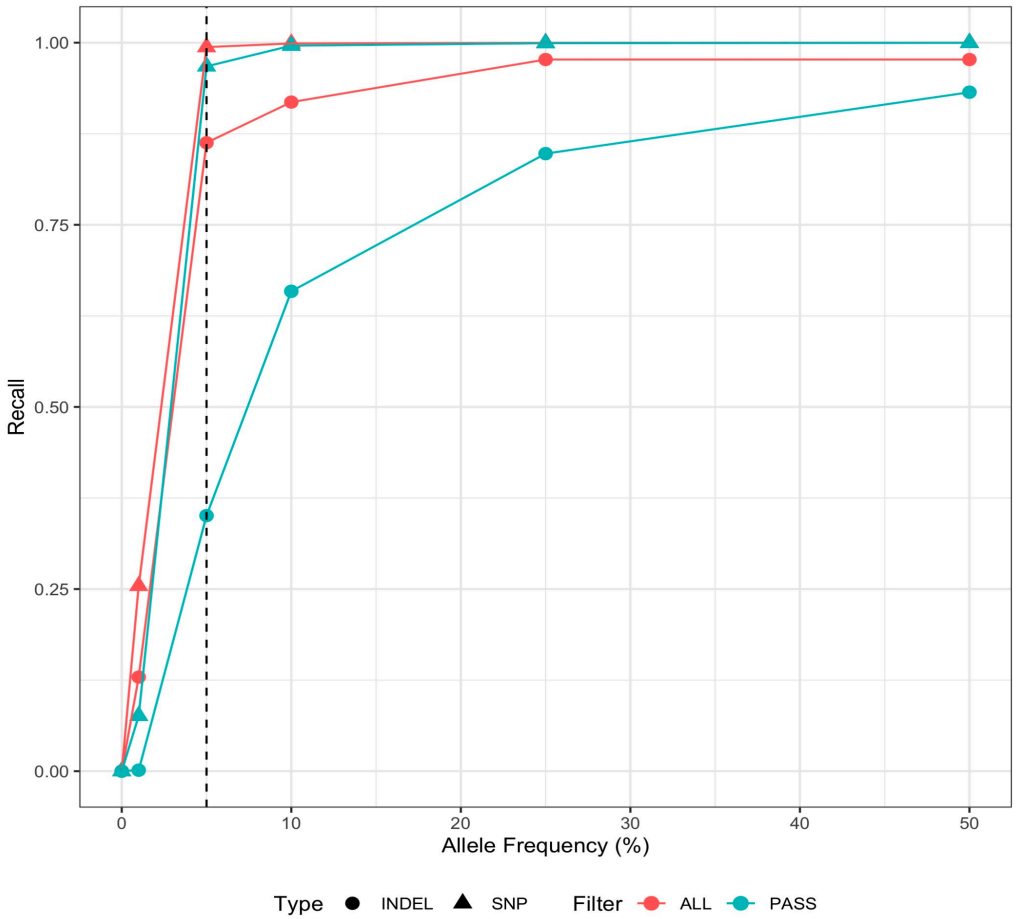
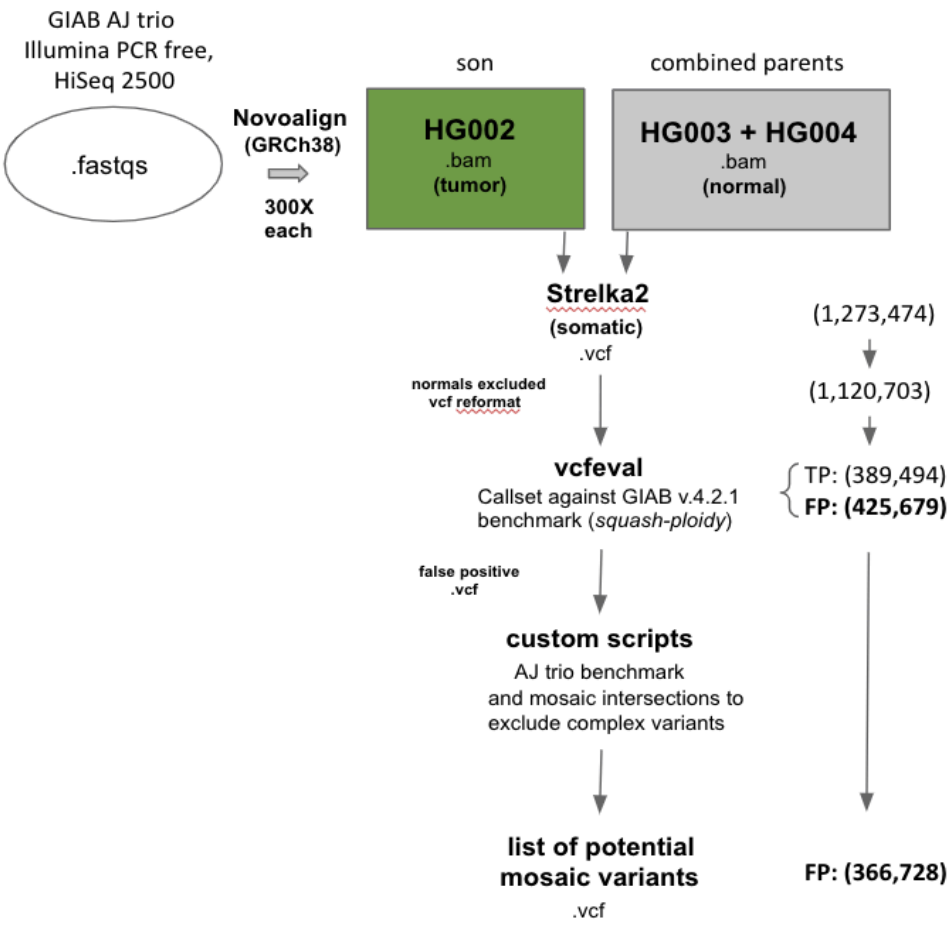
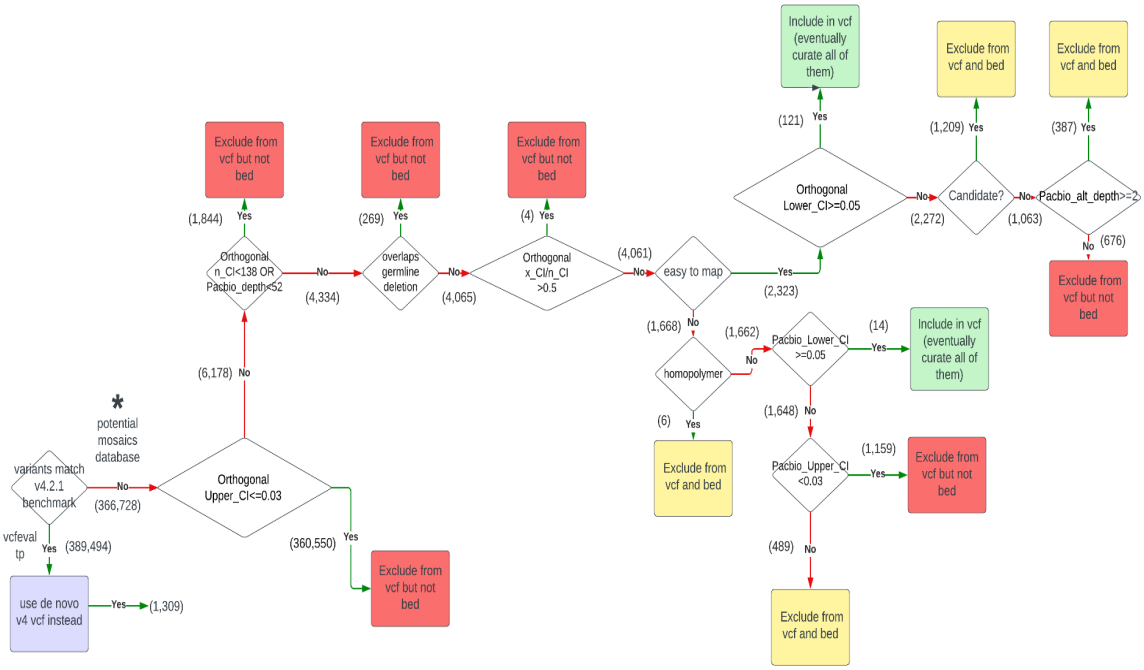
Supplemental Figures



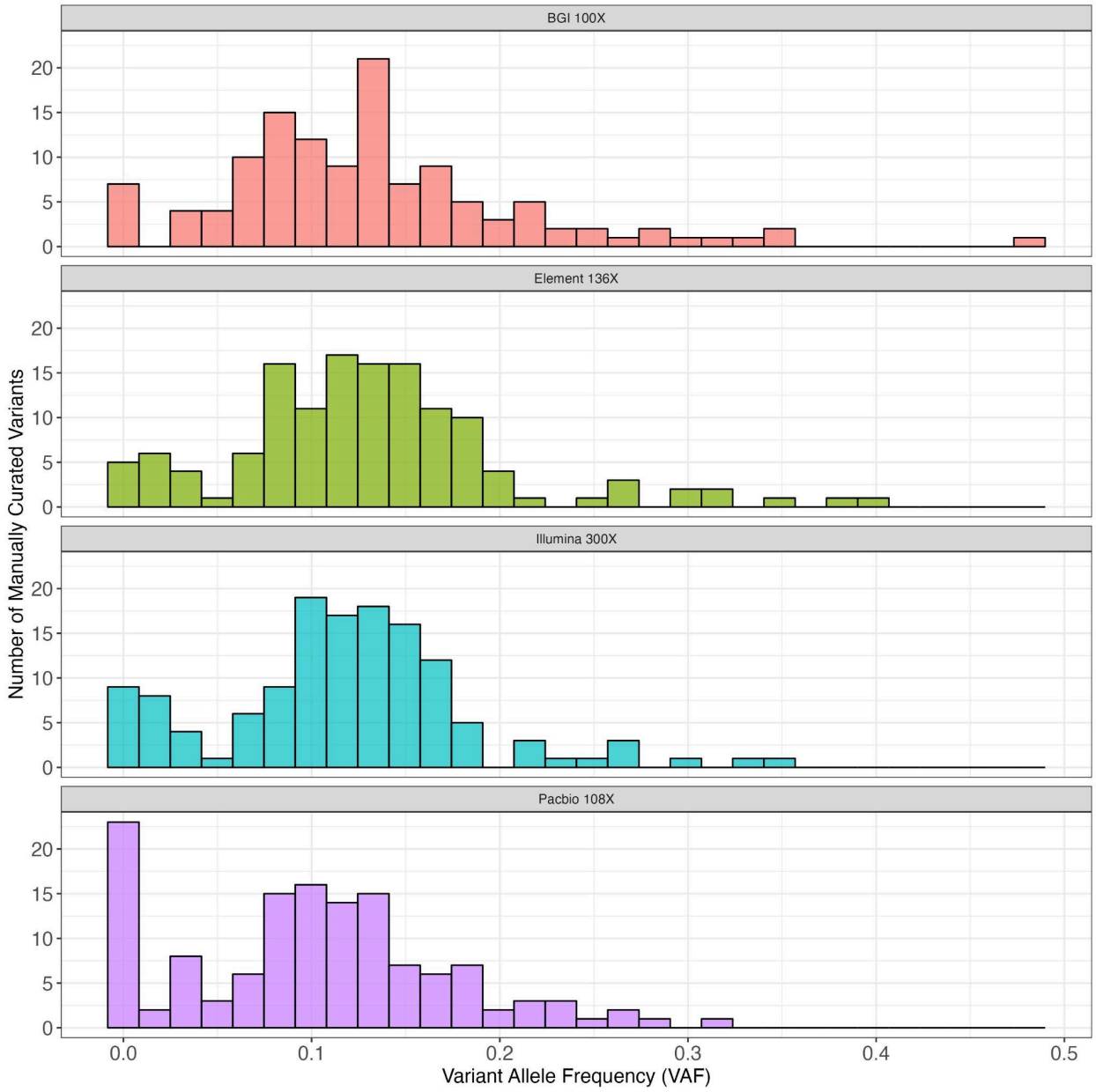
Supplemental Figure 1: Limit of detection (LOD) was established at 5% variant allele fraction (VAF) using Strelka2 callsets from six in-silico mixtures of GIAB reference 300x samples (subset to chromosome 20), HG002 (son) and HG003 (father), as a control. X-axis values indicate the different samples, ranging from 0% AF (HG003 unmixed) to 50% AF (HG002 unmixed). Callsets for each mixture were benchmarked in the precisionFDA app.



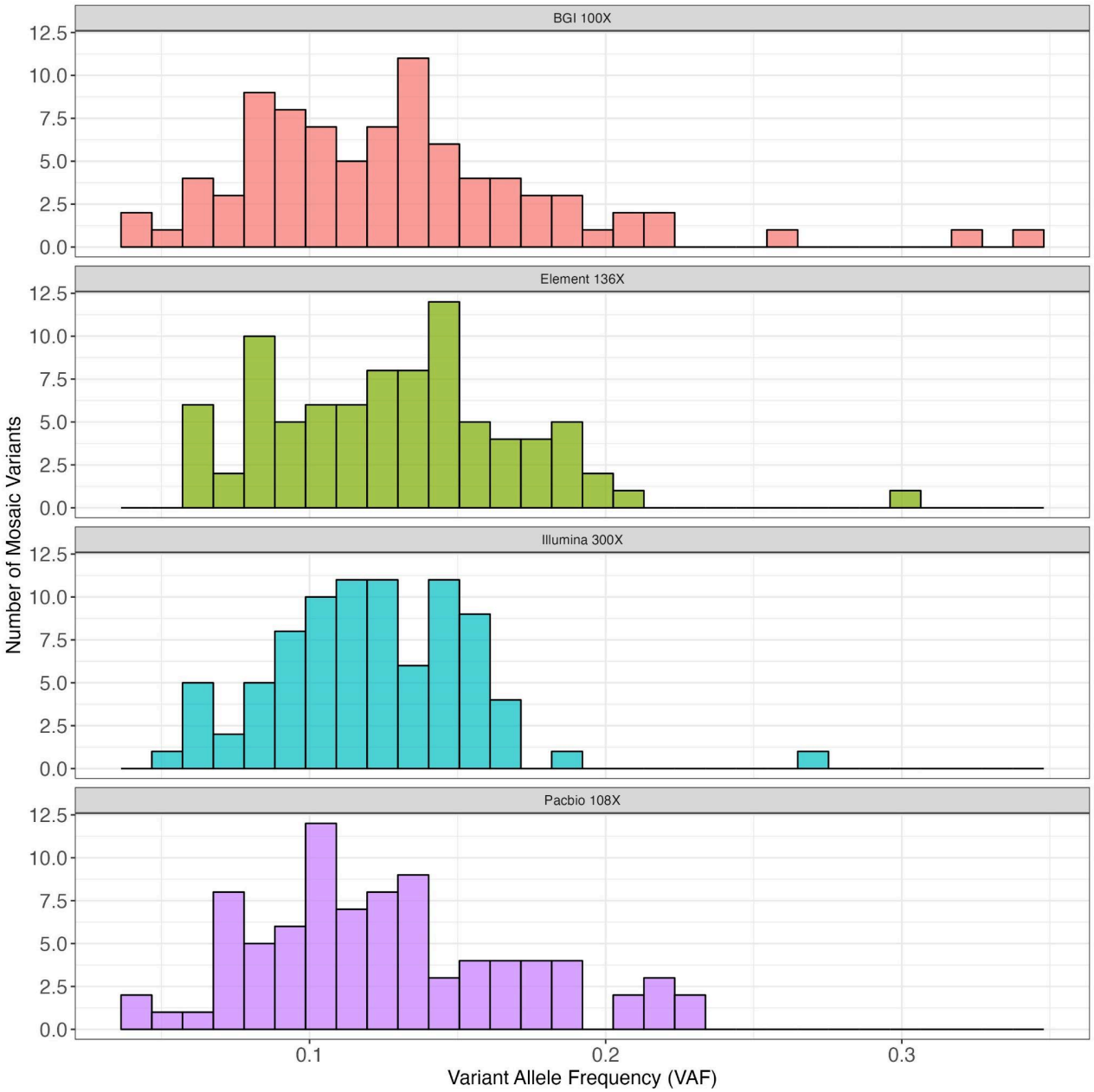
Supplemental Figure 2: Variant counts from the Strelka2 tumor-normal run with the GIAB AJ trio, benchmarking, and filtering steps to generate a list of potential mosaic variants (366,728 vcfeval false positives) for database creation.



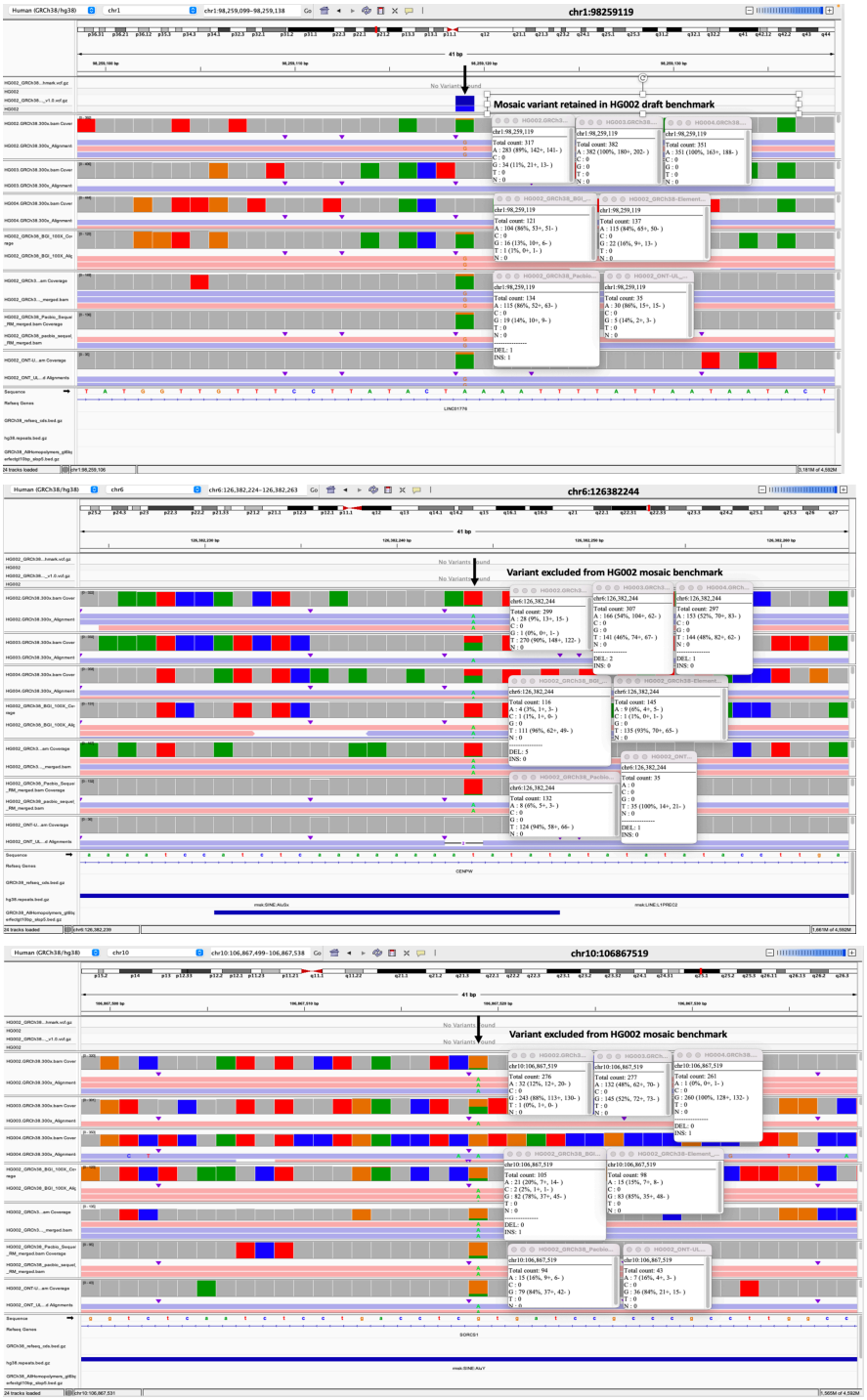
Supplemental Figure 3: Mosaic benchmark decision tree for filtering potential mosaic variants to produce a list for manual curation. A series of heuristics were applied to the potential mosaic database (* starting at bottom left) using combined orthogonal CI thresholds and other attributes for candidate set determination. The tree resulted in three groups of variants: green boxes indicate variants to curate, yellow represents variants excluded from benchmark VCF and BED, and red variants excluded from benchmark VCF but not BED. Values in parentheses are variant counts for each step of the decision tree.



Supplemental Figure 4: VAF distributions for all **manually curated variants** (135) in high-coverage Illumina (300x) and orthogonal (BGI - 100x, Element - 136x, PacBio HiFi - 106x) datasets generated from NIST HG002 reference material.



Supplemental Figure 5: VAF distributions for the 85 HG002 **mosaic benchmark variants** in high-coverage Illumina (300x) and orthogonal (BGI - 100x, Element - 136x, PacBio HiFi - 106x) datasets generated from NIST HG002 reference material.



Supplemental Figure 6: IGV profiles of variants (A) included in and (B,C) excluded from the HG002 mosaic benchmark after manual curation. Profile A depicts an HG002-specific candidate mosaic SNV (chr1:98259119 - not detected in parents), while B and C show a putative and candidate SNV (chr6:126382244 and chr10:106867519). The first three data tracks are the HG002 (son), HG003 (mother), and HG004 (father) 300X coverage Illumina data. The following tracks are for HG002 BGIseq dataset, Element, PacBio Sequel, and ONT-UL datasets.