


Variant Analysis



VCF - Variant Call Format

bcftools converts 23 and Me tsu to VCF

Pharm GKB - curates knowledge about the impact of genetic variation on drug response

Genotype - a set of genes

Phenotype - observable characteristics due to genes

Pharm GKB annotations report the association between a variant and a drug phenotype. Annotations are created manually by curators.

Examples:

Variant	Molecules	Sentence
rs38...	atorvastatin	Genotype GG is associated with ... more so than Genotype TT is

Variant annotations Folder:

1. var-pheno-ann.tsv (variant affects a phenotype)
2. var-drug-ann.tsv (variant affects drug response)
3. var-fa-ann.tsv (in vitro & functional analysis)
4. study-parameters.tsv (cross references other files, info about studies)

Given a VCF file:

- Search for variants that exist in var-pheno-ann
- Parse the sentences for certain words
- consider P-value?
- Generate report that consolidates the info from sentences

variants format

- rs ### . . . (sentences contain A, G, T, C Alleles)
- (CAPITAL)* ## (sentences don't mention Alleles)

Sentence format

for rs ### variant: Allele A is (not) associated with
- .. as compared to Allele C