Clinical Annotations Help File

The pair of files, **clinical_ann.tsv** and **clinical_ann_metadata.tsv**, contain PharmGKB's Clinical Annotations. These annotations are manually created by the PharmGKB curators to provide an evidence-rated, genotype-based summary for a particular pharmacogenetic *variant + drug pair*. The objective is to present a succinct clinical interpretation summarizing the literature evidence for an association between a genetic variant and a drug.

Description of Files:

In the **clinical_ann.tsv file**, the IDs in the "Genotype-Phenotype ID" column match up with the "Genotype-Phenotypes IDs" in the **clinical_ann_metadata.tsv** file. The **clinical_ann.tsv** file has the actual annotations; the **clinical_ann_metadata.tsv** file has the annotations and all of the other data such as the specific rsID, drug(s) and genes in each annotation. It also contains information about the race that the association was found in, if applicable/known. There may also be information about associated disease phenotype(s) when applicable (e.g. the drug/variant association may be applicable to patients with diabetes). Finally, each clinical annotation has an **"Evidence Strength"** , which ranks the level of evidence in the literature curated by PharmGKB for the association.

A description of the fields in each file follows.

clinical_ann.tsv file contains:

- ***Genotype-Phenotype ID***: numbers correspond with those in the "Genotype-Phenotypes IDs" column of the clinical_ann_metadata.tsv file.
- ***Genotype***: the genotype associated with the clinical phenotype in the next column.
- ***Clinical Phenotype***: the clinical annotation for the given genotype.

Example row from allele_phenotype.tsv file:

Genotype-Phenotype ID	Genotype	Clinical Phenotype
613979022	CC	May be less likely to have improved left ventricular ejection fraction after carvedilol treatment.

clinical_ann_metadata.tsv file contains:

- ***Clinical Annotation Id***: unique ID number for each variant/drug annotation.
- ***Location***: the variant location. the most precise location available is used. In order of preference, it will use the RSID from dbSNP, the chromosomal location, the RefSeq Accession ID and location, or haplotype name.
- ***Gene***: HGNC gene symbol, followed by unique PharmGKB identifier for gene in parentheses.
- ***Level of Evidence***: see documentation for definitions below.
- ***Clinical Annotation Types***: tags - curators select from [Dosage, Efficacy, Toxicity/ADR, Other].
- ***Genotype-Phenotypes IDs***: numbers correspond with the **clinical_ann.tsv** file.
- ***Annotation Text***: The clinical annotation for each genotype.
- ***Variant Annotations IDs***: unique ID numbers for PharmGKB variant annotations used to support the clinical

annotation.

- ***Variant Annotations***
- ***PMIDs***: the PMIDs corresponding to the Variant Annotations. Note that a PMID may be repeated in this field if more than one Variant Annotation was written for it.
- ***Evidence Count***: the number of supporting PMIDs.
- ***Related Drugs:*** drug(s) associated with the variant in the annotation; from the PharmGKB drug vocabulary.
- ***Related Diseases***: associated disease phenotype, where applicable. For example, if the variant/drug association was found in patients with a particular phenotype (disease), or if the variant/drug combination causes a particular phenotype. Terms come from the PharmGKB disease vocabulary (derived from MeSH).
- ***Race***: if the variant/drug association was found in one particular race, it is designated here, using the OMB standard categories.

Levels of Evidence:

PharmGKB curators use specific criteria to assess the PharmGKB "variant annotations" and determine the level of evidence.

Level	Criteria
1A	Annotation for a variant-drug combination in a CPIC or medical society-endorsed PGx guideline, or implemented at a PGRN site or in another major health system. **Example:** rs1800460 in TPMT (TPMT*3B) and thiopurines. This association is published as a CPIC guideline and used in multiple clinics.
1B	Annotation for a variant-drug combination where the preponderance of evidence shows an association. The association must be replicated in more than one cohort with significant p-values, and preferably will have a strong effect size. **Example:** rs1801133 in mthfr and methotrexate. PharmGKB has multiple articles for this association with significant P values and several with high odds ratios.
2A	Annotation for a variant-drug combination that qualifies for level 2B where the variant is within a VIP (Very Important Pharmacogene) as defined by PharmGKB. The variants in level 2A are in known pharmacogenes, so functional significance is more likely. **Example:** rs12248560 in CYP2C19 and omeprazole. PharmGKB has multiple articles for this association and it qualifies for level 2 and is in a pharmacogene.
2B	Annotation for a variant-drug combination with moderate evidence of an association. The association must be replicated but there may be some studies that do not show statistical significance, and/or the effect size may be small. **Example:** rs2234922 in EPHX1 and carbamazepine. PharmGKB contains 3 articles, 2 with significant p-values, 1 with no p-value reported; population sizes ranging from 70 to 234; no odds ratios reported.
3	Annotation for a variant-drug combination based on a single significant (not yet replicated) or annotation for a variant-drug combination evaluated in multiple studies but lacking clear evidence of an association. **Example:** rs993648 in CERKL and iloperidone. PharmGKB contains a GWAS article reporting a statistically significant association but it is unreplicated.
4	Annotation based on a case report, non-significant study or in vitro, molecular or functional assay evidence only. **Example:** rs61750900 in UGT2B10 and nicotine. PharmGKB contains 2 in vitro studies for this association.

It is important to understand that these criteria are applied only to literature that has been curated by PharmGKB. There may be more literature in the public domain to support or contradict an association that is not in the 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. PharmGKB will review the evidence from curated literature in the db in non-regular intervals and re-evaluate the evidence strength for each association as more literature becomes available.