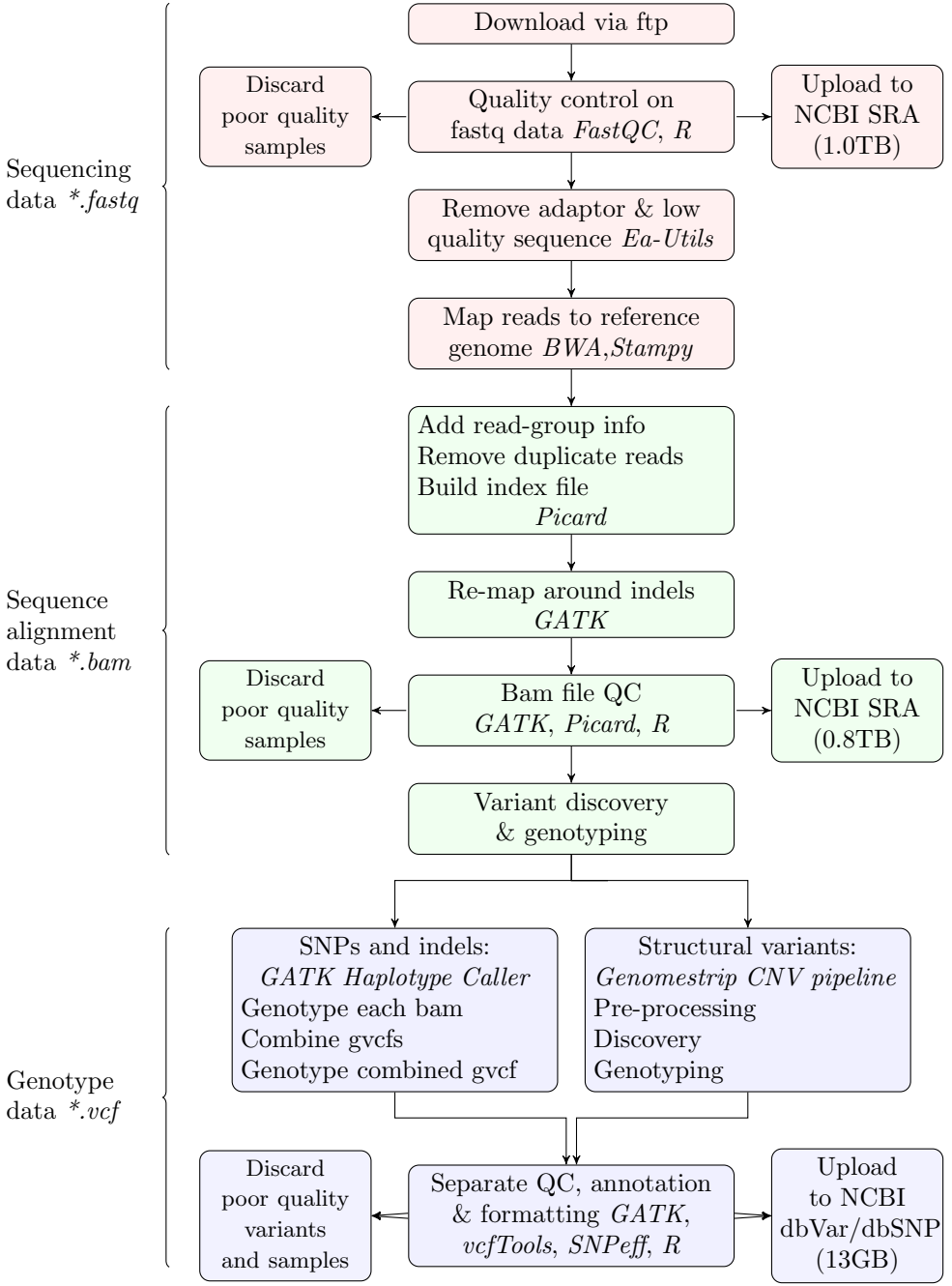
Sequencing data *.fastq
Download via ftp
Discard poor quality samples
Quality control on fastq data FastQC, R
Upload to NCBI SRA (1.0TB)
Remove adaptor & low quality sequence Ea-Utils
Map reads to reference genome BWA,Stampy
Sequence alignment data *.bam
Add read-group info
Remove duplicate reads
Build index file
Picard
Re-map around indels
GATK
Discard poor quality samples
Bam file QC
GATK, Picard, R
Upload to NCBI SRA (0.8TB)
Variant discovery & genotyping
Genotype data *.vcf
SNPs and indels:
GATK Haplotype Caller
Genotype each bam
Combine gvcfs
Genotype combined gvcf
Structural variants:
Genomestrip CNV pipeline
Pre-processing
Discovery
Genotyping
Discard poor quality variants and samples
Separate QC, annotation & formatting GATK, vcfTools, SNPeff, R
Upload to NCBI dbVar/dbSNP (13GB)