PRECISION HEALTH
UNIVERSITY OF MICHIGAN

MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN
DATA OFFICE FOR CLINICAL & TRANSLATIONAL RESEARCH

Pharmacogenomic Star Allele and Activity Phenotype Inferences for the Michigan Genomics Initiative Biobank

About this guide

This guide provides information about the production and evaluation of pharmacogenomic star allele and activity phenotype inferences in the Michigan Genomics Initiative (MGI) biobank.

High level summary

The star allele nomenclature system describes polymorphism in pharmacologically important genes and can inform on gene activity. The MGI inferred star alleles and activity phenotypes for 56,984 MGI participants in freeze 3. These inferences were made based on genotype array data using the state-of-the-art "Stargazer" software[1].

Example Use Case

An investigator could use star alleles and activity phenotypes to test for pharmacogenomic relationships in the MGI.

What are star alleles and activity phenotypes?	The star allele system is a nomenclature system that is used to summarize variants into alleles for different pharmacogenes, which are genes that interact with pharmaceuticals. Star alleles can be translated to pharmacogene activity phenotypes (e.g. poor metabolizer, ultrarapid metabolizer, etc.) based on definitions that are available from pharmacogenomics consortia[2].
Which pharmacogenes are included in these data?	Star alleles and activity phenotypes are inferred for the following pharmacogenes: SULT1A1, CYP2R1, CYP26A1, GSTP1, GSTM1, VKORC1, CYP1A1, CYP4A11, CYP4B1, CYP1A2, NAT2, ABCB1, CYP1B1, CYP4A22, NAT1, SLCO1B3, CYP2D6, CYP2E1, IFNL3, UGT1A4, NUDT15, CYP2A6, CYP2A13, CYP2F1, CYP2W1, SLC15A2, CYP2S1, UGT1A1, CYP3A5, CYP3A4, CYP2J2, XPC, UGT2B7, CYP3A43, TPMT, CYP4F2, CYP19A1, CYP3A7, UGT2B15, CYP17A1, DPYD, CYP2C8, SLC22A2, SLCO2B1, CYP2B6, POR, CYP2C9, CACNA1S, PTGIS, CFTR, CYP2C19, SLCO1B1, TBXAS1, RYR1
How are star alleles inferred for participants of the MGI?	SNP array data imputed with the Haplotype Reference Consortium r1.1 or Trans-Omics for Precision Medicine r2 panels are used as input for the star allele inference software Stargazer[1,3,4]. As a quality control measure, poorly imputed variants with an estimated $R^2 < 0.3$ are removed from the input data. Star allele inferences that are made from data where no imputation quality filters were applied are available upon special request.
How are activity phenotypes inferred for participants of the MGI?	To infer activity phenotypes, star allele inferences are translated using diplotype to phenotype tables accessed from PharmGKB.org[2]. For star alleles where PharmGKB diplotype to phenotype mappings are not available, activity phenotypes are inferred by the Stargazer software. At present, the activity phenotype mappings that are described by the tables available from PharmGKB are preferred as PharmGKB is a longstanding, curated resource for pharmacogenomics.
How were these data evaluated?	Concordance between star allele inferences made by commercially available clinical pharmacogenomics tests and Stargazer were evaluated for a small subset of MGI participants. The concordance tests were performed for CYP2B6, CYP2C9, CYP3A4, CYP2D6, CYP1A2, CYP2C19, TPMT, UGT1A4, UGT2B15, and UGT1A1. A star allele was considered concordant if the same unambiguous star allele was both called by a clinical pharmacogenomic test and inferred by Stargazer. Using this definition, we found the mean concordance rate to be 96% with rates ranging from 77 to 100%. Tables that describe these concordance evaluations are available upon request.

Are there limitations/considerations for these data?	For any given gene, some star alleles may not be inferred in the MGI data set. The total number of star alleles that can be interrogated for each gene is limited to the variants that were either directly typed on the SNP array or were imputed with an estimated $R^2 > 0.3$. Star alleles that are defined at least in part by rare variants are more likely to be not inferred in the MGI data set. A table that describes which star alleles are interrogated for each gene is available upon request. This table indicates where important star allele inferences may be missed in MGI. Star alleles that are defined by duplications, deletions, or gene fusions are currently not inferred in the MGI data set. Genes that are characterized by extensive structural variation, such as CYP2D6, may result in incorrect star allele and activity phenotype inferences in MGI. Using this MGI service to analyze star alleles for genes where structural variation is not uncommon may require careful consideration. Star allele and activity phenotype inferences for any given MGI data freeze may become outdated over time. Both star allele and activity phenotype inferences for the MGI cohort are updated only upon release of a new freeze.

References

1. Lee, S. *et al.* Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **21**, 361–372 (2019).
2. Whirl-Carrillo, M. *et al.* Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* **92**, 414–417 (2012).
3. TOPMed Imputation Server. https://imputation.biodatacatalyst.nhlbi.nih.gov/#!pages/about.
4. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).