FASTQ
95% population
5% wild-type

split FASTQ (FASTX Barcode Splitter)

population FASTQ

wild-type FASTQ

trim adapters (Cutadapt)

trimmed population FASTQ

trimmed wild-type FASTQ

map reads (BWA)

population BAM

wild-type BAM

mark PCR duplicates (modified Picard MarkDuplicates)

recalibrate base quality scores (GATK)

marked PCR duplicates (modified Picard MarkDuplicates)

recalibrated wild-type BAM

deduped population BAM

deduped, recalibrated population BAM

recalibrated wild-type BAM

recalibrate base quality scores (GATK)

marked overlapping bases in paired-end reads (custom Java program)

deduped, recalibrated, overlap-added population BAM

deduped, recalibrated, overlap-added population BAM

determined allele counts (custom Python script implementing PySAM)

per-site population allele counts

wild-type covariate table

recalibrated, overlap-added wild-type BAM

per-site wild-type allele counts
per-site population allele counts

merge all wild-type counts

merged per-site wild-type allele counts

compare population and wild-type counts

sites with non-reference allele proportion greater than wild-type

one-tailed Fisher’s Exact Test between population and wt counts

significant sites

multiple hypothesis testing correction

significant sites, FDR-corrected

heuristic filters

final call set

Supplementary Figure 3.