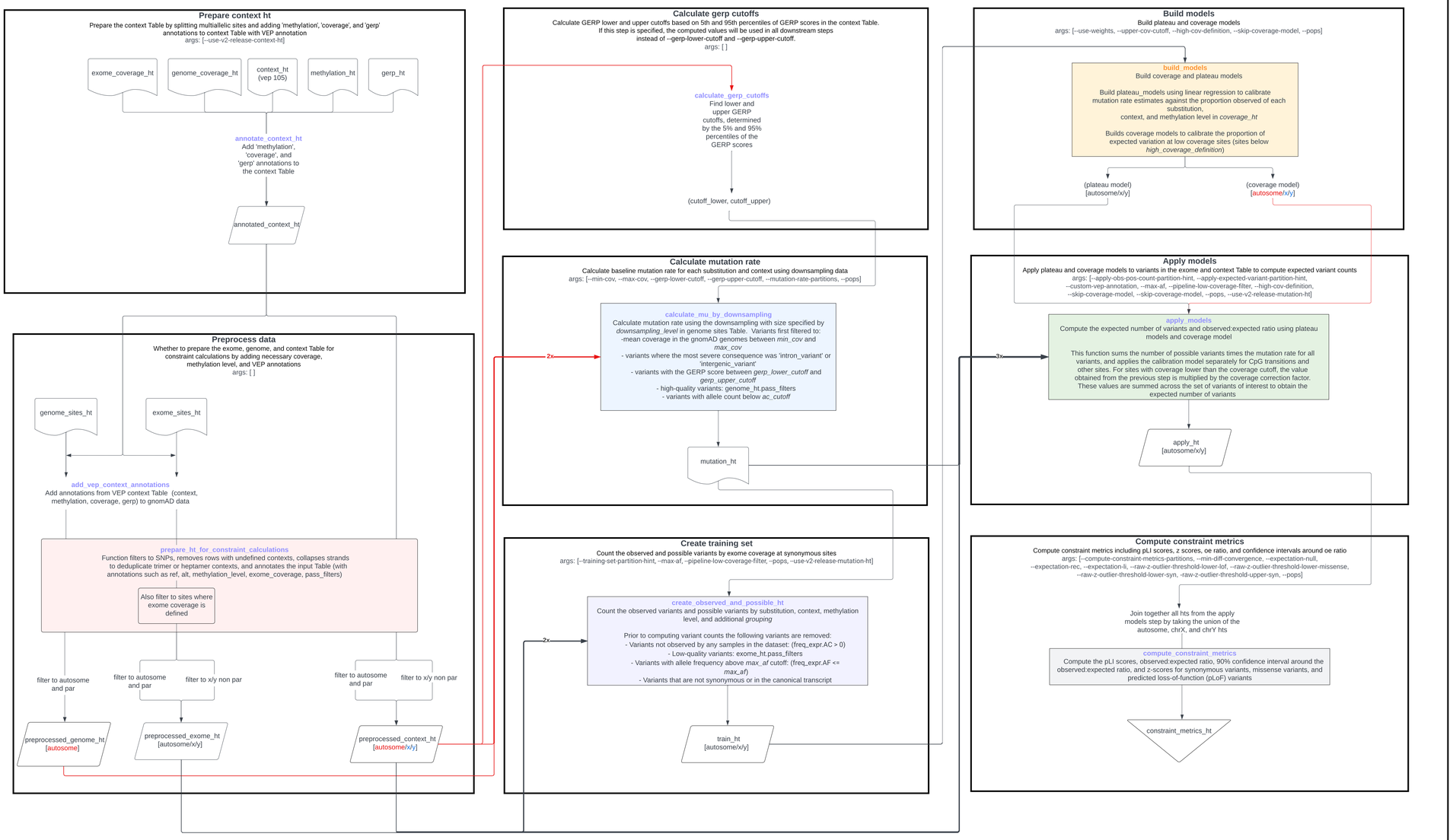
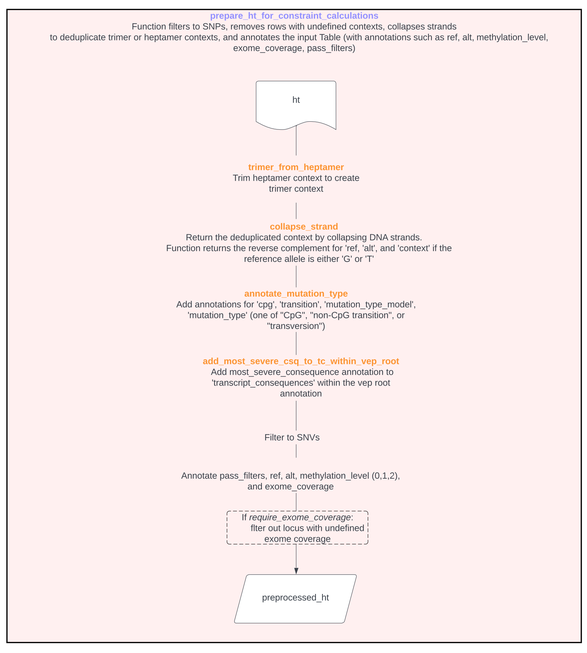
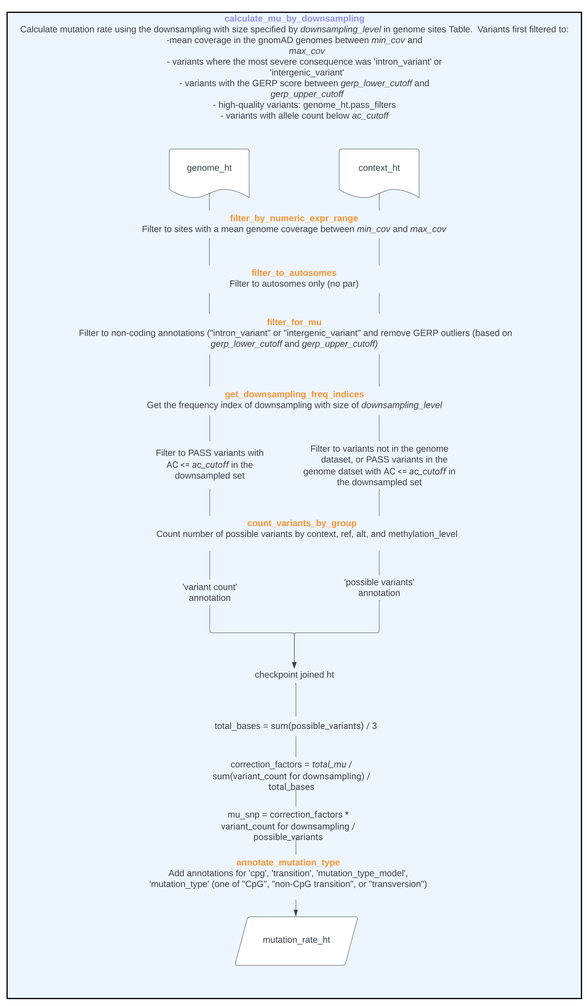
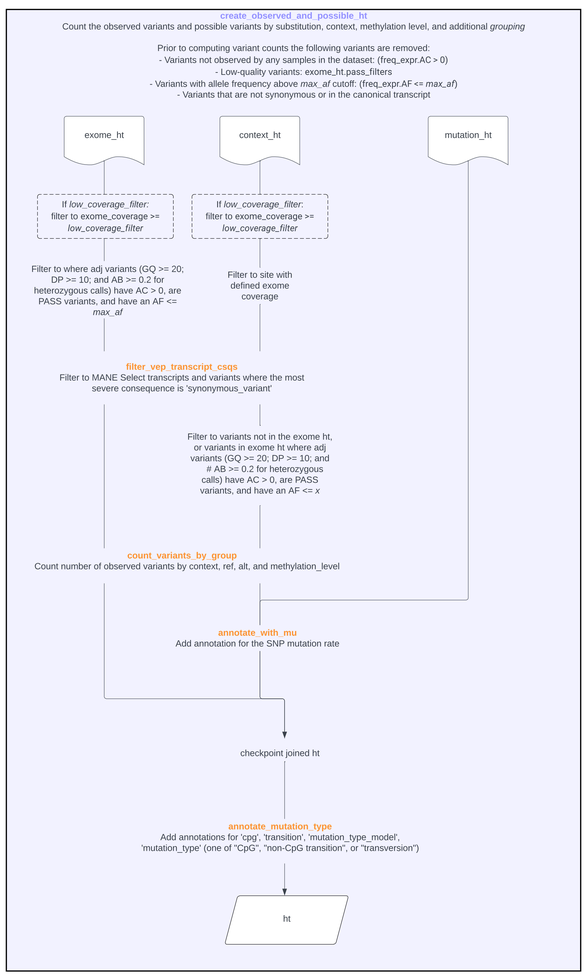
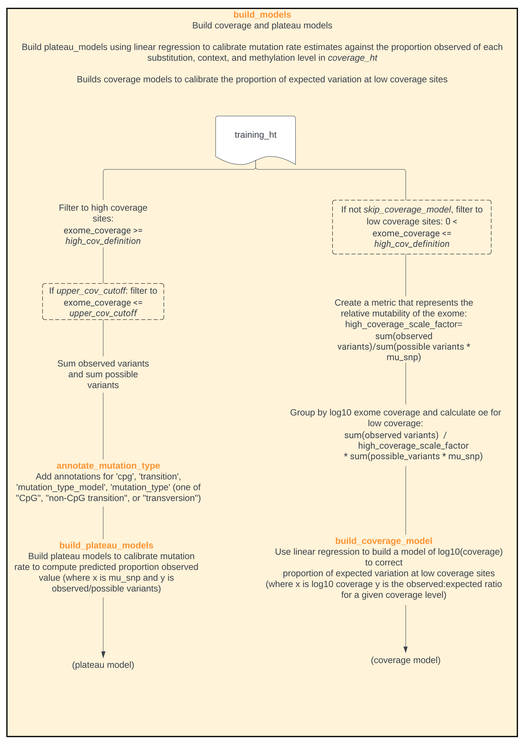
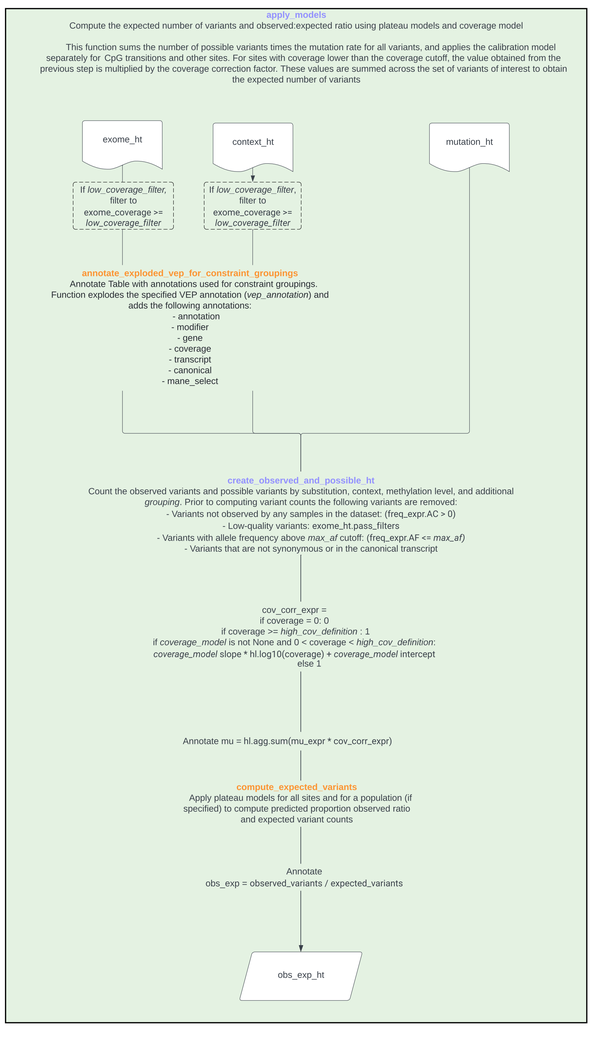
Pipeline



Notes

- Purple text indicates a function that originates from https://github.com/broadinstitute/gnomad-constraint
- Orange text indicates a function that originates from https://github.com/broadinstitute/gnomad_methods
- *Italicized* text indicates an argument which is supplied to a function
- Numbers breaking arrows indicate the number of resources represented by the arrow if greater than one
- Output within ( ) indicates a non-Hail Table output
- The main output of the pipeline that is also used for the gnomAD browser is "constraint_metrics_ht"
- For output with the format of [autosome/x/y], red arrows extending from the resource indicate that only the autosomes are used for the next step, while black arrows indicate that all three resources (autosome, chrX, and chrY) are used in the next step (however, each resource is passed to the next function separately). Note that there are no occurences where only chrX and chrY are passed to the next function.
- Although coverage models are created separately for autosomes, chrX, and chrY, at the moment only the autosome coverage model is used, even for genes on chrX and chrY. Further testing is still needed to determine if separate coverage models for chrX and chrY are appropriate.

Main function arguments

--use-v2-release-context-ht: Whether to use the annotated context Table for the v2 release.
--pipeline-low-coverage-filter: Lower median coverage cutoff to use throughout the pipeline. Default is 30.
--min-cov: Minimum coverage required to keep a site when calculating the mutation rate. Default is 15.
--max-cov: Maximum coverage required to keep a site when calculating the mutation rate. Default is 60.
--gerp-lower-cutoff: Minimum GERP score for a variant to be included when calculating the mutation rate. Default is -3.9885.
--gerp-upper-cutoff: Maximum GERP score for a variant to be included when calculating the mutation rate. Default is 2.6607.
--mutation-rate-partitions: Number of partitions to which the mutation rate Table should be repartitioned.
--training-set-partition-hint: Target number of partitions for aggregation when counting variants for training datasets.
--max-af: Maximum variant allele frequency to keep.
--pops: Populations on which to train models, build models, apply models, or compute metrics on.
--use-v2-release-mutation-ht: Whether to use the mutation rate computed for the v2 release.
--use-weights: Whether to generalize the models to weighted least squares using 'possible_variants'.
--upper-cov-cutoff: Upper median coverage cutoff. Default is 100.
--high-cov-definition: Lower median coverage cutoff to define high coverage sites. Default is 30.
--skip-coverage-model: Omit computing and applying the coverage model.
--apply-obs-pos-count-partition-hint: Target number of partitions for aggregation when counting observed and expected variants.
--apply-expected-variant-partition-hint: Target number of partitions for sum aggregators after applying models to get expected variant counts.
--custom-vep-annotation: Custom VEP annotation to be used to annotate transcripts when applying models.
--compute-constraint-metrics-partitions: Number of partitions to which the unioned Table should be repartitioned.
--min-diff-convergence: Minimum iteration change in pLI score to consider the EM model convergence criteria as met.
--expectation-null: Expected observed/expected rate of truncating variation for genes completely tolerated by natural selection. Default is 1.0.
--expectation-rec: Expected observed/expected rate of truncating variation for recessive disease genes. Default is 0.706.
--expectation-li: Expected observed/expected rate of truncating variation for severe haploinsufficient genes. Default is 0.207.
--raw-z-outlier-threshold-lower-lof: Value at which the raw z-score is considered an outlier for lof variants. Default is -8.0.
--raw-z-outlier-threshold-lower-missense: Value at which the raw z-score is considered an outlier for missense variants. Default is -8.0.
--raw-z-outlier-threshold-lower-syn: Value at which the raw z-score is considered an outlier for synonymous variants. Default is -8.0.
--raw-z-outlier-threshold-upper-syn: Value at which the raw z-score is considered an outlier for synonymous variants. Default is 8.0.

Functions