Variant Annotations Help File

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. All 3 files contain variant/drug associations:

1. The **var_pheno_ann.tsv** file contains associations in which the variant affects a phenotype, with or without drug information.
2. The **var_drug_ann.tsv** file contains associations in which the variant affects a drug dose, response, metabolism, etc
3. The **var_fa_ann.tsv** file contains in vitro and functional analysis-type associations.

A fourth file, **study_parameters.tsv**, is cross-referenced against the other files and contains information about the study population size, race and statistics for the variant annotations.

It is important to understand that the PharmGKB curators routinely review several high profile journals for articles to curate. However, there may be more literature in the public domain to support or contradict a pharmacogenetic association that has not been curated by PharmGKB db. 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.

A description of the fields in each file follows.

1. **var_drug_ann.tsv** file contains:

- ***Annotation ID***: unique ID number for each variant/drug annotation
- ***Variant***: dbSNP ID or haplotype
- ***Gene***: HGNC symbol (PharmGKB Accession ID)
- ***Chemical***: Drug name (PharmGKB Accession ID)
- ***Literature ID***: PMID
- ***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
- ***StudyParameters***: corresponds with the Study Parameters ID in the study_parameters.tsv file
- ***Alleles***: variant alleles in annotation

2. **var_fa_ann.tsv** file contains:

- ***Annotation ID***: unique ID number for each variant/drug annotation
- ***Variant***: dbSNP ID or haplotype
- ***Gene***: HGNC symbol (PharmGKB Accession ID)
- ***Chemical***: Drug name (PharmGKB Accession ID)
- ***Literature ID***: PMID
- ***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
- ***StudyParameters***: corresponds with the Study Parameters ID in the study_parameters.tsv file
- ***Alleles***: variant alleles in annotation

3. **var_pheno_ann.tsv** file contains:

- ***Annotation ID***: unique ID number for each variant/drug annotation
- ***Variant***: dbSNP ID or haplotype
- ***Gene***: HGNC symbol (PharmGKB Accession ID)
- ***Chemical***: Drug name (PharmGKB Accession ID)
- ***Literature ID***: PMID
- ***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
- ***StudyParameters***: corresponds with the Study Parameters Id in the study_parameters.tsv file
- ***Alleles***: variant alleles in annotation

4. **study_parameters.tsv** file contains:

- ***Study Parameters Id***: Unique ID number and pairs with column "StudyParameters" in the other files
- ***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 Operator***: options [=, <]
- ***P Value***: free text and is the p-value that is reported in the paper
- ***Ratio Stat Type***: option [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
- ___Race(s)___: OMB+Race(s), Unknown, Mixed Population or Other. OMB: [American Indian or Alaskan Native, Asian, Black or African American, Hispanic or Latino (not really OMB Race), Native Hawaiian or Other Pacific Islander, White]

See Office of Management and Budgets for definitions of OMB races.