

File descriptions

[1] ukb23145_300k_QQFE.annotations.txt.gz

File contains variant annotations derived from snpEff using the Ensembl v85 gene definitions to determine their functional impact on transcripts and genes:

- Each row contains an annotation for a given variant and gene pair.
- First column contains the variant identifier using the chromosome and position of the variant (based on GRCh38) as well as the reference and alternate allele (CHR:POS:REF:ALT).
- Second column contains the gene identifier using the gene name and Ensembl gene ID.
- Third column contains the annotations derived from snpEff effect predictions where, for each gene, the predictions involving protein-coding transcripts with an annotated start and stop were rolled up into a single functional impact prediction. This was done by selecting the most deleterious functional effect class. In addition, missense predicted variants were further ranked for deleteriousness using 5 scoring algorithms (based on dbNSFP v3.2).

[2] ukb23145_300k_QQFE.sets.txt.gz

File contains the list of gene sets based on the annotation file above:

- Each row contains a list of variants assigned to a gene.
- First column contains the gene identifier (corresponding to that in the annotation file above).
- Second column contains the chromosome for the gene.
- Third column contains a bp position assigned to the gene.
- Fourth column contains the list of variants which constitute the gene set.

[3] ukb23145_300k_QQFE.masks

File contains the definition for a single mask (M1) consisting of loss-of-function (LoF) annotated variants.

[4] ukb23145_300k_QQFE.90pct10dp_gc_variants.txt

File contains the variants failing the "90pct10dp" depth filter in the variant identifier format (CHR:POS:REF:ALT) in the PLINK file.

[5] `ukb23145_300k_OQFE.variant_ID_mappings.txt`

File contains the mappings of variant IDs in pVCF and PLINK in two tab-delimited columns with a header line. It is provided to allow tracing variants in PLINK back to the original line in pVCF.

- Column #1 contains variant IDs in the original pVCF.
- Column #2 contains variant IDs in the PLINK.