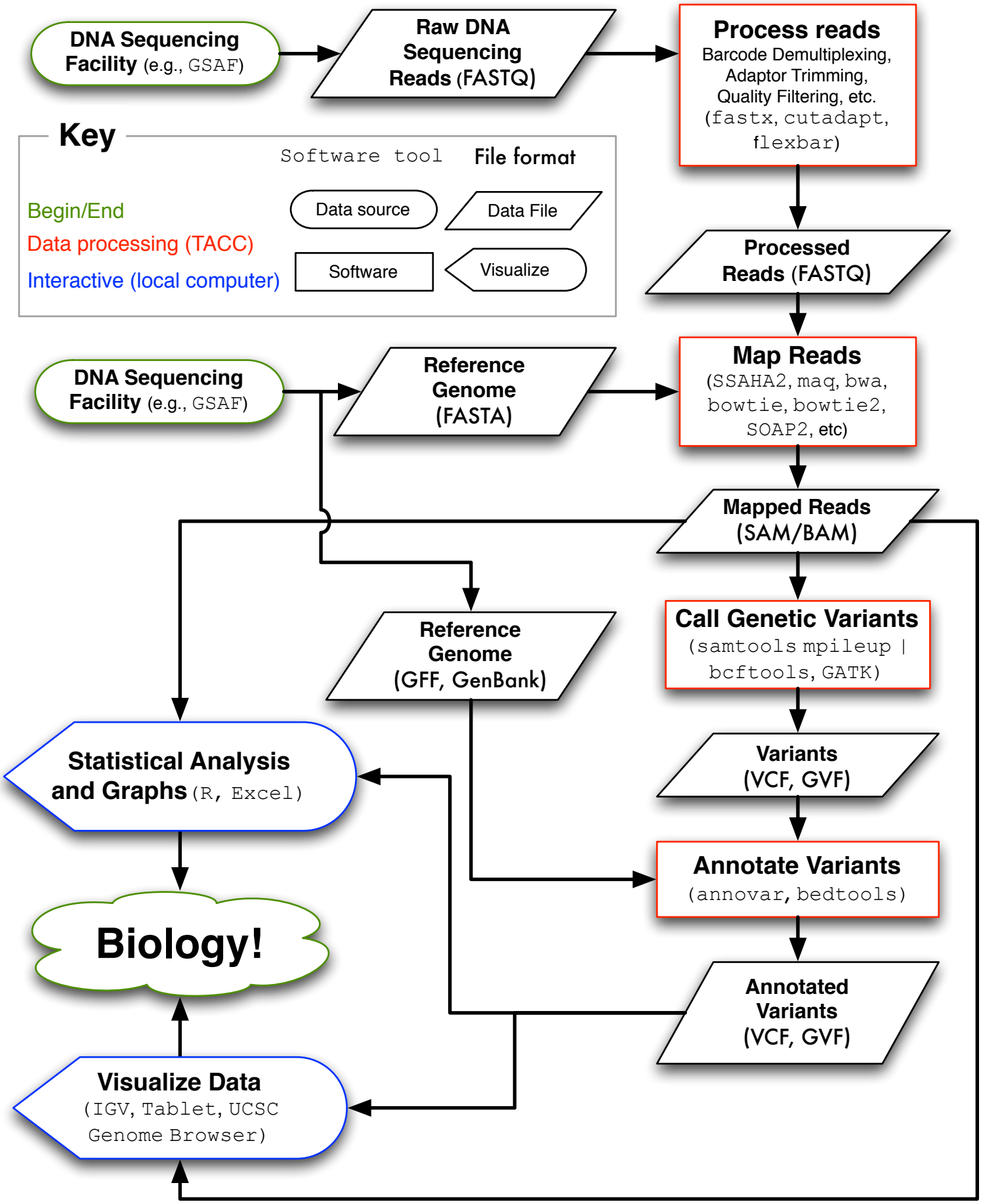
DNA Sequencing Facility (e.g., GSAF)
Raw DNA Sequencing Reads (FASTQ)
Process reads
Barcode Demultiplexing, Adaptor Trimming, Quality Filtering, etc.
(fastx, cutadapt, flexbar)
Key
Software tool
File format
Begin/End
Data processing (TACC)
Interactive (local computer)
Data source
Data File
Software
Visualize
Processed Reads (FASTQ)
DNA Sequencing Facility (e.g., GSAF)
Reference Genome (FASTA)
Map Reads
(SSAHA2, maq, bwa, bowtie, bowtie2, SOAP2, etc)
Mapped Reads (SAM/BAM)
Reference Genome (GFF, GenBank)
Call Genetic Variants
(samtools mpileup | bcftools, GATK)
Statistical Analysis and Graphs (R, Excel)
Variants (VCF, GVF)
Annotate Variants
(annovar, bedtools)
Biology!
Annotated Variants (VCF, GVF)
Visualize Data
(IGV, Tablet, UCSC Genome Browser)