (anti-cancer) drugs like fluoropyrimidines (5-fluorouracil or 5FU and capecitabine). (Kniffin, 2010)

The frequency of heterozygosity for the c.1236G>A mutation was found to be 2.6%, 3.3%, and 1.9% in the Dutch, German, and Tunisian populations, respectively. Nine cases of the c.1236G>A mutation were discovered in 203 German patients diagnosed with cancer between 2002 and 2009 and evaluated for 5FU-associated side effects. Six of them had severe grade III/IV 5FU-related toxicity, primarily neutropenia, leucopenia, and diarrhea. A haplotype

with three intronic polymorphisms (c.483 + 18G>A, c.959-51T>G, c.680 + 139G>A) and the c.1236G>A mutation was linked to severe 5FU toxicity. (André B. P. van Kuilenburg, 2010)

There was a strong association for the hapB3-specific variant 1236G>A (13% vs. 5%; P = 0.08), which showed high LD. The exonic SNP 1236G>A was significantly associated with toxicity when it was included in the multivariate analysis. It was strongly associated with grade 3 to 4 diarrhea in 568 patients with advanced colorectal cancer treated with capecitabine-based chemotherapy ($P < 0.05$; FDR < 0.3), whereas 2194G>A showed an intermediate association ($P < 0.05$; FDR $0.3 \leq x < 0.4$). Positive predictive values ranged between 50% and 41%. (Maarten J Deenen, 2011)

In the locus-by-locus analysis, all polymorphisms found to be associated with severe 5-FU toxicity were located within this haplotype block, and four of them were combined in haplotype B3 (IVS5+18G>A, IVS6+139G>A, IVS9-51T>G, and c.1236G>A). The only homozygous carrier experienced lethal 5-FU toxicity, while four of the seven patients developed severe 5-FU toxicity (57%). Furthermore, the frequency of haplotype B3 was found to be positively related to increasing toxicity grade. Surprisingly, the haplotype consists of one synonymous SNP in exon 11 (c.1236G>A) and three intronic polymorphisms (IVS5+18G>A, IVS6+139G>A, and IVS9-51T>G), with no nonsynonymous or splice-site mutations. (Ursula Amstutz, 2009)

A synonymous sequence variant (c.1236G>A) in exon 11 of DPYD was found in the associated haplotype hapB3. In two prospective cohorts of 500 patients receiving FP-based chemotherapy, the c.1129-5923C>G mutation was found to be in perfect linkage with c.1236G>A, with a carrier frequency of 4.6% and a 3.7-fold increased relative risk of severe toxicity. When this variant was used alone, the sensitivity for detecting early-onset FP toxicity increased by 15%. The hapB3 haplotype, which is linked to c.1129-5923C>G, is a major contributor to early onset severe toxicity in Caucasian carriers. (Satoshi Matsusaka, 2015)

Capecitabine or 5-FU dose reduction is recommended and required for heterozygous patients carrying the c.1236G>A/HapB3 variant because they will have one decreased activity allele and one fully functional allele, resulting in DPD enzyme activity of 75% of normal. This variant has a gene activity score of 1.5, which corresponds to a recommended starting dose of 75% of the standard dose (25% dose reduction). According to the toxicity data, a 50% dose

reduction would be too large given the measured enzyme activities. To prevent toxicity as well as underdosing, 75% of the normal dose i.e. 25% dose reduction for heterozygous patients appears appropriate for the first treatment cycle. Following the initial dose reduction, the patient should be closely monitored, and the dose can be adjusted based upon the toxicity that occurs. (Linda M Henricks, 2015)

The mean DPD activity in DPYD wild-type patients was 9.4 nmol/mg per h (SD 3.6), which was consistent with previously reported results. The mean DPD activity for the 35 c.1236G>A variant carriers was 7.5 nmol/mg per h (2.8; a 20% drop compared to wild-type). However, there was a significant heterogeneity in this activity, implying that some patients require a higher dose reduction, while others may tolerate the full dose. These findings are consistent with a wide range of pharmacokinetic exposure observed in c.1236G>A carriers.

Personalized dose titration is essential to ensure that all patients receive an adequate and safe dose. In heterozygous c.1236G>A carriers, dosage reductions of 25% were insufficient to reduce the risk of fluoropyrimidine-related toxicity to the observed risk in wild-type patients. A higher initial dose reduction of 50% for c.1236G>A carriers should be considered, with further individual dose titrations, and the CPIC recommends a 50% dose reduction based on individual drug dosage tolerance. (Linda M Henricks C. A.-J., 2018)

Heterozygous carriers of the c.1236G>A variant have enzyme activity reductions ranging from 20% to 35%. Reductions of 20% to 70% have been observed in homozygous carriers as well. The most common variant in Caucasians is c.1236G>A, which affects 2.6% to 6.3% of the population. Because it reduces DPYD activity, c.1236G>A is classified as a decreased-function variant.

Across six studies, the frequency of overall severe toxicity ranged from 30.0% to 92.9% in heterozygous c.1236G>A carriers treated with a standard fluoropyrimidine dose and 8.2% to 85.0% in wild-type patients treated with a standard dose. In two studies, two homozygous c.1236G>A carriers experienced severe toxicity (100%), compared to 8.5% (57) and 32.3% (13) in wild-type patients. In both studies, the risk of severe toxicity was higher in homozygous carriers than in wild-type patients (RR 3.10, 95% CI 1.47-3.31; 13 and RR 11.76, 95% CI 4.73-14.91; 57).

In one study, seventeen (22.1%) heterozygous c.1236G>A carriers and 184 (9.8%) wild-type patients treated with a standard fluoropyrimidine dose experienced severe neutropenia (RR 2.26, 95% CI 1.38-3.40). The single homozygous carrier found in this study did not have severe neutropenia.

The most commonly reported gastrointestinal toxicity was severe diarrhea. In one study, 11 (14.3%) c.1236G>A heterozygous carriers and 234 (12.5%) wild-type patients treated with a standard fluoropyrimidine dose experienced severe diarrhea (RR 1.14, 95% CI 0.61-1.92). A single homozygous carrier (100.0%) experienced severe nausea and/or vomiting, whereas 88 (4.7%) wild-type patients did not (RR 21.3, 95% CI 9.29-25.49). According to a second study, 14 (50.0%) carriers and 125 (23.1%) wild-type patients had severe diarrhea (RR 2.16, 95% CI 1.35-3.34).

According to one study, 26 (92.9%) heterozygous c.1236G>A carriers and 459 (85.0%) wild-type patients treated with a standard fluoropyrimidine dose experienced severe (grade 3) hand-foot syndrome (RR 1.09, 95% CI 0.91-1.95).

One homozygous c.1236G>A carrier died as a result of severe fluoropyrimidine-related toxicity in a single study that reported mortality. (Ontario-Health, 2021)

c.1236G>A variant is prevalent in 4-6% of European (Caucasian) population, 1.4% of South Asian population and absent in East Asian population. It is an Intermediate Metabolizer according to CPIC and its adjusted relative risk is 1.59 in Caucasian carriers. Variant allele frequency of c.1236G>A is 0.021 in European (Finnish) population, 0.012 in European (Non-Finnish) population, 0.003 in African/African American population, 0.017 in South Asian population, 0.005 in Latino/Admixed American population, 0.007 in Ashkenazi Jewish population, 0.015 in Other population and its Total(overall) variant allele frequency is 0.014. It is absent (0.000) in East Asian population.

c.2194G>A variant is prevalent in 8.4% of South Asian population and has shown deleterious effect on in silico analysis of South Asian population. It causes Increased grade ≥2 fluoropyrimidine toxicity in Indian carriers. According to CPIC, it has no effect on DPD enzyme activity in Caucasian population. DPD activity was reduced by 29% in 6 African-American carriers of c.2194G>A (p.V732I) compared to wild-type carriers (0.361 versus 0.508 nmol/min/mg, p=0.049).

In European-American participants carrying the same variant, there was no associated reduction in DPD activity. Variant allele frequency of c. 2194G>A is 0.021 in European (Finnish) population, 0.046 in European (Non-Finnish) population, 0.024 in African/African American population, 0.098 in South Asian population, 0.019 in East Asian population, 0.026 in Latino/Admixed American population, 0.107 in Ashkenazi Jewish population, 0.058 in Other population and its Total(overall) variant allele frequency is 0.045. (Cassandra White, 2021)

c.2194G>A is associated with clinically-relevant ADRs like nausea/vomiting ($p = 0.830$), diarrhea ($p = 0.725$), stomatitis ($p = 0.053$ OR 1.514), dermatitis ($p = 0.152$), alopecia ($p = 0.886$), leucopenia ($p = 0.003$ OR 1.895), neutropenia ($p = 0.127$), febrile neutropenia ($p = 0.062$ OR 1.838), anemia ($p = 0.595$), thrombocytopenia ($p = 0.049$ OR 1.796), HFS ($p = 0.371$), and fever ($p = 0.084$). Toxicity risk factor is indicated by OR > 1 if associated with a p-value < 0.05. Boige et al. studied a group of 1545 patients and discovered a link between ADRs and the c.2194G>A variant. The statistical analysis found a link between G≥3 ADRs caused by 5-FU and c.2194G>A (OR = 1.7; p 0.001); more specifically, G≥3 hematologic adverse events (OR = 1.9) and G≥3 neutropenia (OR = 1.8) were linked to c.2194G>A. (Marzia Del Re, 2019)

The most commonly found SNP was c.2194G>A as it was present in 46 of 366 patients (12.5%): 28 patients in the cases group (60% of 46 patients) and 18 patients in the controls group (40%).There is a significant link between c.2194G>A and subjects who experienced severe toxicities as this SNP was distributed differently in the two groups.

A case-cohort analysis was performed on 568 previously untreated patients with advanced colon cancer who were assigned to capecitabine therapy combined with oxaliplatin and bevacizumab with or without cetuximab in the CAIRO trial. The c.2194G>A variant was linked to grade 3-4 diarrhea but had a low predictive value of 41%. In 1545 colorectal cancer patients who received standard adjuvant FOLFOX4 or FOLFOX4 in combination with cetuximab in the PETACC-8 trials, increased fluorouracil-related AEs, including hematologic AEs and neutropenia (OR 1.8, 95% CI 1.3–2.4) was observed in patients carrying the c.2194G>A SNP.

Colon cancer patients were enrolled in the TOSCA trial's pharmacogenetics study for 3 or 6 months of either FOLFOX or XELOX adjuvant chemotherapy. FAEs were more common in c.2194G>A carriers in the association analysis of 508 patients, and this DPYD variant had a negative effect on the time to neutropenia. The TTT analysis revealed the clinical significance

of the c.2194G>A SNP in DPYD variant carriers, with a significantly shorter TTT occurrence. TTTs were 7.0, 3.0, and 2.1 months for c.2194G>A GG, GA, and AA genotype carriers, respectively.

Zygosity	Variant	c.DNA	% of standard dose
Wild type	Haplotype B3	c.1236 GG	100
	DPYD V732I	c.2194 GG	100
Heterozygous	Haplotype B3	c.1236 GA	75
	DPYD V732I	c.2194 GA	85
Homozygous	Haplotype B3	c.1236 AA	50
	DPYD V732I	c.2194 AA	70

Table no. 23 – DPYD and Haplotype B3 variants and their standard recommended doses.

A statistically significant link was discovered between c.2194G>A and neutropenia. It was observed in 50% of patients with c.2194G>A (23 out of 46) vs. 21% of patients with WT genotype (67 out of 320) (OR = 3.75, 95% CI 1.98-7.10; p 0.0001). Patients with the c.2194G>A variant had an even earlier onset of AE than the WT, as measured by a median TTT (time to toxicity) of 6 cycles vs. 10 (p = 0.0022), with a significant difference in Kaplan-Meier curves between the two groups. In terms of the significant correlation with hematologic toxicity, these findings are consistent with the PETACC8 and TOSCA studies. (Francesco Iachetta, 2019)

HLA-B – c. 74-22C>T

GENE	HLA-B
HGVSc	NM_005514.7:c.74-22C>T
HGVSp	-
VARIANT	G > A
GENOTYPE	Homozygous
POLYPHEN	-
SIFT	-
GENE SCORES	-
PROTEIN CHANGE	-
VEP CONSEQUENCE	Intron Variant
MAX AF	0.0605

Gene Zone	Intron
Protein Type	-
Splicing Region	NO
In Dataset	YES
CADD	13.91
EXAC Frequency	4.2382
GNOMAD Frequency	3.4719
DBSNP Frequency	-
No. of Heterozygotes	3
Heterozygote frequency	7.317073171
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	1.829268293

Table no. 24 – HLA-B – c. 74-22C>T

HLA-B – c.97T>G p.Tyr33Asp (Y33D)

GENE	HLA-B
HGVSc	NM_005514.7:c.97T>G
HGVSp	NP_005505.2:p.Tyr33Asp
VARIANT	A > C
GENOTYPE	Heterozygous

POLYPHEN	benign 0.005
SIFT	tolerated_low_confidence 0.21
GENE SCORES	-
PROTEIN CHANGE	Y33D
VEP CONSEQUENCE	Missense Variant
MAX AF	0.0754

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	YES
CADD	0.01
EXAC Frequency	4.63
GNOMAD Frequency	5.8449
DBSNP Frequency	-
No. of Heterozygotes	8
Heterozygote frequency	19.51219512
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	4.87804878

Table no. 25 – HLA-B – c.97T>G

HLA-B – c.141C>T p.(Ile47=)

GENE	HLA-B
HGVSc	NM_005514.7:c.141C>T
HGVSp	NP_005505.2:p.(Ile47=)
VARIANT	G > A
GENOTYPE	Heterozygous
POLYPHEN	-
SIFT	-
GENE SCORES	-
PROTEIN CHANGE	-
VEP CONSEQUENCE	Synonymous variant
MAX AF	0.0978

Gene Zone	Coding exon
Protein Type	Synonymous
Splicing Region	NO
In Dataset	YES
CADD	9.471
EXAC Frequency	7.4514
GNOMAD Frequency	5.7783
DBSNP Frequency	-
No. of Heterozygotes	32
Heterozygote frequency	78.04878049
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	19.51219512

Table no. 26 – HLA-B – c.141C>T

HLA-B – c.301A>G p.Ser101Gly (S101G)

GENE	HLA-B
HGVSc	NM_005514.7:c.301A>G
HGVSp	NP_005505.2:p.Ser101Gly
VARIANT	T > C
GENOTYPE	Heterozygous
POLYPHEN	benign 0
SIFT	tolerated_low_confidence 0.23
GENE SCORES	-
PROTEIN CHANGE	S101G
VEP CONSEQUENCE	Missense Variant
MAX AF	0.0586

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	YES
CADD	0.918
EXAC Frequency	4.5549

GNOMAD Frequency	3.7244
DBSNP Frequency	-
No. of Heterozygotes	18
Heterozygote frequency	43.90243902
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	10.97560976

Table no. 27 – HLA-B – c.301A>G

HLA-B – c.309G>C p.(Arg103=)

GENE	HLA-B
HGVSc	NM_005514.7:c.309G>C
HGVSp	NP_005505.2:p.(Arg103=)
VARIANT	C > G
GENOTYPE	Heterozygous
POLYPHEN	-
SIFT	-
GENE SCORES	-
PROTEIN CHANGE	-
VEP CONSEQUENCE	Synonymous variant
MAX AF	0.0825

Gene Zone	Coding exon
Protein Type	Synonymous
Splicing Region	NO
In Dataset	YES
CADD	9.174
EXAC Frequency	2.5521
GNOMAD Frequency	3.0411
DBSNP Frequency	97.2402
No. of Heterozygotes	15
Heterozygote frequency	36.58536585
No. of Homozygotes	4
Homozygote frequency	9.756097561
Allele Frequency	14.02439024

Table no. 28 – HLA-B – c.309G>C

HLA-B – c.311delA p.Asn104ThrfsTer47 (N104X)

GENE	HLA-B
HGVSc	NM_005514.7:c.311delA
HGVSp	NP_005505.2:p.Asn104ThrfsTer47
VARIANT	GT > G
GENOTYPE	Heterozygous
POLYPHEN	-
SIFT	-
GENE SCORES	-
PROTEIN CHANGE	N104X
VEP CONSEQUENCE	Frameshift variant
MAX AF	0.0838

Gene Zone	Coding exon
Protein Type	Frame Shift
Splicing Region	NO
In Dataset	YES
CADD	9.509
EXAC Frequency	3.2488
GNOMAD Frequency	3.3109
DBSNP Frequency	3.4633
No. of Heterozygotes	7
Heterozygote frequency	17.07317073
No. of Homozygotes	33
Homozygote frequency	80.48780488
Allele Frequency	44.51219512

Table no. 29 – HLA-B – c.311delA

HLA-B – c.314delT p.Leu105ArgfsTer46 (L105X)

GENE	HLA-B
HGVSc	NM_005514.7:c.314delT
HGVSp	NP_005505.2:p.Leu105ArgfsTer46
VARIANT	CA > C
GENOTYPE	Heterozygous

POLYPHEN	-
SIFT	-
GENE SCORES	-
PROTEIN CHANGE	L105X
VEP CONSEQUENCE	Frameshift variant
MAX AF	0.0838

Gene Zone	Coding exon
Protein Type	Frame Shift
Splicing Region	NO
In Dataset	YES
CADD	19.97
EXAC Frequency	3.244
GNOMAD Frequency	3.2546
DBSNP Frequency	6.5895
No. of Heterozygotes	3
Heterozygote frequency	7.317073171
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	1.829268293

Table no. 30 – HLA-B – c.314delT

HLA-B – c.319G>T p.Gly107Cys (G107C)

GENE	HLA-B
HGVSc	NM_005514.7:c.319G>T
HGVSp	NP_005505.2:p.Gly107Cys
VARIANT	C > A
GENOTYPE	Heterozygous
POLYPHEN	possibly_damaging 0.869
SIFT	deleterious_low_confidence 0
GENE SCORES	-
PROTEIN CHANGE	G107C
VEP CONSEQUENCE	Missense Variant
MAX AF	0.0874

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	YES
CADD	23.3
EXAC Frequency	4.2489
GNOMAD Frequency	4.4768
DBSNP Frequency	-
No. of Heterozygotes	5
Heterozygote frequency	12.19512195
No. of Homozygotes	36
Homozygote frequency	87.80487805
Allele Frequency	46.95121951

Table no. 31 – HLA-B – c.319G>T

HLA-B – c.319_320insCC p.Gly107AlafsTer45 (G107AX)

GENE	HLA-B
HGVSc	NM_005514.7:c.319_320insCC
HGVSp	NP_005505.2:p.Gly107AlafsTer45
VARIANT	C > CGG
GENOTYPE	Heterozygous
POLYPHEN	-
SIFT	-
GENE SCORES	-
PROTEIN CHANGE	G107AX
VEP CONSEQUENCE	Frameshift variant
MAX AF	0.0646

Gene Zone	Coding exon
Protein Type	Frame Shift
Splicing Region	NO
In Dataset	YES
CADD	23.1
EXAC Frequency	2.9093

GNOMAD Frequency	3.419
DBSNP Frequency	-
No. of Heterozygotes	11
Heterozygote frequency	26.82926829
No. of Homozygotes	28
Homozygote frequency	68.29268293
Allele Frequency	40.85365854

Table no. 32 – HLA-B – c.319_320insCC

HLA-B – c.354_355delCC p.Leu119ProfsTer19 (L119X)

GENE	HLA-B
HGVSc	NM_005514.7:c.354_355delCC
HGVSp	NP_005505.2:p.Leu119ProfsTer19
VARIANT	AGG > A
GENOTYPE	Heterozygous
POLYPHEN	-
SIFT	-
GENE SCORES	-
PROTEIN CHANGE	L119X
VEP CONSEQUENCE	Frameshift variant
MAX AF	0

Gene Zone	Coding exon
Protein Type	Frame Shift
Splicing Region	YES
In Dataset	YES
CADD	23.4
EXAC Frequency	7.3031
GNOMAD Frequency	8.6855
DBSNP Frequency	-
No. of Heterozygotes	26
Heterozygote frequency	63.41463415
No. of Homozygotes	3
Homozygote frequency	7.317073171
Allele Frequency	19.51219512

Table no. 33 – HLA-B – c.354_355delCC

HLA-B – c.361A>T p.Ser121Cys (S121C)

GENE	HLA-B
HGVSc	NM_005514.7:c.361A>T
HGVSp	NP_005505.2:p.Ser121Cys
VARIANT	T > A
GENOTYPE	Heterozygous
POLYPHEN	possibly_damaging 0.876
SIFT	tolerated_low_confidence 0.17
GENE SCORES	-
PROTEIN CHANGE	S121C
VEP CONSEQUENCE	Missense Variant
MAX AF	0.0227

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	YES
CADD	0.002
EXAC Frequency	1.0155
GNOMAD Frequency	1.4367
DBSNP Frequency	96.8779
No. of Heterozygotes	2
Heterozygote frequency	4.87804878
No. of Homozygotes	1
Homozygote frequency	2.43902439
Allele Frequency	2.43902439

Table no. 34 – HLA-B – c.361A>T

HLA-B – c.411T>C p.(His137=)

GENE	HLA-B
HGVSc	NM_005514.7:c.411T>C
HGVSp	NP_005505.2:p.(His137=)
VARIANT	A > G
GENOTYPE	Homozygous

POLYPHEN	-
SIFT	-
GENE SCORES	-
PROTEIN CHANGE	-
VEP CONSEQUENCE	Synonymous variant
MAX AF	0.0547

Gene Zone	Coding exon
Protein Type	Synonymous
Splicing Region	NO
In Dataset	YES
CADD	0.062
EXAC Frequency	3.7801
GNOMAD Frequency	3.3778
DBSNP Frequency	-
No. of Heterozygotes	5
Heterozygote frequency	12.19512195
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	3.048780488

Table no. 35 – HLA-B – c.411T>C

HLA-B – c.412G>C p.Asp138His (D138H)

GENE	HLA-B
HGVSc	NM_005514.7:c.412G>C
HGVSp	NP_005505.2:p.Asp138His
VARIANT	C > G
GENOTYPE	Homozygous
POLYPHEN	benign 0.025
SIFT	tolerated_low_confidence 0.5699
GENE SCORES	-
PROTEIN CHANGE	D138H
VEP CONSEQUENCE	Missense Variant
MAX AF	0.0547

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	YES
CADD	0.357
EXAC Frequency	3.5677
GNOMAD Frequency	3.6749
DBSNP Frequency	51.0258
No. of Heterozygotes	1
Heterozygote frequency	2.43902439
No. of Homozygotes	-
Homozygote frequency	0
Allele Frequency	0.609756098

Table no. 36 – HLA-B – c.412G>C

(Luis Ramudo-Cela, 2021)

SLCO1B1 – c.1929A>C p.Leu643Phe (L643F)

GENE	SLCO1B1
HGVSc	NM_006446.4:c.1929A>C
HGVSp	NP_006437.3:p.Leu643Phe
VARIANT	A > C
GENOTYPE	Heterozygous
POLYPHEN	benign 0.005
SIFT	tolerated 1
GENE SCORES	HI 0.055
PROTEIN CHANGE	L643F
VEP CONSEQUENCE	Missense Variant
MAX AF	0.0654

Gene Zone	Coding exon
Protein Type	Nonsynonymous
Splicing Region	NO
In Dataset	NO
CADD	3.415
EXAC Frequency	4.6322
GNOMAD Frequency	4.5844
DBSNP Frequency	4.6241
No. of Heterozygotes	20
Heterozygote frequency	48.7804878
No. of Homozygotes	12
Homozygote frequency	29.26829268
Allele Frequency	26.82926829

Table no. 37 – SLCO1B1 – c.1929A>C

(Luis Ramudo-Cela, 2021)

The SLCO1B1 gene encodes a membrane-bound sodium-independent organic anion transporter protein 1B1 (OATP1B1). The active cellular influx of many endogenous and xenobiotic compounds is regulated by this protein. OATP1B1 is a genetically polymorphic transporter that is primarily expressed on human hepatocytes' sinusoidal (basolateral) membrane. OATP1B1 variants have been linked to changes in the pharmacokinetics of substrate drugs, treatment response, and the risk of drug-induced toxicities in clinical studies.

The c.1929A>C (p.Leu643Phe) variant is common and has a phenotype of an extensive transporter but has a low minor allele frequency of 1%-5%. (Hannah H. Lee, 2017)

The polymorphism SLCO1B1 c.1929A>C has been associated with elevated activity, which results in increased hepatic drug uptake, including increased hepatic uptake of the antifolate methotrexate (P = 0.028) in cancer patients and increased atorvastatin uptake. When compared to individuals with the SLCO1B1 c.1929A/A genotype, the SLCO1B1 c.1929A/C genotype was associated with a 75% reduction in Cmax (P 0.001) of rosuvastatin. This decrease in Cmax in SLCO1B1 c.1929C allele carriers could indicate an increase in hepatic uptake due to increased transporter activity. These findings are consistent with the previously reported increased hepatic uptake of methotrexate and atorvastatin by the OATP1B1 Leu643Phe variant. Thus, the c.1929A>C variant plays a role in rosuvastatin response as well. (Nyarai D. Soko, 2011)

A genome wide association study of Simvastatin pharmacokinetics was conducted and many variants were screened for the drug-gene interactions and the results specific to SLCO1B1 missense variant c.1929A>C with a 36% decrease in Simvastatin acid AUC. These haplotypes have been linked to higher OATP1B1 expression in the liver. Additionally, they have been linked to lower OATP1B1 biomarker concentrations, as well as increased methotrexate clearance in humans. This variant is an increased function variant because it increases hepatic uptake of simvastatin acid, and thus lowering its systemic exposure. (Anssi J. H. Mykkänen, 2022)

Variants that are not found in any of the NCBI databases

Sr. No.	GENE	HGVSc	HGVSp	CHANGE	VARIANT	GENOTYPE	POLYPHEN	SIFT	VEP CONSEQUENCE	MAX AF
1	CYP2C19	c.1266T>C	p.(=)	-	T > C	Heterozygous	-	-	Synonymous variant	0
2	CYP2D6	c.1107C>T	p.(=)	-	G > A	Heterozygous	-	-	Synonymous variant	0.000194
3	DPYD	c.672T>C	p.(=)	-	A > G	Heterozygous	-	-	Synonymous variant	< 0.0001
4	HLA-B	c.74-40G>T	-	-	C > A	Heterozygous	-	-	Intron variant	0.0015
5		c.117C>T	p.(=)	-	G > A	Heterozygous	-	-	Synonymous variant	0.0015
6		c.228A>G	p.Ile76Met	I76M	T > C	Homozygous	benign 0.0399	tolerated_low_confidence 0.97	Missense variant	0.068
7		c.233C>T	p.Ala78Val	A78V	G > A	Homozygous	benign 0.046	deleterious 0.029	Missense variant	0.0665
8		c.239G>A	p.Gly80Glu	G80E	C > T	Homozygous	probably_damaging 0.948	tolerated_low_confidence 0.079	Missense variant	0
9		c.259A>C	p.Asn87His	N87H	T > G	Homozygous	possibly_damaging 0.5239	deleterious_low_confidence 0.009	Missense variant	0
10		c.265C>G	p.Gln89Glu	Q89E	G > C	Homozygous	benign 0.1829	deleterious_low_confidence 0.029	Missense variant	0
11		c.290C>T	p.Thr97Ile	T97I	G > A	Homozygous	benign 0.3429	tolerated_low_confidence 0.17	Missense variant	0
12		c.296G>T	p.Arg99Leu	R99L	C > A	Homozygous	benign 0.439	deleterious_low_confidence 0.019	Missense variant	0
13		c.341C>A	p.Ala114Asp	A114D	G > T	Heterozygous	benign 0.001	tolerated_low_confidence 0.25	Missense variant	0.0015
14		c.368A>G	p.Tyr123Cys	Y123C	T > C	Heterozygous	probably_damaging 0.924	deleterious_low_confidence 0.05	Missense Variant	< 0.0001
15		c.397C>T	p.Leu133Phe	L133F	G > A	Heterozygous	benign 0.02	tolerated_low_confidence 0.18	Missense variant	0.0144
16		c.991A>G	p.Met331Val	M331V	T > C	Heterozygous	benign 0	tolerated_low_confidence 0.09	Missense variant	0.0043

Table no. 38 – Variants that are not found in any of the NCBI databases.

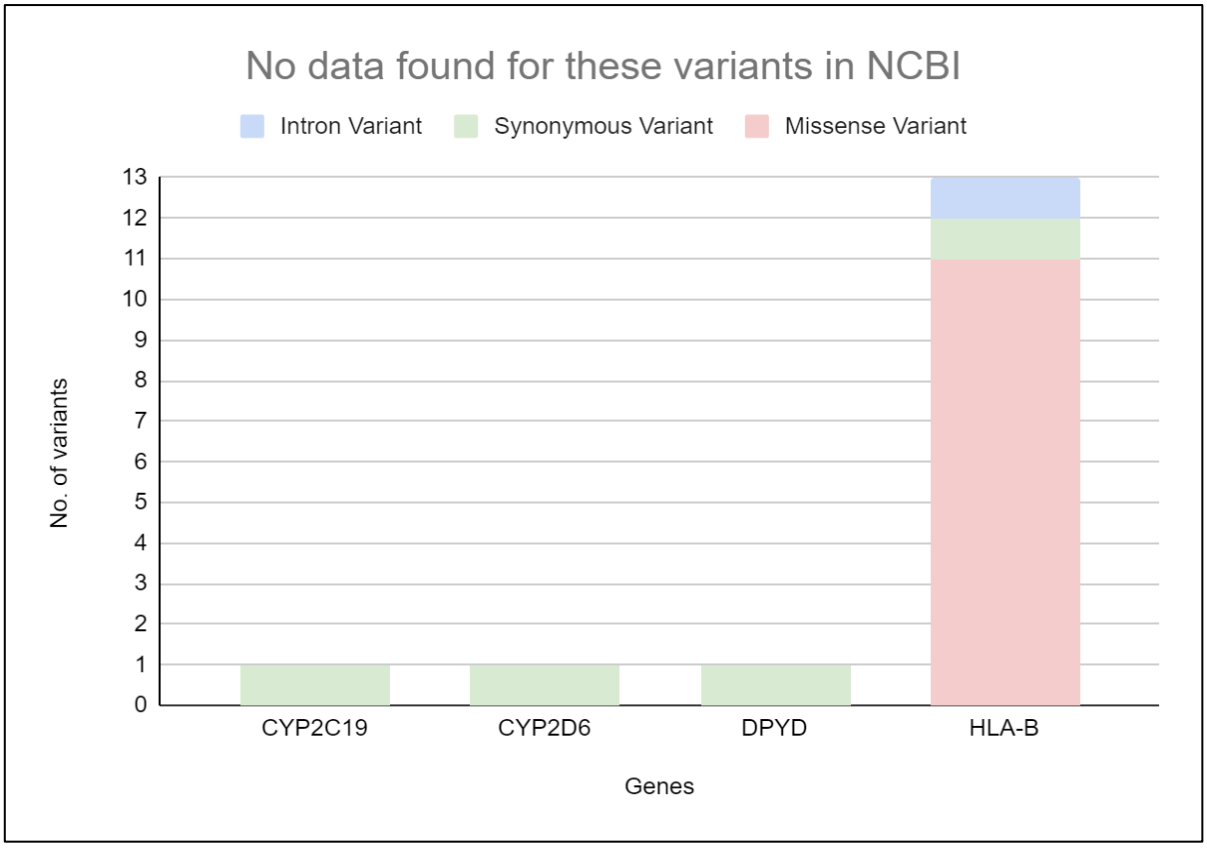

Figure 27 – Column chart showing variants for which no data was found in the NCBI databases.

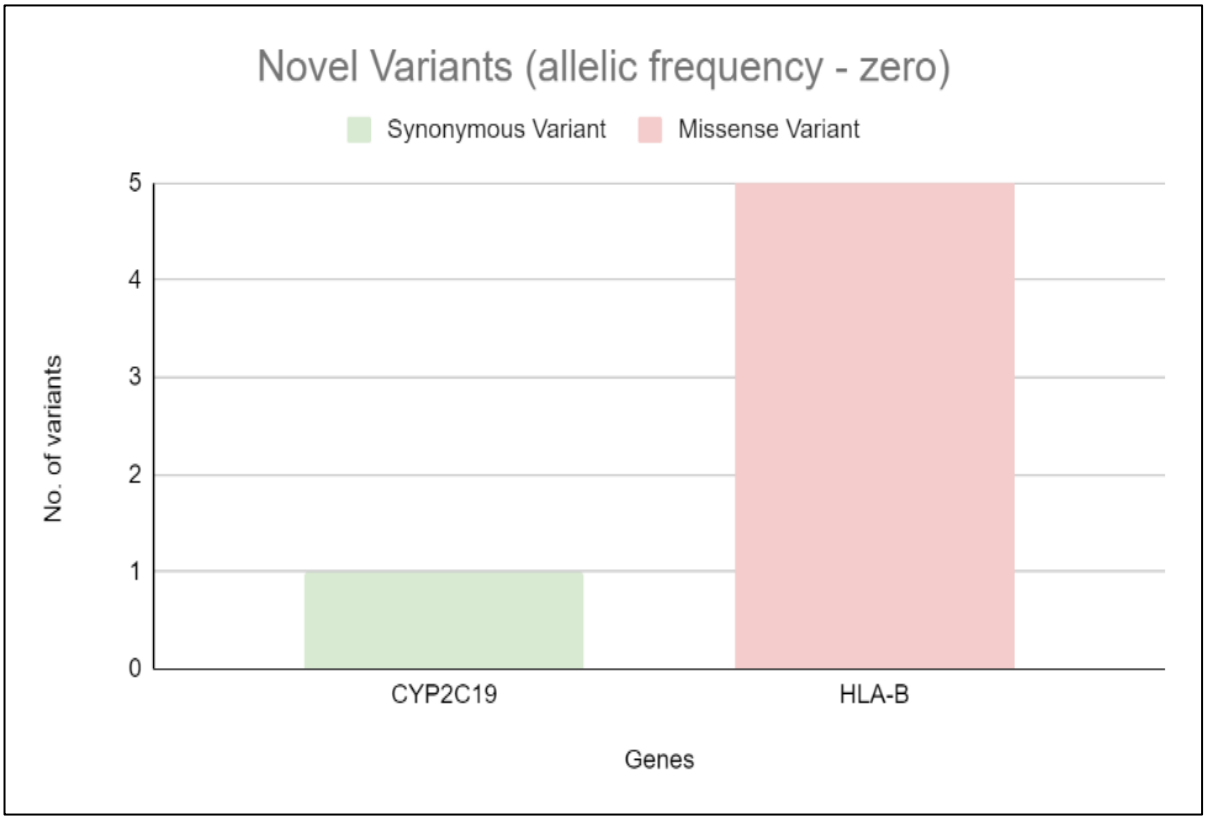

Figure 28 – Column chart showing Novel Variants with Zero allelic frequency.

RESULTS AND CONCLUSION

- The comprehensive study of the Pharmacogenomic gene panel of Zoroastrian-Parsi family members has provided interesting insights. One, in particular, appears to have the potential to reveal a lot more.
- There are 16 variants that are unique and have not been reported before in any of the NCBI databases as of now. Out of these 16 variants, 6 variants are of interest as they have Zero Allelic Frequency.
- c.1266T>C is a CYP2C19 gene synonymous variant that is heterozygous whereas, c.239G>A, c.259A>C, c.265C>G, c.290C>T, and c.296G>T are all HLA-B gene missense variants that are homozygous.
- These variants were found to be novel variants across 6 Pharmacogenomic genes of the sample database of 22 individuals. These variants have been scored as possibly damaging, probably damaging and deleterious by Polyphen and SIFT predictions.
- Testing and detailed analysis of these 6 variants could detect if these variants have a statistically significant impact on protein structure folding. If these variants cause protein structure misfolding, potentially, these variants could also be disease associated.
- There is a possibility that these novel variants are unique to the Zoroastrian-Parsi population. A larger study can be conducted to verify whether the results of this study hold true for the rest of the ZP population.

REFERENCES

André B. P. van Kuilenburg, J. M.-C.-G.-J. (2010). Intragenic deletions and a deep intronic mutation affecting pre-mRNA splicing in the dihydropyrimidine dehydrogenase gene as novel mechanisms causing 5-fluorouracil toxicity. *Human Genetics*, 128(5), 529–538. doi:https://doi.org/10.1007/s00439-010-0879-3

Anssi J. H. Mykkänen, S. T.-H. (2022). Genomewide Association Study of Simvastatin Pharmacokinetics. *Clinical Pharmacology and Therapeutics*, 112(3), 676–686. doi:https://doi.org/10.1002/cpt.2674

Boglárka Zámbó, O. M. (2020). Cellular expression and function of naturally occurring variants of the human ABCG2 multidrug transporter. *Cellular and Molecular Life Sciences*, 77(2), 365–378. doi:https://doi.org/10.1007/s00018-019-03186-2

Boglárka Zámbó, Z. B. (2018). Clinically relevant mutations in the ABCG2 transporter uncovered by genetic analysis linked to erythrocyte membrane protein expression. *Scientific Reports*, 8(1), 7487. doi:https://doi.org/10.1038/s41598-018-25695-z

Cassandra White, R. J. (2021). Ethnic Diversity of DPD Activity and the DPYD Gene: Review of the Literature. *Pharmacogenomics and Personalized Medicine*, 14, 1603–1617. doi:https://doi.org/10.2147/PGPM.S337147

Francesco Iachetta, C. B. (2019). The clinical relevance of multiple DPYD polymorphisms on patients candidate for fluoropyrimidine based-chemotherapy. An Italian case-control study. *British Journal of Cancer*, 120(8), 834–839. doi:https://doi.org/10.1038/s41416-019-0423-8

Hannah H. Lee, a. R. (2017). Interindividual and interethnic variability in drug disposition: polymorphisms in organic anion transporting polypeptide 1B1 (OATP1B1; SLCO1B1). *British Journal of Clinical Pharmacology*, 83(6), 1176–1184. doi:https://doi.org/10.1111/bcp.13207

Kniffin, C. L. (2010). *DIHYDROPYRIMIDINE DEHYDROGENASE; DPYD*. Retrieved from Online Mendelian Inheritance in Man: https://omim.org/entry/612779

Linda M Henricks, C. A.-J. (2015). Translating DPYD genotype into DPD phenotype: using the DPYD gene activity score. *Pharmacogenomics*, 16(11):1277-86. doi:https://doi.org/10.2217/pgs.15.70

Linda M Henricks, C. A.-J. (2018). DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *Lancet Oncology*, 19(11):1459-1467. doi:https://doi.org/10.1016/S1470-2045(18)30686-7

Lingxin Zhang, V. S. (2020). CYP2C9 and CYP2C19: Deep Mutational Scanning and Functional Characterization of Genomic Missense Variants. *Clinical and Translational Science*, 13(4), 727–742. doi:https://doi.org/10.1111/cts.12758

Luis Ramudo-Cela, F. B.-F.-M.-M.-H. (2021). Results of the implementation of a pharmacogenomics platform based on NGS technologies. Combining clinical and research approaches. . *Farmacia Hospitalaria*, 45(Suppl 1):S11-37. doi: 10.7399/fh.11762

Maarten J Deenen, J. T.-J. (2011). Relationship between single nucleotide polymorphisms and haplotypes in DPYD and toxicity and efficacy of capecitabine in advanced colorectal cancer. *Clinical Cancer Research*, 17 (10): 3455–3468. doi:https://doi.org/10.1158/1078-0432.CCR-10-2209

Marzia Del Re, S. C. (2019). DPYD*6 plays an important role in fluoropyrimidine toxicity in addition to DPYD*2A and c.2846A>T: a comprehensive analysis in 1254 patients. *The Pharmacogenomics Journal*, 19(6), 556–563. doi:https://doi.org/10.1038/s41397-019-0077-1

Mun-Fai Loke, H. W.-L.-C. (2019). Deep sequencing analysis to identify novel and rare variants in pain-related genes in patients with acute postoperative pain and high morphine use. *Journal of Pain Research*, 12, 2755–2770. doi:https://doi.org/10.2147/JPR.S213869

Myung-Eui Seo, B.-J. M. (2023). Comprehensive in vitro and in silico assessments of metabolic capabilities of 24 genomic variants of CYP2C19 using two different substrates. *Frontiers in Pharmacology*, 14, 1055991. doi:https://doi.org/10.3389/fphar.2023.1055991

Nyarai D. Soko, C. M. (2011). A cost effective RFLP method to genotype Solute carrier organic anion 1B1 (SLCO1B1) c.1929A>C (p.Leu643Phe, rs34671512); a variant with potential effect on rosuvastatin pharmacokinetics. *BMC research notes*, 11(1), 384. doi:https://doi.org/10.1186/s13104-018-3469-4

Ontario-Health. (2021). DPYD Genotyping in Patients Who Have Planned Cancer Treatment With Fluoropyrimidines: A Health Technology Assessment. *Ontario Health Technology Assessment Series*, 21(14), 1–186.

Satoshi Matsusaka, H.-J. L. (2015). Pharmacogenomics of fluorouracil -based chemotherapy toxicity. *Expert Opinion on Drug Metabolism and Toxicology*, 11(5):811-21. doi:https://doi.org/10.1517/17425255.2015.1027684

Ursula Amstutz, S. F. (2009). Dihydropyrimidine dehydrogenase gene variation and severe 5-fluorouracil toxicity: a haplotype assessment. *Pharmacogenomics*, 10(6):931-44. doi:https://doi.org/10.2217/pgs.09.28