ensembl\_queries

# Sample queries for ensembl gene annotations

## Find all variants in ensembl gene regions

The following query can be used to determine what gene regions a variant lies in and to create a subset for downstream annotation for finding amino acid changes.

Update or remove the following variables, as needed:  
- vcf.filter: filter out variants that did not pass filtering by the variant caller  
- vcf.qual: filter out variants below a certain quality score

SELECT vcf.sample\_id as vcf\_sample\_id, vcf.chromosome as vcf\_chrom, vcf.position as vcf\_pos,vcf.ref as vcf\_ref, vcf.alt as vcf\_alt, vcf.id as vcf\_rsID, ens.start as ensembl\_start, ens.stop as ensembl\_end, ens.feature, ens.gene\_name as ensembl\_gene\_name, ens.gene\_id as ensembl\_geneid, ens.gene\_biotype as ensembl\_gene\_biotype, ens.transcript\_name as ensembl\_tx\_name,ens.transcript\_id as ensembl\_trans\_id, ens.exon\_id as ensembl\_exonid, ens.strand as ensembl\_strand  
FROM p7\_ptb.illumina\_variant as vcf, public\_hg19.ensembl\_genes as ens   
WHERE vcf.filter = "PASS"   
WHERE vcf.qual > 100  
AND vcf.chromosome = ens.chromosome   
AND vcf.position BETWEEN ens.start AND ens.stop

This query is equivalent to:

SELECT vcf.sample\_id as vcf\_sample\_id, vcf.chromosome as vcf\_chrom, vcf.position as vcf\_pos,vcf.ref as vcf\_ref, vcf.alt as vcf\_alt, vcf.id as vcf\_rsID, ens.start as ensembl\_start, ens.stop as ensembl\_end, ens.feature, ens.gene\_name as ensembl\_gene\_name, ens.gene\_id as ensembl\_geneid, ens.gene\_biotype as ensembl\_gene\_biotype, ens.transcript\_name as ensembl\_tx\_name,ens.transcript\_id as ensembl\_trans\_id, ens.exon\_id as ensembl\_exonid, ens.strand as ensembl\_strand  
FROM p7\_ptb.illumina\_variant as vcf  
JOIN  
public\_hg19.ensembl\_genes as ens   
ON vcf.chromosome = ens.chromosome   
WHERE vcf.filter = "PASS"  
AND vcf.qual > 100  
AND vcf.position BETWEEN ens.start AND ens.stop

## Find specific genes in the ensembl gene subset created above

Update or remove the following variables, as needed:  
- vcf.filter: filter out variants that did not pass filtering by the variant caller  
- vcf.qual: filter out variants below a certain quality score  
- ens.gene\_name: enter each gene of interest as a list, with each gene in quotes, comma-separated

SELECT vcf.sample\_id as vcf\_sample\_id, vcf.chromosome as vcf\_chrom, vcf.position as vcf\_pos,vcf.ref as vcf\_ref, vcf.alt as vcf\_alt, vcf.id as vcf\_rsID, ens.start as ensembl\_start, ens.stop as ensembl\_end, ens.feature, ens.gene\_name as ensembl\_gene\_name, ens.gene\_id as ensembl\_geneid, ens.gene\_biotype as ensembl\_gene\_biotype, ens.transcript\_name as ensembl\_tx\_name,ens.transcript\_id as ensembl\_trans\_id, ens.exon\_id as ensembl\_exonid, ens.strand as ensembl\_strand  
FROM p7\_ptb.illumina\_variant as vcf, public\_hg19.ensembl\_genes as ens   
WHERE vcf.filter = "PASS"   
AND vcf.qual > 100  
AND vcf.chromosome = ens.chromosome   
AND vcf.position BETWEEN ens.start AND ens.stop   
AND ens.gene\_name IN ("RMRPP1","PPIAP13","NDST2","RP11-574K11.8","RPL39P25")

## Find a specific gene in the ensebml gene subset

Update or remove the following variables, as needed:  
- vcf.filter: filter out variants that did not pass filtering by the variant caller  
- vcf.qual: filter out variants below a certain quality score  
- ens.gene\_name: enter a gene of interest in quotes

SELECT vcf.sample\_id as vcf\_sample\_id, vcf.chromosome as vcf\_chrom, vcf.position as vcf\_pos,vcf.ref as vcf\_ref, vcf.alt as vcf\_alt, vcf.id as vcf\_rsID, ens.start as ensembl\_start, ens.stop as ensembl\_end, ens.feature, ens.gene\_name as ensembl\_gene\_name, ens.gene\_id as ensembl\_geneid, ens.gene\_biotype as ensembl\_gene\_biotype, ens.transcript\_name as ensembl\_tx\_name,ens.transcript\_id as ensembl\_trans\_id, ens.exon\_id as ensembl\_exonid, ens.strand as ensembl\_strand  
FROM p7\_ptb.illumina\_variant as vcf, public\_hg19.ensembl\_genes as ens   
WHERE vcf.filter = "PASS"   
AND vcf.qual > 100  
AND vcf.chromosome = ens.chromosome   
AND vcf.position BETWEEN ens.start AND ens.stop   
AND ens.gene\_name IN ("RMRPP1")