**Post Imputation Analysis**

**Chromosome chosen:** Chr 22

**Number of SNPs prior to imputation:** 16, 218

**Number of SNPs post imputation (without any filtering) :** 775,275

**Number of SNPs after using an R2 > 0.8 and MAF > 0.5 :** 124,700

**Plotted allele frequency of a random subset of the output dose file.**

**Chart, histogram

Description automatically generated**

Also plotted the MAF and the Rsq for all the imputed SNPs as shown below: (This is the raw output before filtering)

Chart

Description automatically generated with medium confidence

Chart, histogram

Description automatically generated

Other effective methods of quality estimation which I found cited by various papers which we could use include:

1. Concordance score
2. Proportion of non-missing genotypes under a given INFO score
3. Hellinger score
4. Mask data from random chromosomes ->impute -> average genotype errors

Next steps :

1. Script to calculate AF in the dose files of the imputed output
2. Script to calculate MAF according to the population