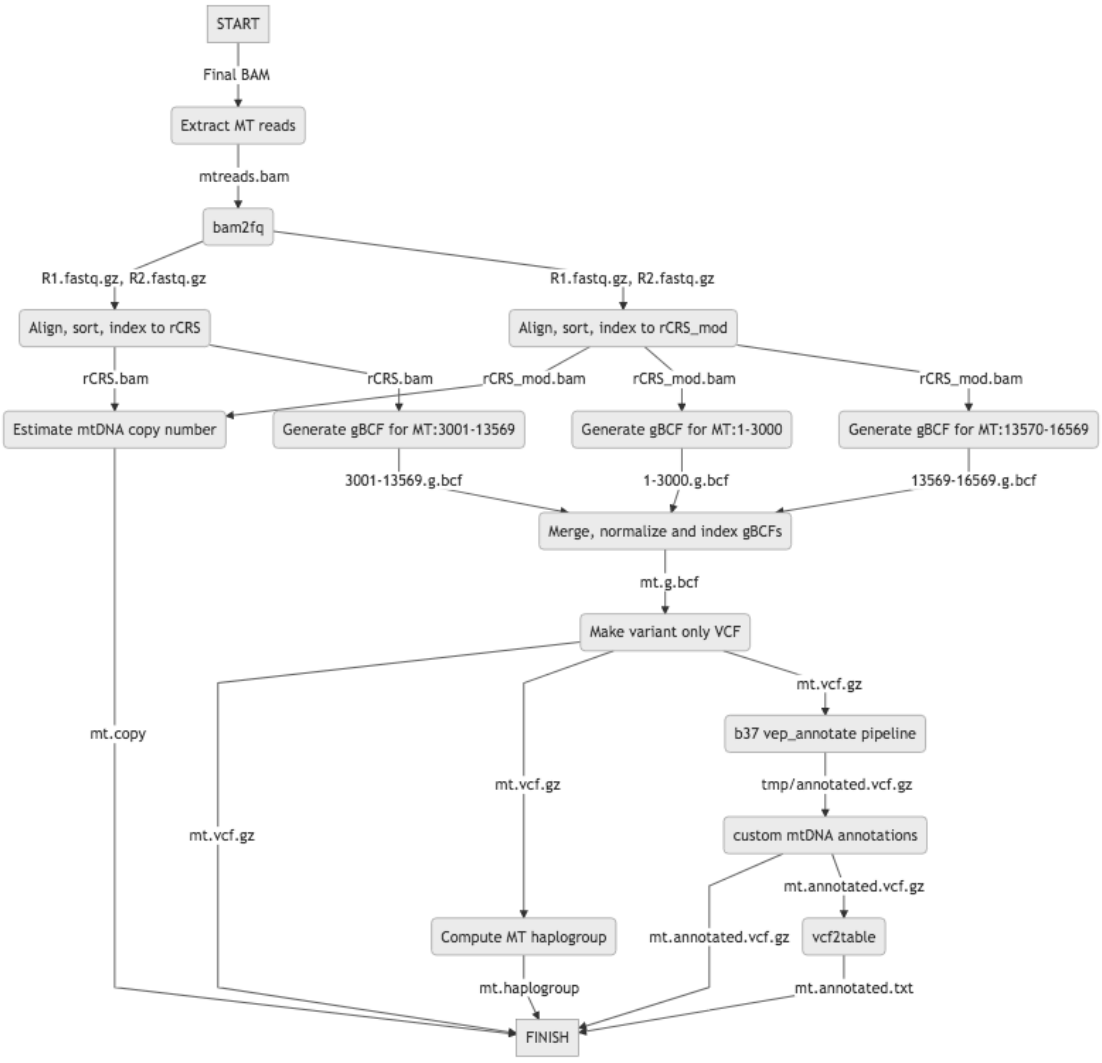
The v1 NYGC mtDNA pipeline for a single sample consists of the following steps,

- Extract the reads uniquely aligned to mitochondria from the final bam file (using samtools v1.7)
- Align MT reads to rCRS mitochondrial reference, taking into consideration the circular nature of the mitochondrial genome (using bwa v0.7.17)
- Estimate the mitochondrial copy number using per base MT coverage and average autosomal coverage (using mosdepth v0.2.3)
- Call mitochondrial SNPs and InDels with AF >= 10% and generate normalized gBCF with coverage & quality information for each base of the mitochondrial genome. (using Freebayes v1.2.0 and bcftools v1.8)
- Estimate mitochondrial haplogroup from the called SNPs (using Haplogrep v2.1.1 and Phylotree build 17)
- Functional annotation of the MT variants (using VEP build 93.2)
- Optionally, merge multiple sample level gBCFs into a complete project level BCF.



Final Deliverables:

- Per sample:
    - *.mt.copy - MT DNA copy number
    - *.mt.haplogroup - MT haplogroup
    - *.mt.g.bcf - per base data and MT variant calls
    - *.mt.g.bcf.csi - index for above file
    - *.mt.vcf.gz - MT variant calls
    - *.mt.vcf.gz.tbi - index for above file
    - *.mt.annotated.vcf.gz - MT variant calls with functional annotations
    - *.mt.annotated.vcf.gz.tbi - index for above file
    - *.mt.annotated.txt - Tab delimited file for MT variant calls with functional annotations

- Per project (optional deliverables):
    - *.mt.combined.bcf.gz - combined project level MT variants
    - *.mt.combined.bcf.gz.csi - index for the above file