ClinVar is a repository of clinically relevant genome variant annotations. Information in the repository is available through the NCBI webpage or by downloads in XML or VCF format. The ClinVar VCF, which represents population-scale variant information, can be used to annotate a local VCF containing data from individual samples, using software such as ‘vcfanno’ (<https://doi.org/10.1101/041863)>. However, the ClinVar VCF 1.0 format has a problem that can throw kinks into an annotation.

Here’s what you want:

ClinVar VCF:

chr pos ref alt info

7 879832 A C,G,T GENEINFO=TP53;CLNALLE=1,2,3;CLNSIG=5,0,2

Your local VCF, before you’ve annotated it with the ClinVar VCF:

chr pos ref alt

CLNALLE tells you alleles for which there is info in subsequent INFO tags.

CLNSIG denotes clinical significance; 5 = Pathogenic, 0 = Uncertain, 2 = Benign

7 879832 A C

7 879832 A G

=====================================================

Your local VCF, after you’ve annotated it with the ClinVar VCF:

chr pos ref alt info

7 879832 A C GENEINFO=TP53; CLNSIG=5

7 879832 A G GENEINFO=TP53; CLNSIG=0

=====================================================

Here’s the problem:

Sometimes there is no CLNSIG information for the first allele and annotation gets out of sync:

ClinVar VCF:

chr pos ref alt info

7 879832 A C,G,T GENEINFO=TP53;CLNALLE=2,3;CLNSIG=0,2

Your local VCF, before you’ve annotated it with the ClinVar VCF:

chr pos ref alt

7 879832 A C

7 879832 A G

=====================================================

Your local VCF, after you’ve annotated it with the ClinVar VCF:

chr pos ref alt info

7 879832 A C GENEINFO=TP53; CLNSIG=0

7 879832 A G GENEINFO=TP53; CLNSIG=2

=====================================================

Steps:

1) Figure out how to decompose and annotate a VCF file (containing data from indiv samples) with ClinVar VCF. Identify and show instances of the problem.

2) Use Python to parse ClinVar VCF and correct the error that causes the problem.

3) Illustrate, with examples, how the problem has been fixed.