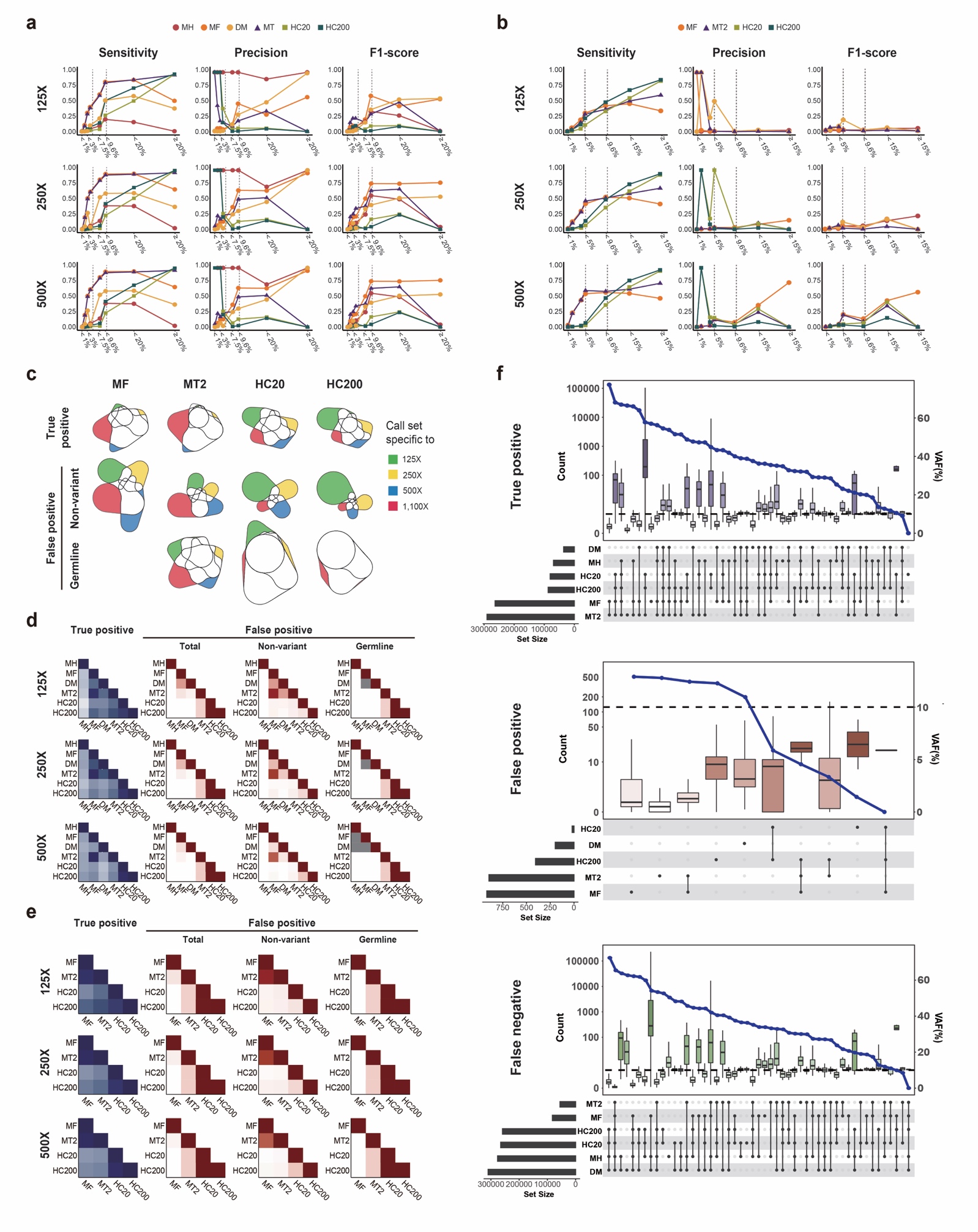
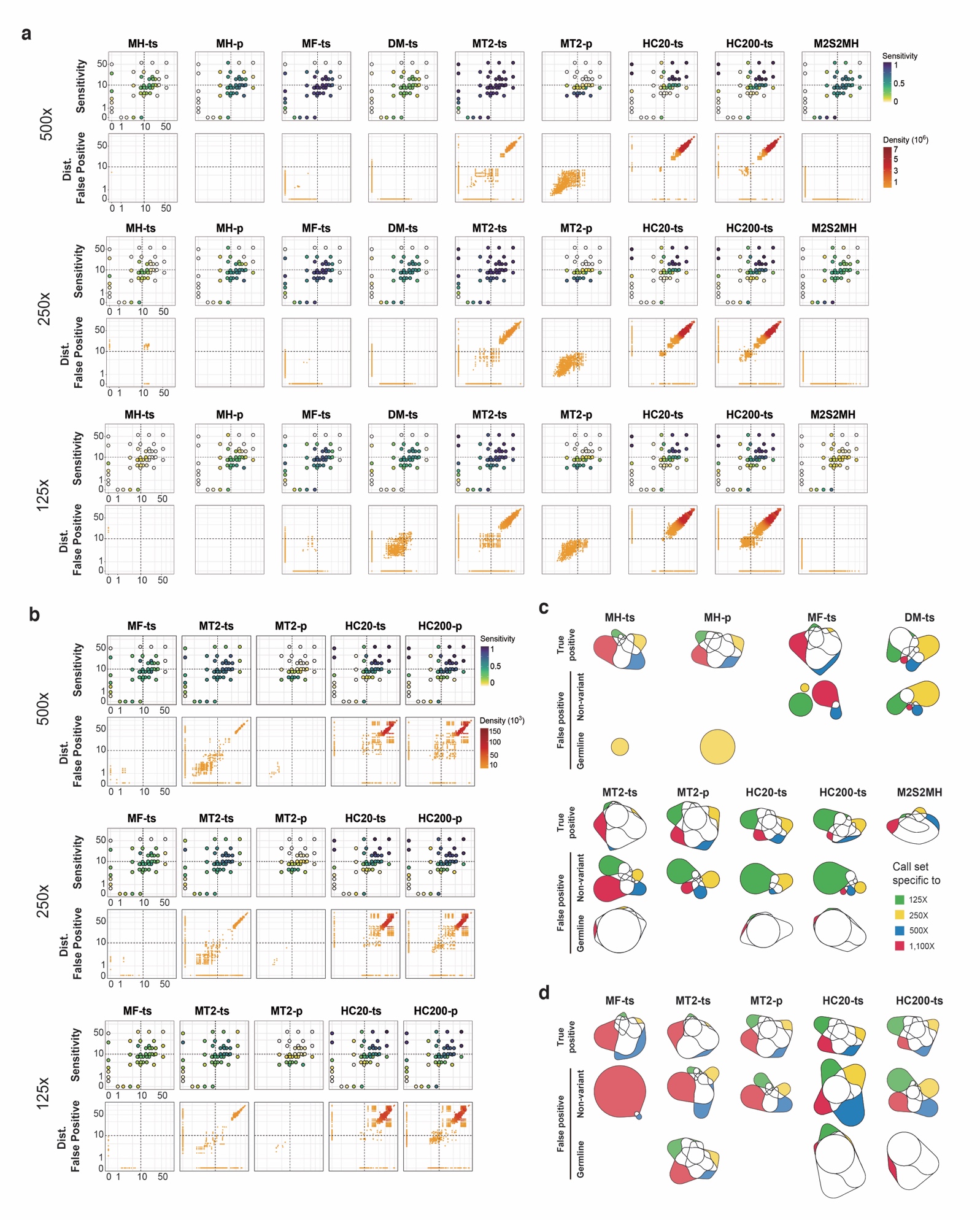
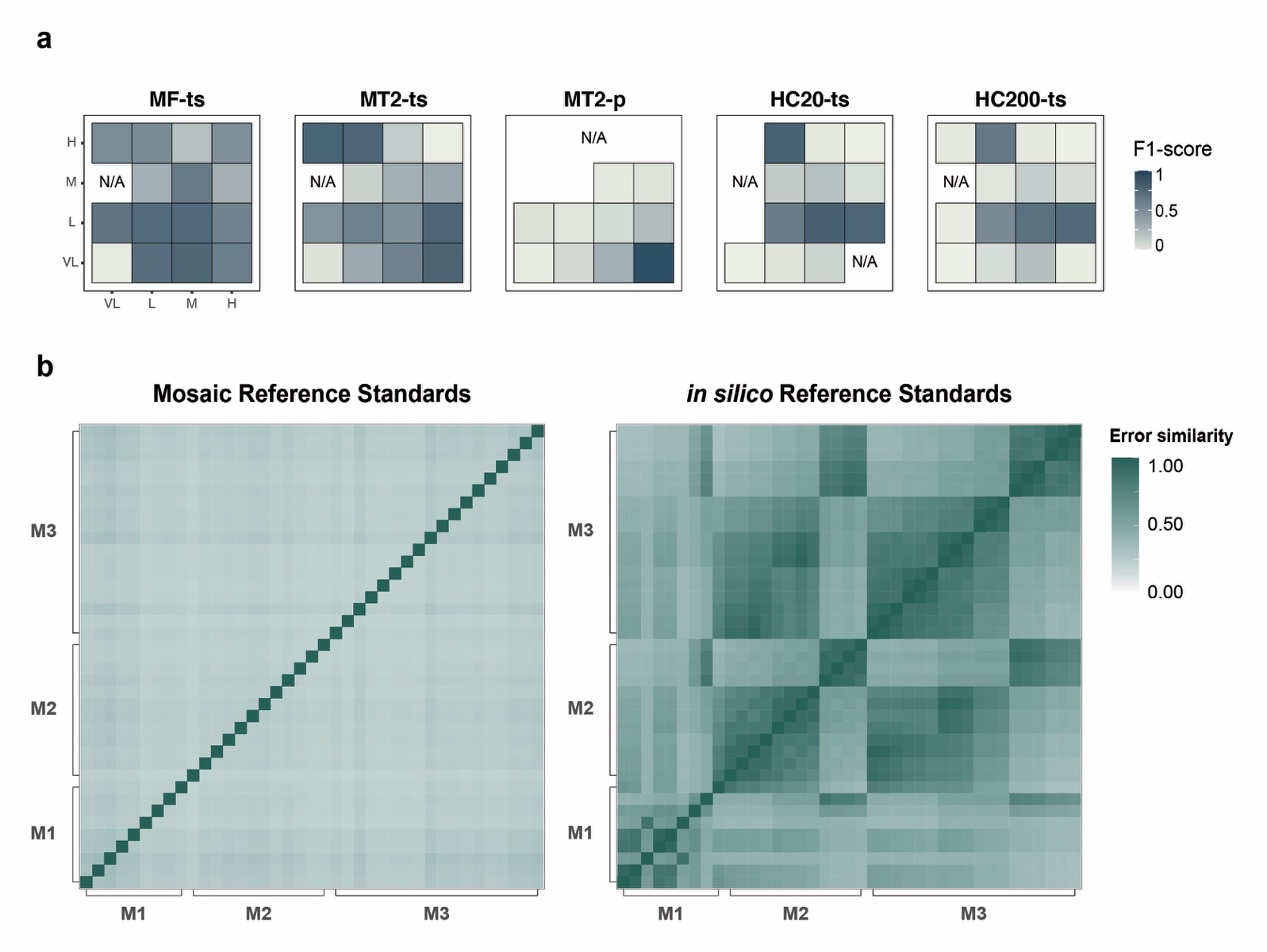
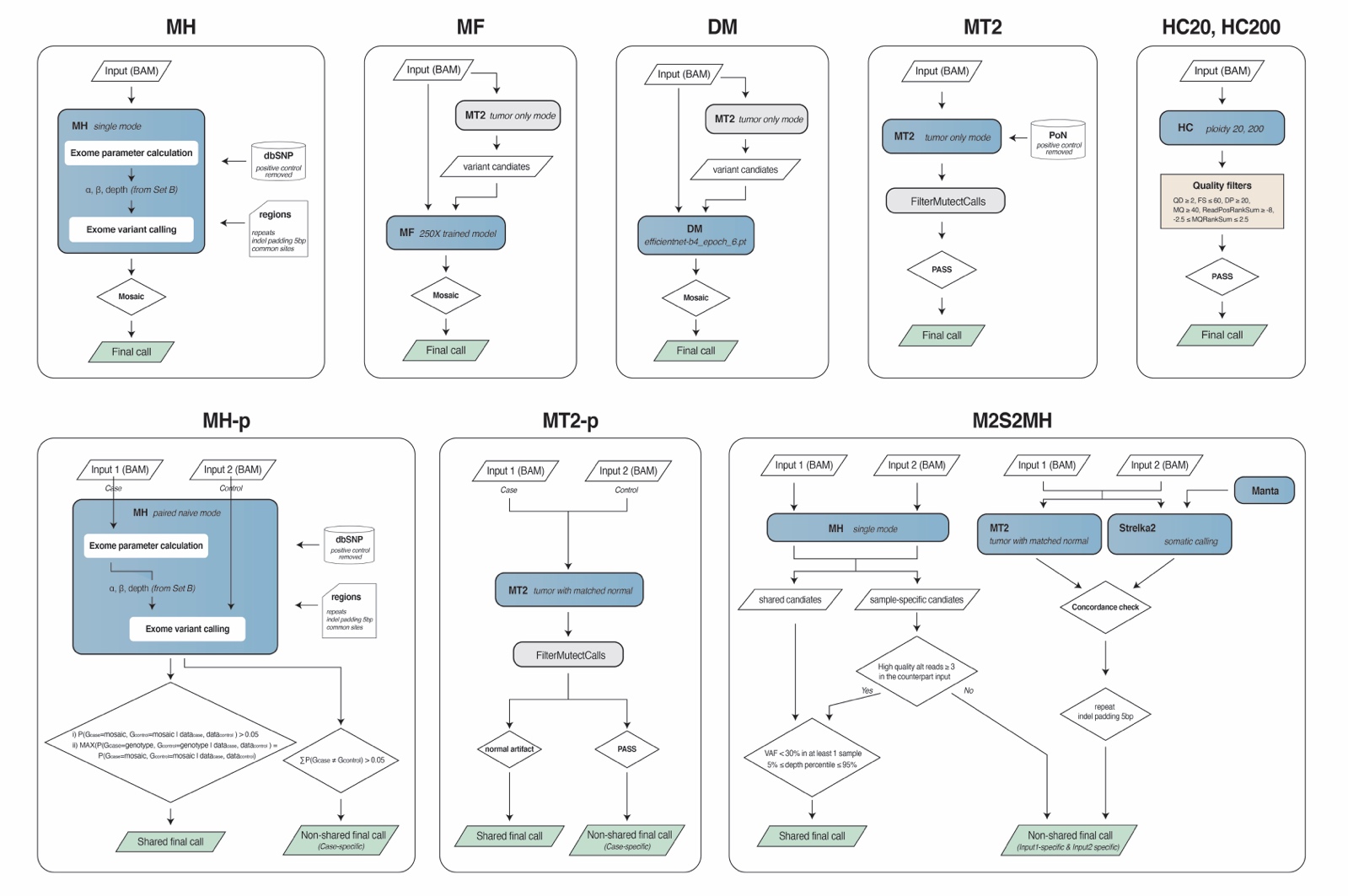
**Supplementary Figures**

**Supplementary figure 1 (a)** Evaluation of SNV calling performance in single sample in three different read depths (125X, 250X, and 500X) **(b)** valuation of INDEL calling performance in single sample in three different read depths (125X, 250X, and 500X) **(c)** Diagram illustrating the consistency of the INDEL call sets for single sample analysis **(d)** SNV Call set similarity between methods for all read depths (125X, 250X, and 500X). **(e)** INDEL Call set similarity between methods for all read depths (125X, 250X, and 500X). **(f)** The composition of three different types of variant call sets (true positives, false positives, and false negatives) in detection approaches. On the right y-axis, variant allele frequencies (VAF) of each combination are shown.

**Supplementary figure 2 (a)** Sensitivity and false positive distribution of the nine approaches of shared SNV in three different read depths (125X, 250X, and 500X) **(b)** Sensitivity and false positive distribution of the nine approaches of shared INDEL in three different read depths (125X, 250X, and 500X **(c)** Diagram illustrating the consistency of the shared SNV call sets towards all read depths **(d)** Diagram illustrating the consistency of the shared INDEL call sets towards all read depths ****

**Supplementary figure 3 (a)** The partitioned F1-scores of shared mosaic INDELs were calculated in sixteen areas with the combinations of the four VAF range groups (very-low: <5%, low: ≥5% and <10%, medium: ≥10% and <25%, and high: ≥10% and <25%). One of the sixteen areas couldn’t be evaluated as the positive controls could not be assigned, and areas with none of true positive in approaches also could not be shown (N/A). **(b)** Comparison of error profiles between the mosaic reference standards and *in silico* reference standards, which were generated by the same combinations and mixing ratios in the 39 products. Unexpected alternative alleles in the non-variant negative controls were calculated with the pileups. The similarity was calculated by Jaccard Index. M1, M2, and M3 refers to the three different categories of the reference standards depending on the mixing combination. VL very low, L low, M medium, H high

****

**Supplementary figure 4** Schematic pipelines of the evaluated detection approaches for the single sample analysis approaches (upper), which were also utilized for the two-single analysis in shared variant detection. Pipelines for the paired-sample analysis are also shown (lower). Detailed information and the versions used can be found in the **Method**.

**Supplementary Notes**

**Noise in the reference standards**

To investigate the effect of noises due to the mixing procedure, we gathered the germline variants of the six cell lines, MRC5, RPE, CCD-18co, HBEC30-KT, THLE-2, and FHC. The merged number of germline variants that was 62,719 (Target region of the whole exome sequencing, SureSelect Human All Exon V6, Agilent Technologies, Inc., CA, USA). When subtracted by the number of the positive controls (18,873), 43,866 remain in the mixture. It means there would be a variant in every 843 base pairs in average. In the MRC5 cell line, which has pure genotype, 24,926 germline variants existed and there would be one variant in every 1,484 base pairs. Considering that variant calling algorithms use 100-200 base pair-window size to evaluate alignment patterns of flanking regions or variant clustered regions, the effect by the noise would be ignorable.