Functional genetic analysis of Single Nucleotide Polymorphisms in drug metabolism genes

Katherine G. Marlow[1], Abigail Weir[2], Salsabil Ahmed[3], Srikanta Kumar Rath[4], and Amit Kumar Mitra[5, 6]

[1] Undergraduate Student, Department of Biological Sciences, Auburn University
[2] Undergraduate Student, Department of Chemistry and Biochemistry, Auburn University
[3] Graduate Student, Harrison College of Pharmacy, Auburn University
[4] Division of Toxicology and Experimental Medicine, CSIR-Central Drug Research Institute, India
[5]Assistant Professor, Harrison College of Pharmacy, Auburn University
[6] Director, Center for Pharmacogenomics and Single -Cell Omics Initiative (AUPharmGx), Auburn University

Single Nucleotide Polymorphism (SNP) is a variation of a nucleotide at a single position in a DNA sequence. Certain SNPs are associated with disease and can be used to provide a better understanding of how mutations affect function in the body [1]. One of the major causes of inter-individual differences in drug effects is genetic variations in drug metabolism enzymes (DMEs) [2]. Certain mutations could lead to the development of adverse drug reactions, caused by the change in efficacy or toxicity of the drug. Having this genetic information will aid in prescribing the best treatment for a beneficial outcome while also avoiding unwanted effects in a particular patient in a timely manner (also known as Precision Medicine)[3]. The objective of this research is to understand the pharmacogenomic biomarkers underlying variable drug response and differences in drug metabolism genes. Our SNP discovery work focused on the Indian subpopulations that have not yet been explored. Mutational analysis was conducted to find the most damaging mutations. This determined the SNPs that would be most effective to study for functional analysis using bacteria transformation. The results of this research will help in understanding how different sub populations compared to that of the original Human Genome Project results. Furthermore, it will attempt to explain how these genetic differences can affect disease predisposition, drug reactions, and metabolism.

This project's aims are broken down into two main phases. In **Phase 1**, genotype and allele frequencies were calculated from the sequencing data that has been previously generated by the Mitra Lab. In **Phase 2**, the mutation types (coding or non-coding regulatory SNPs) and resulting amino acid changes were evaluated using databases like National Center for Biotechnology Information (NCBI) and Ensembl. SNPs were then shortlisted for functional analysis based on the predicted effect on protein function using PolyPhen-2.

[1]Corresponding author: kgm0045@auburn.edu

The results of this research will help in understanding how these genetic differences can affect drug metabolism and treatment outcomes, and also allow us to explore population differences in DMEs.

Following SNP discovery and validation in two separate datasets, information about each SNP's gene, chromosome position, alleles, clinical significance, gene consequence, flanks, and aggregate allele frequencies was collected from the NCBI SNP database (https://www.ncbi.nlm.nih.gov/snp/). The allele frequencies from reported SNP data were also calculated. The potential damaging effects of each SNP were evaluated using SIFT Prediction data (https://sift.bii.a-star.edu.sg/www/SIFT_dbSNP.html) and HumDiv and HumVar models from PolyPhen-2 data (http://genetics.bwh.harvard.edu/pph2/).

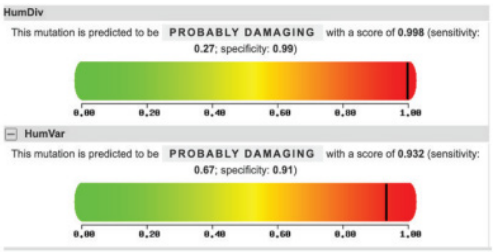


Fig. 1 Polyphen-2 data on SNP rs28399434 in CYP2A6, a drug metabolism gene.

Information gathered about the genes associated with the listed SNPs includes drug label annotations and clinical annotations from PharmGKB (https://www.pharmgkb.org/). PharmGKB data included drug molecules and phenotypes already associated with the SNPS. If available, CPIC guidelines were noted (https://cpicpgx.org/). Finally, the genes were searched on GeneCards for more information about drugs already associated with the SNPs (https://www.genecards.org/). The reference sequence of each gene was found using NCBI gene data (https://www.ncbi.nlm.nih.gov/snp/). Each gene was then annotated to find the start of the gene, the exon positions, and the mutation positions.

A condensed list of SNPs was created and was used for the focus of further study [4, 5]. The SNPs were chosen based on damaging HumDiv and HumVar models, as well as deleterious SIFT predictions. The allele frequencies were also considered when creating the condensed list. In addition, the importance of the gene that the SNP references were considered based on its relation to drug metabolism and other cellular mechanisms like DNA damage repair pathways [6].

Table 1 Information about SNPs chosen for further research.

SNP	Gene Name	Position	Consequence
rs28399434	CYP 2A6	chr19:40850414 (GRCh38.p13)	CYP2A6 : Missense Variant
rs1058930	CYP 2C8	chr10:95058362 (GRCh38.p13)	CYP2C8 : Missense Variant
rs2307186	XRCC1	chr19:43575439 (GRCh38.p13)	XRCC1: Missense Variant

All PCR were carried out using reagents from New England Biolabs (New England Biolabs, Ipswich, MA, USA). A 25 µL reaction mix was made by adding 2.5 µL 10X Taq Buffer, 1 µL dNTP solution mix, 1 µL forward primer, 1 µL reverse primer, 0.5 µL Taq DNA polymerase, 100 ng DNA sample, and 17 µL purified water. The optimal annealing temperature was predicted by inspecting the GC content of the primer sequence to determine the melting point of the primers. A gradient was chosen to include the estimated optimal annealing temperature and temperatures above and below the estimated optimal annealing temperature. The initial denaturation step was at 95 °C for 7 min and 34 cycles of PCR consisting of denaturation at 94 °C for 30 sec. For primers used in the present reaction, an annealing temperature gradient was made using temperatures of 57 °C, 57.9 °C, 59.2 °C, and 61.1 °C. The rest of the cycling PCR reaction had an extension step at 72 °C for 45 sec for 30 cycles, and a final extension step at 72 °C for 7 min.

The reaction mix for the PCR was prepared the same as in the Gradient PCR above, adjusting the annealing temperature to 57.9 °C. Reaction mixes were also made for various multiple myeloma cell line DNA samples including U266P, JIM3, U266VR, RPMI8226VR, MMISVR, and UTMC2.

Bidirectional DNA sequencing was performed on SeqStudio Genetic Analyzer System with SmartStart (Applied Biosystems, Foster City, USA). The DNA samples were cleaned and prepared using 2 µL ExoSAP-IT (Thermo Fisher Scientific, Waltham, MA, USA) and 5 µL of the sample. The sample was incubated at 37° C for 4 minutes, 80° C for 1 minute, and held at 4° C. The cycling sample was prepared using 3.5 µL purified H2O, 2 µL BigDye Terminator, 1.5 µL BigDye Buffer, 1 µL primer, and 2 µL cleaned PCR sample. The samples were mixed and thermal cycling was run. After the cycling was completed, 45 µL SAM solution, 10 µL of XTerminator solution, and 10 µL of the cleaned sample was added to the well. The plate was sealed and vortexed for 30 minutes at 2000 speed. The plate was then centrifuged for 2 minutes at 2000 RPM. Sequencing results were exported and viewed using DNASTAR software (DNASTAR, Madison, WI, USA).

Bioinformatics analysis was done to analyze the frequencies of the SNPs between different Indian subpopulations. The first SNP analyzed was rs28399434, which was the original CYP2A6 mutation that was predicted to be damaging within a population and an individual. SNPs rs4997557 and rs8192730 were other mutations found after sequencing that warrant further research.

The results obtained from this dual-staged SNP exploration study exemplify the uniqueness of the Indian subpopulation clusters with respect to certain mutations, genotypes, and allele frequency patterns. It also

shows the variation between different subdivisions of the country. These findings underscore the fact that the Indian population must be investigated for its plausible existence as a separate entity from the commonly inferred major global population clusters. The data generated from this study may have wide-ranging applications for further epidemiological and public health-related research on the Indian population.

In addition, several SNPs within our genes of interest show a predicted damaging effect on an individual and require functional validation within in vitro models, that we plan to do next using myeloma cell lines mentioned above. Continued research will include site-directed mutagenesis and transfection into mammalian cells. Additionally, these mutations may affect drug metabolism and require further functional study using a reference in-vivo model system.

Statement of Research Advisor

Katie's research involves the functional analysis of key drug metabolizing gene mutations using two parallel approaches: Bioinformatic analysis of high-risk variations, and recombinant DNA technologies. We are planning to further this work with a particular focus on world populations and under-served subpopulations.
-Amit K. Mitra, Harrison College of Pharmacy

References

[1] Grierson G. A. 1927 "A linguistic survey of India." Superintendent of Government Printing, Calcutta, India.

[2] Meyer, Urs A., and Ulrich M. Zanger. "Molecular Mechanisms of Genetic Polymorphisms of Drug Metabolism." Annual Review of Pharmacology and Toxicology, vol. 37, no. 1, 1997, pp. 269–296., doi:10.1146/annurev.pharmtox.37.1.269.

[3] Ahmed, Shabbir, et al. "Pharmacogenomics of Drug Metabolizing Enzymes and Transporters: Relevance to Precision Medicine." Genomics, Proteomics & Bioinformatics, vol. 14, no. 5, 2016, pp. 298–313., doi:10.1016/j.gpb.2016.03.008.

[4] Indian Genome Variation Consortium. "Genetic landscape of the people of India: a canvas for disease gene exploration." J. Genet., vol. 87, 2008, pp. 3-20

[5] Mitra AK, Singh A; Indian Genome Variation Consortium; Rath SK. "LIG1 polymorphisms: the Indian scenario." J. Genet. 2014 Aug;93(2):459-69. doi: 10.1007/s12041-014-0415-7

[6] Mappangara I, Yusuf I, Aspar Mappahya A, Qanitha A. "CYP2A6 gene polymorphism and severity of coronary atherosclerosis in Indonesian male smokers: A pilot study." Medicine (Baltimore). 2022 Sep 16;101(37):e30308. doi: 10.1097

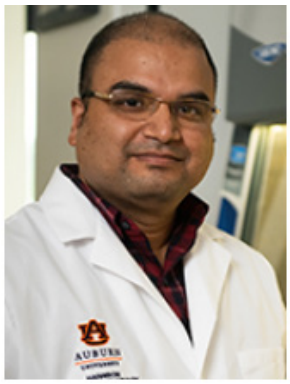
Authors Biography

Katherine G. Marlow is a junior-year student pursuing a B.S. degree in Genetics at Auburn University. She has played key research roles in understanding pharmacogenomic biomarkers in drug metabolism genes and DNA repair genes.

Abigail Weir is a junior-year undergraduate pursuing a B.S. degree in Biochemistry at Auburn University. She has performed mutational analysis of mutations in DNA repair genes such as LIG1 and MSH2.

Salsabil Ahmed is a graduate student pursuing an M.S. degree in Pharmacy at Auburn University. She works as a Graduate Research Assistant for Center for Pharmacogenomics and Single-Cell Omics (AUPharmGx), in the Drug Discovery and Development Harrison College of Pharmacy.

Dr. Srikanta K. Rath is a Chief Scientist at, the Division Of Toxicology & Experimental Medicine, CSIR-Central Drug Research Institute, India.



Dr. Amit K. Mitra is an Assistant Professor at the Harrison College of Pharmacy at Auburn University. He is also the founding Director of the Center for Pharmacogenomics and Single-Cell Omics (AU-PharmGx). Dr. Mitra's current research involves integrating single-cell multi-omics with in vitro drug response modeling, and functional genomics with translational bioinformatics to investigate inter-tumor and intra-tumor (subclonal) heterogeneity in human cancers.