The unsolved cases were rerun in June 2021.

The pipeline outline:

**1. Alignment**  
bwa mem aligner; default parameters.

Alignment to the GRC38 genome without alt contigs (GCA\_000001405.15\_GRCh38\_no\_alt\_analysis\_set)

**2. Removing duplicates**

Picard MarkDuplicates

**3. Variant calling**

GATK pipeline (v4.1.9).

(Call variants for every nucleotide of a single file (\*g.vcf files); combine the files, call variants on all samples.

The hard filters are implemented separately for the indels and SNPs.

For SNPs:

-filter "QD < 2.0" --filter-name "QD2" \

-filter "QUAL < 30.0" --filter-name "QUAL30" \

-filter "SOR > 3.0" --filter-name "SOR3" \

-filter "FS > 60.0" --filter-name "FS60" \

-filter "MQ < 40.0" --filter-name "MQ40" \

-filter "MQRankSum < -12.5" --filter-name "MQRankSum-12.5" \

-filter "ReadPosRankSum < -8.0" --filter-name "ReadPosRankSum-8"

For indels:

-filter "QD < 2.0" --filter-name "QD2" \

-filter "QUAL < 30.0" --filter-name "QUAL30" \

-filter "FS > 200.0" --filter-name "FS200" \

-filter "ReadPosRankSum < -20.0" --filter-name "ReadPosRankSum-20"

**4. Variant annotation**

VEP104, ensemble annotation.

Plugins in use: CADD, dbNSFP, ExACpLI, loFtool, DisGeNET, REVEL, Mastermind.

Filtering for \_annotated.csv file:

Variants with gnomAF less then 0.05 or not knonw.

Remove the variants without any consequences, intron variants, upstream gene variants, downstream gene variants, intergenic variants, 5’ and 3’ UTR variants, synonymous varisnts, noncoding transcripts variants.

Remove the genes that are not in MitoCarta3.0

Extra filtering:

Clinvar\_out: all variants that are annotated as pathogenic in ClinVar database, no further filters implemented

Morbid: variants in the Morbid genes (as defined by OMIM) excluding the MitoCarta genes; the filtering is the same as for Mitocarta genes (noncoding, pseudogenes, synonymous variants etc are removed from the file)