Annovar

Annovar is installed on fourierseq at /opt/annovar. Annovar is a useful tool for functional annotation of SNPs. It will annotate whether the SNP is known/novel, whether it is an exonic region or otherwise, the amino acid changes it would cause and scores to determine the damage potential of the SNP.

To generate a summary annovar report on your SNPs, using Human hg19 genome, follow the steps below:

1. Convert vcf SNP file to annovar format

   ```
   convert2annovar.pl -format vcf4 --includeinfo out.vcf > out.vcf.annovar
   ```

2. Prepare database for hg19 by downloading various annotation files.

   ```
   annotate_variation.pl -downdb -buildver hg19 gene hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 band hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 1000g2010nov hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 snp132 hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 avsift hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 mce46way hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 ljb_pp2 hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 ljb_lrt hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 ljb_gerp++ hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 ljb_phylop hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 ljb_sift hg19_annovar_db/
   annotate_variation.pl -downdb -buildver hg19 ljb_all hg19_annovar_db/
   ```

3. Run `summarize_annovar` to annotate your SNPs

   ```
   summarize_annovar.pl out.vcf.annovar hg19_annovar_db/ --buildver hg19
   --verdbsnp 132 --ver1000g 1000g2010nov --genetype refgene
   ```

4. Output files *exome_summary.csv and *genome_summary.csv contain summary annotation for all SNPs in exonic regions and in non-exonic (intergenic, intronic and UTR regions) respectively.

   Annovar can also accept custom annotation files in the form of gff3, vcf and bed formats. For gff3 files, annovar requires that all lines have an ID.
tag (which isn't a standard in many gff files and so may need manipulation).