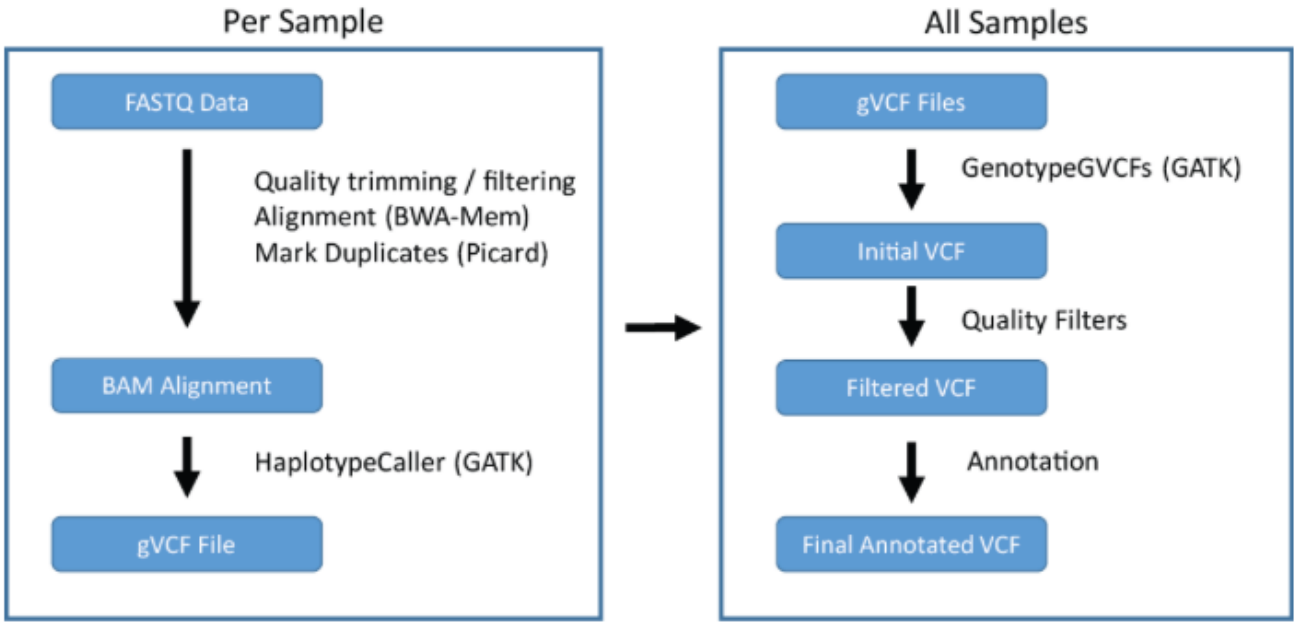
Short Variant Catalog:

The primary dataset generated by MCC is a catalog of short variants (summarized to the right). Raw sequence data are analyzed using a vetted pipeline designed to produce high-confidence genotype calls. The resulting variants are annotated with snpEff to predict function, overlap with regulatory elements and association with phenotypes and diseases.

- Use the Genome Browser to view and search data.
- Unlike many datasets, MCC has genotype-level data, often connected to living animals from pedigreed breeding colonies
- Download raw sequence data from the SRA.

Data Processing / Variant Calling Strategy

We analyze our sequence and variant data using a modified version of the Broad Institute / GATK SNP and Indel Discovery Best Practices, adapted for marmosets.



Sequence Data

All sequence and variant data generated through MCC are publicly available. Sequence data are available on the NCBI SRA database to increase findability and accessibility to the broad researcher community.

The table below displays a list of all sample data currently available through MCC, with basic animal information and SRA accessions per each MCC ID. Note: you can hit the 'export' button to export this to excel or click any of the column headers to filter the data.
Click here to view instructions on downloading data from SRA.

Summary of Current Release:

Current Version:	2.0
Total Passing Variants:	44,883,206
Total Animals:	461
Private Variants:	15,225,277