1. Convert the raw 23andMe data to vcf.

```
/media/daria/DaryaNika/progs/plink_linux_x86_64_20201019/plink --23file
SNP_raw_v4_Full_20170514175358.txt --recode vcf --out snps_clean
--output-chr MT --snps-only just-acgt
```

How many variants were found?

```
grep -v '#' snps_clean.vcf | wc -l
595401
```

2. Annotation

a) **SnpEff/SnpSift**

```
java -jar /home/daria/snpEff/snpEff.jar GRCh37.75 snps_clean.vcf  >
snps_snpeff.vcf
```

First we will compare it with ClinVar database. We download vcf with ClinVar variants: https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh37/clinvar.vcf.gz

```
java -jar /home/daria/snpEff/SnpSift.jar annotate clinvar.vcf
snps_clean.vcf > snps_clean_snpsift_clinvar.vcf
```

```
cat snps_clean_snpsift_clinvar.vcf | grep CLNDN
```

b) **GWAS catalog**

We can download it from here: https://www.ebi.ac.uk/gwas/api/search/downloads/full

```
java -jar /home/daria/snpEff/SnpSift.jar gwasCat -db
gwas_catalog_v1.0-associations_e100_r2021-02-25.tsv snps_clean.vcf >
snps_clean_gwascat.vcf
```

```
snps_clean_gwascat.vcf | grep GWASCAT_TRAIT
```

All are heterozygotes

```
grep -v '^#' snps_clean_snpsift_clinvar.vcf | grep -e 'CLNDN' | cut -f10 |
sort | uniq
0/1
```

Estimate number of non-synonymous

```
 grep -v '^#' snps_clean_snpsift_clinvar.vcf | grep -e 'CLNDN' | grep -v
'|synonymous_variant' | wc -l
713
```

```
cat snps_clean_gwascat.vcf | grep GWASCAT_TRAIT | awk 'OFS = "\t" {if ($10
== "1/1") {print $1, $2, $3, $4, $5, $6, $7, $8, $9, $10}}'
```

3. MT Haplogroup
https://dna.jameslick.com/mthap/
Result: **H(T152C)**

Another method using haplogrep:
Extract mMT variants
```
grep -e 'MT' snps_clean.vcf > snps_clean_MT.vcf
```

Add a header
```
 grep -e '^#' snps_clean.vcf | cat - snps_clean_MT.vcf > snps_clean_MT_header.vcf
```

Running haplogrep classification:
```
./haplogrep classify --in ../snps_clean_MT_header.vcf --format vcf --out ../haplotype.class.txt
```

4. Y Haplogroup
https://ytree.morleydna.com/extractFromAutosomal
Result: R1a1a

Using yhaplo:

yhaplo installation
```
git clone https://github.com/23andMe/yhaplo.git
python -m pip install --editable .
```

running yhaplo
```
yhaplo -i ../snps_clean.vcf -o ../yhaplo_out
```

Result: FAM001_ID001 R-Page7 R-M417 R1a1a1