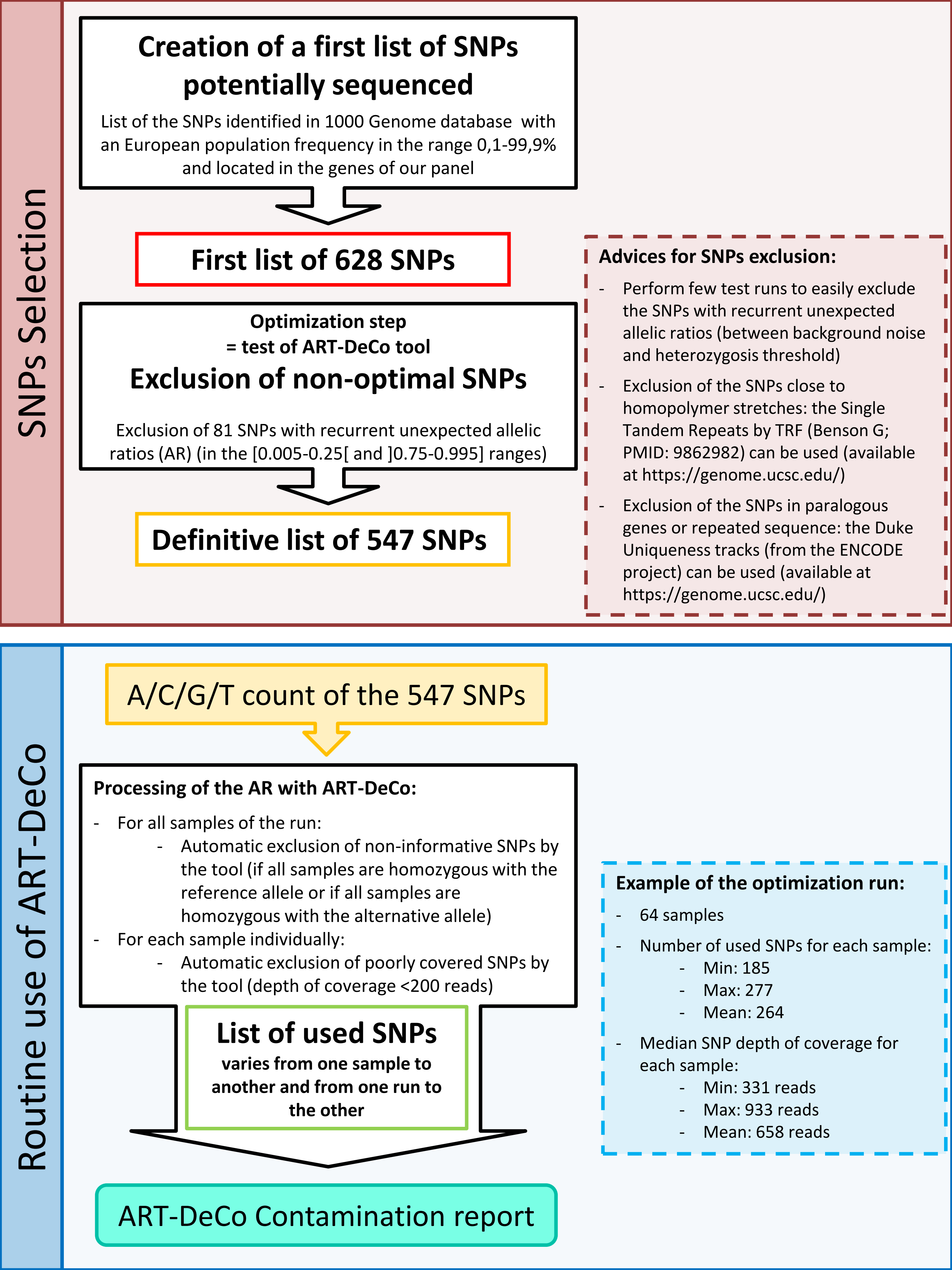

SNPs Selection

Creation of a first list of SNPs potentially sequenced

List of the SNPs identified in 1000 Genome database with an European population frequency in the range 0,1-99,9% and located in the genes of our panel

First list of 628 SNPs

Optimization step
= test of ART-DeCo tool

Exclusion of non-optimal SNPs

Exclusion of 81 SNPs with recurrent unexpected allelic ratios (AR) (in the [0.005-0.25[and]0.75-0.995] ranges)

Definitive list of 547 SNPs

Advices for SNPs exclusion:

- Perform few test runs to easily exclude the SNPs with recurrent unexpected allelic ratios (between background noise and heterozygosis threshold)
- Exclusion of the SNPs close to homopolymer stretches: the Single Tandem Repeats by TRF (Benson G; PMID: 9862982) can be used (available at https://genome.ucsc.edu/)
- Exclusion of the SNPs in paralogous genes or repeated sequence: the Duke Uniqueness tracks (from the ENCODE project) can be used (available at https://genome.ucsc.edu/)

Routine use of ART-DeCo

A/C/G/T count of the 547 SNPs

Processing of the AR with ART-DeCo:

- For all samples of the run:
    - Automatic exclusion of non-informative SNPs by the tool (if all samples are homozygous with the reference allele or if all samples are homozygous with the alternative allele)
- For each sample individually:
    - Automatic exclusion of poorly covered SNPs by the tool (depth of coverage <200 reads)

List of used SNPs
varies from one sample to another and from one run to the other

ART-DeCo Contamination report

Example of the optimization run:

- 64 samples
- Number of used SNPs for each sample:
    - Min: 185
    - Max: 277
    - Mean: 264
- Median SNP depth of coverage for each sample:
    - Min: 331 reads
    - Max: 933 reads
    - Mean: 658 reads