Variant Annotations Help File

The set of files comprised of **var_pheno_ann.tsv**, **var_drug_ann.tsv**, **var_fa_ann.tsv** and **study_parameters.tsv** contain PharmGKB's variant annotations and associated information. Variant annotations are separated into 3 files because PharmGKB has 3 different standardized templates to capture information depending on the specifics of the association. Each file has slightly different fields, though some fields are common to all 3. Please refer to the PharmGKB website for more information about variant annotations, and how they are used to create clinical annotations based on score.

It is important to understand that the PharmGKB curators routinely review several high profile journals for articles to curate. There may be more literature in the public domain to support or contradict an association that is not in the PharmGKB database. PharmGKB does its best to manually curate high profile literature but does not contain curated literature from every domain-based journal, or all of PubMed. PharmGKB reviews evidence from curated literature in non-regular intervals and re-evaluates the evidence strength for each association as more literature becomes available.

Description of Files:

- **var_pheno_ann.tsv**: Contains associations in which the variant affects a phenotype, with or without drug information.
- **var_drug_ann.tsv**: Contains associations in which the variant affects a drug dose, response, metabolism, etc
- **var_fa_ann.tsv**: Contains *in vitro* and functional analysis-type associations.
- **study_parameters.tsv**: Contains information about the study population size, biogeographical group and statistics for the variant annotations; this file is cross-referenced against the 3 variant annotation files.
- **LICENSE.txt**: The PharmGKB license for using PharmGKB data, including clinical annotations.
- **CREATED_xxxx-xx-xx.txt**: This file indicates the date that all files in this group were created from the database.
- **README.pdf** file: This document.

A description of the fields in each file follows.

var_drug_ann.tsv:

- ***Variant Annotation ID***: unique ID number for each variant/drug annotation
- ***Variant/Haplotypes***: dbSNP ID or haplotype(s)
- ***Gene***: HGNC symbol
- ***Drug(s)***: Drug name
- ***PMID***: PubMed identifier
- ***Phenotype Category***: options [efficacy, toxicity, dosage, metabolism/PK other]
- ***Significance***: yes or no – determined by if the author stated the association was significant
- ***Notes:*** curator notes field
- ***Sentence***: structured sentence
- ***Alleles***: variant alleles in annotation
- ***Specialty Population***: tags for any special populations this annotation is relevant to (e.g. pediatric)

var_fa_ann.tsv:

- ***Variant Annotation ID***: unique ID number for each variant/drug annotation
- ***Variant/Haplotypes***: dbSNP ID or haplotype(s)
- ***Gene***: HGNC symbol
- ***Drug(s)***: Drug name
- ***PMID***: PubMed identifier
- ***Phenotype Category***: options [efficacy, toxicity, dosage, metabolism/PK other]
- ***Significance***: yes or no – determined by if the author stated the association was significant
- ***Notes***: curator notes field
- ***Sentence***: structured sentence
- ***Alleles***: variant alleles in annotation
- ***Specialty Population***: tags for any special populations this annotation is relevant to (e.g. pediatric)

var_pheno_ann.tsv:

- ***Variant Annotation ID***: unique ID number for each variant/drug annotation
- ***Variant/Haplotypes***: dbSNP ID or haplotype(s)
- ***Gene***: HGNC symbol
- ***Drug(s)***: Drug name
- ***PMID***: PubMed identifier
- ***Phenotype Category***: options [efficacy, toxicity, dosage, metabolism/PK other]
- ***Significance***: yes or no – determined by if the author stated the association was significant
- ***Notes***: curator notes field
- ***Sentence***: structured sentence
- ***Alleles***: variant alleles in annotation
- ***Specialty Population***: tags for any special populations this annotation is relevant to (e.g. pediatric)

study_parameters.tsv:

- ***Study Parameters ID***: Unique ID number for each "Study Parameters" entry
- ***Variant Annotation ID***: the ID number of the Variant Annotation the Study Parameters is associated with
- ***Study Type***: options [cohort, case/control, case series, cross-sectional, clinical trial, meta-analysis, GWAS, replication, prospective, retrospective, linkage, trios
- ***Study Cases***: number of cases in the paper
- ***Study Controls***: number of controls in the paper (with genotypes and phenotypes, used in the association analysis)
- ***Characteristics***: free text where the curator can record gender, disease, age group or other distinguishing characteristics about the group studied
- ***Characteristics Type***: options [disease, drug, age group, gender, study cohort]
- ***Frequency in Cases***: allele frequency in the cases
- ***Allele of Frequency in Cases***: which allele the above frequency is for
- ***Frequency in Controls***: allele frequency in controls
- ***Allele of Frequency in Controls***: which allele the above frequency is for
- ***P Value***: the operator (=, <, etc...) and p-value that is reported in the paper
- ***Ratio Stat Type***: options [OR, RR, HR]
- ***Ratio Stat:*** free text number that pairs with above field

- ***Confidence Interval Start***: free text number
- ***Confidence Interval Stop***: free text number
- ***Biogeographical Groups***: population groups from PMID:30506572. More information found at https://www.pharmgkb.org/page/biogeographicalGroups. options [African America/Afro-Caribbean, American, Central/South Asian, East Asian, European, Latino, Near Eastern, Oceanian, Sub-Saharan African, Unknown, Multiple Groups]