# Allele Frequency Data Speciifcation

The below allele frequency fields are already included in the built_with_change_types.tsv file. None need to be removed.

24. Variant_frequency_LOVD
25. Minor_allele_frequency_ESP
26. EUR_Allele_frequency_1000_Genomes
27. AFR_Allele_frequency_1000_Genomes
28. AMR_Allele_frequency_1000_Genomes
29. EAS_Allele_frequency_1000_Genomes
30. Allele_frequency_1000_Genomes
31. SAS_Allele_Frequency_1000_Genomes
32. Allele_frequency_ExAC

For field number 25, the ESP data will be converted from percentage to decimal format for consistency with other frequency columns.

All fields from 1000 Genomes necessary for visualization are present and correctly formatted.

There are additional fields from ExAC that will be pulled through the pipeline:

| AC_AFR | AN_AFR | Het_AFR | Hom_AFR |
| AC_AMR | AN_AMR | Het_AMR | Hom_AFR |
| AC_EAS | AN_EAS | Het_EAS | Hom_EAS |
| AC_FIN | AN_FIN | Het_FIN | Hom_FIN |
| AC_NFE | AN_NFE | Het_NFE | Hom_NFE |
| AC_OTH | AN_OTH | Het_OTH | Hom_OTH |
| AC_SAS | AN_SAS | Het_SAS | Hom_SAS |

This data can be found in the "exec.brca12.sorted.hg38.vcf" file present in the output. AC is short for "Allele Count," which is the number of alleles found in a population. AN is short for "Allele Number," which is the total number of sample alleles. "Het" and "Hom" are short for homozygosity, which gives the user the ability to understand the allele counts in terms of individuals in the population.

Allele frequency is calculated by dividing AC/AN, which can be done on the fly for the Allele Frequency Tile. Accordingly, ACs and ANs are sufficient for the pipeline. Overall AC can also be calculated by summing together all ACs of the populations. Overall AN can also be calculated this way.

Rergarding AF discrepancies, the overall Allele Frequency calculated by ExAC sometimes uses data that BRCA Exchange does not pull through its pipeline.

Further information about ExAC data can be found at:
<http://exac.broadinstitute.org/faq>