

## *supplementary materials*

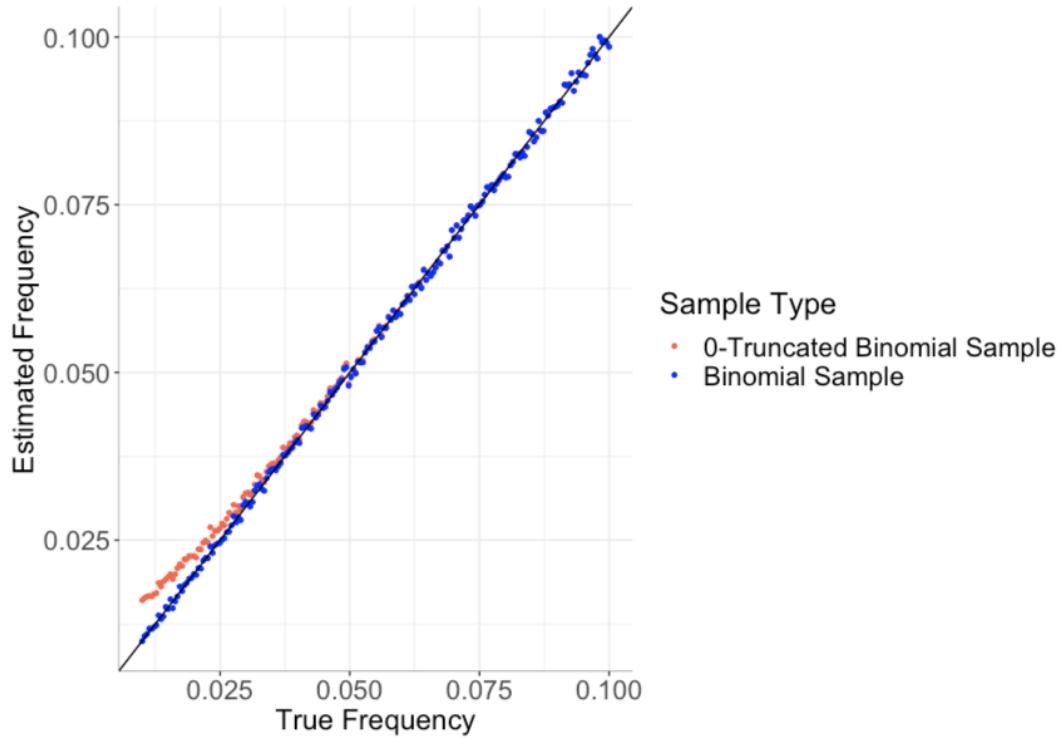


Fig. S1: Comparing the estimated frequencies, estimated by the proportion of alternative reads (or success) versus the true frequencies for both binomial (represented by the blue dots) and 0-truncated binomial samples (represented by the red dots). The plot suggests that the proportion of alternative reads is a biased estimator for 0-truncated binomial samples.

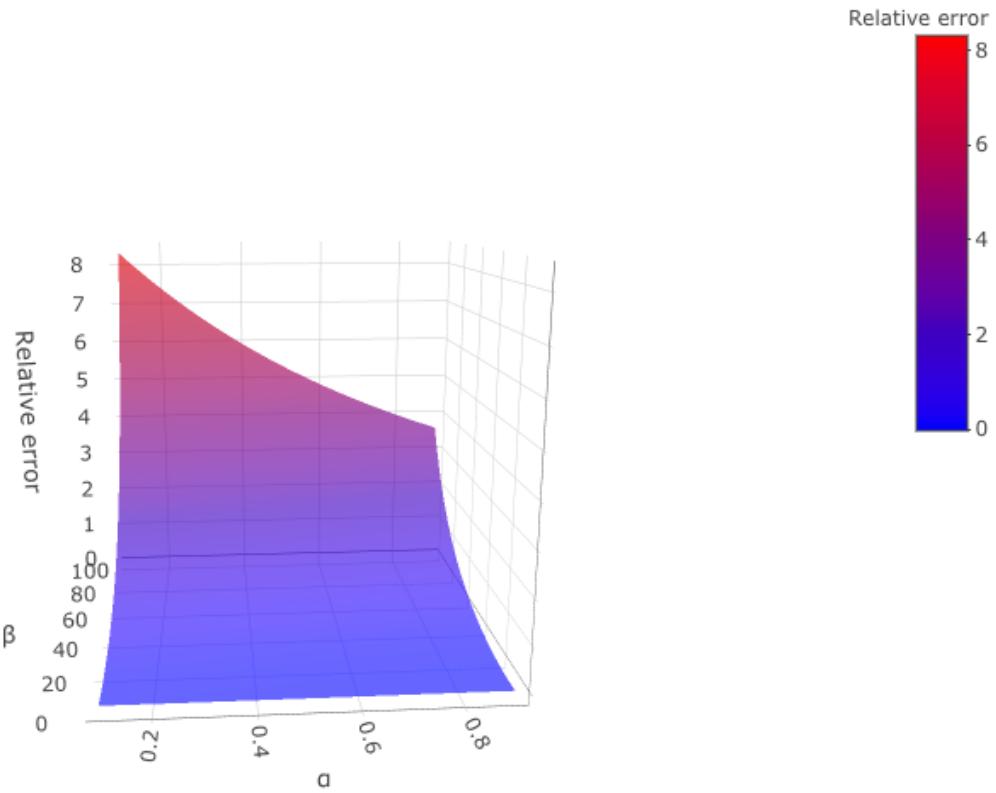


Fig. S2: The surface represents the relative error, calculated from our simulated data, over different parameter values. Positive relative errors correspond to overestimation. As we see, for beta distributions with parameter values within these ranges the amount of underestimation error is small, but in cases in which the beta density function is sharp the counting method may result in an overestimation error 8 times greater than the true TMB.

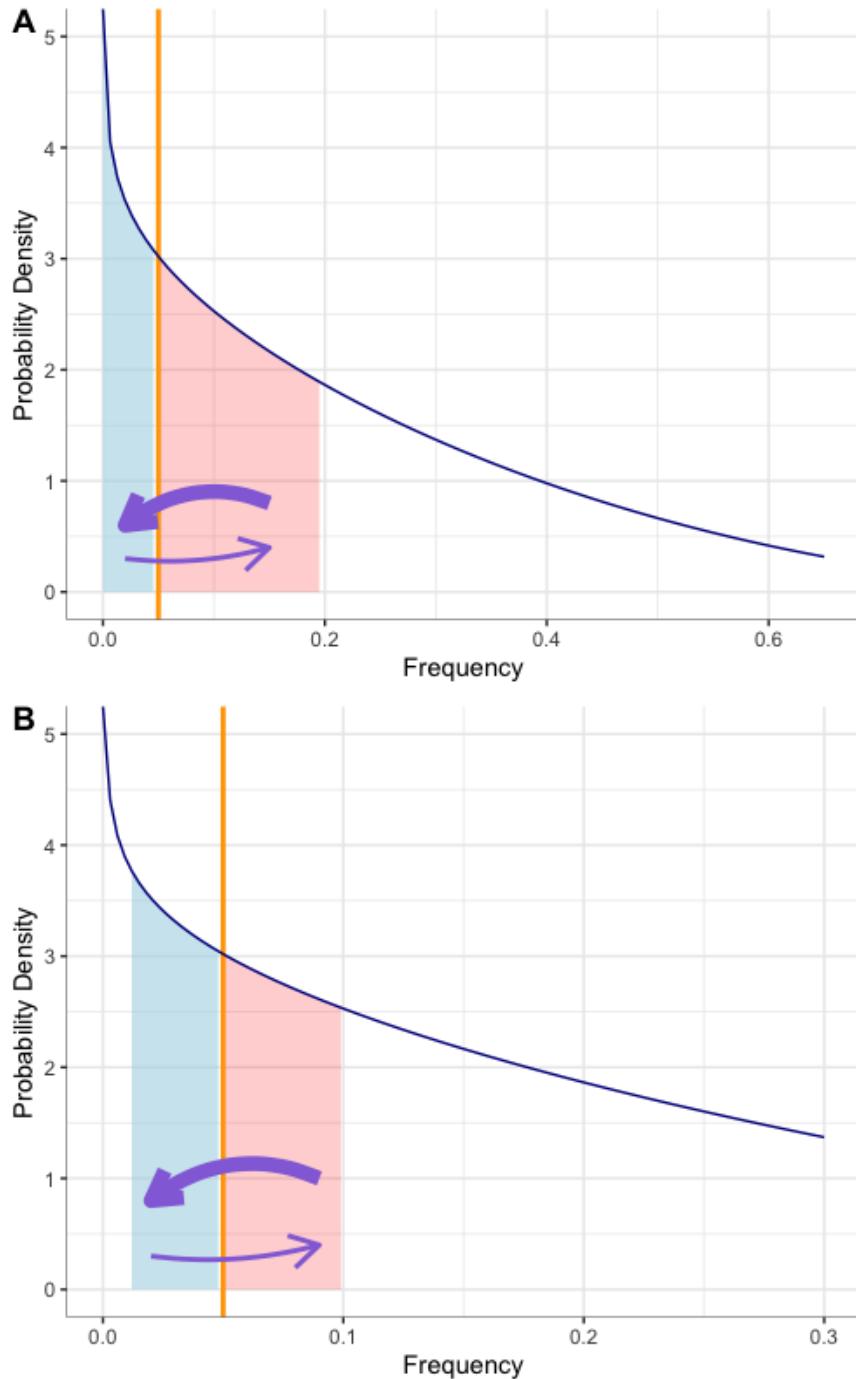


Fig. S3: Cases in which the counting method may underestimate TMB. The shaded red area represents the true frequencies (of somatic mutations in a tumor sample) that their corresponding estimated frequency might be less than 0.05. The blue area represents those true frequencies that their estimated frequency might cross the point of threshold. As the blue area is less than the red area, the plot suggests that the TMB estimated by counting the number of mutations with frequency above 0.05 may underestimate true TMB. However the underestimation error in B is smaller than A, as in B the tumor sample has been sequenced at higher depth which result in a smaller interval around the threshold that is not truncated on the left.

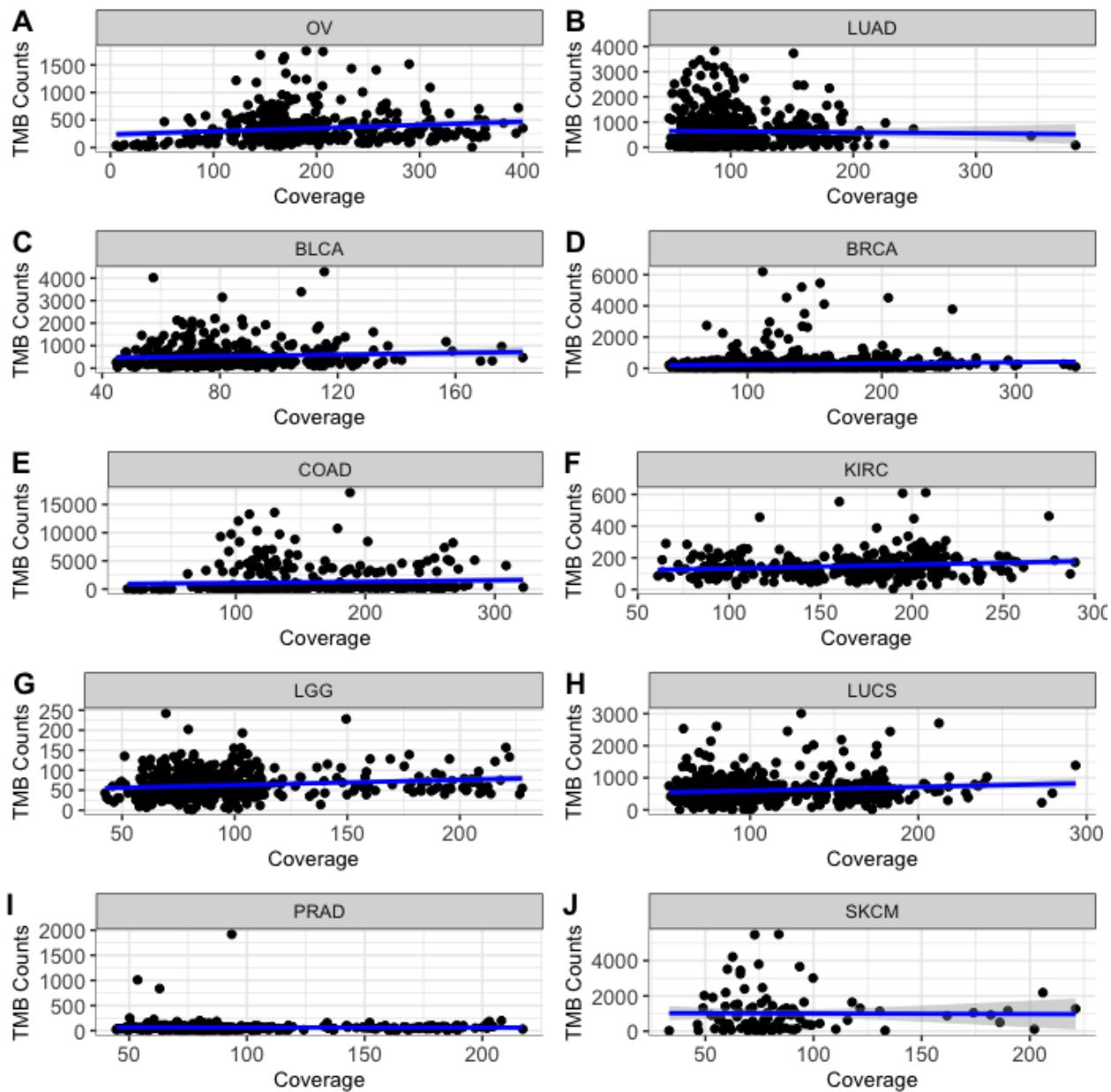


Fig. S4: (A-I), Spearman correlation analysis to test for relationship between TMB and coverage. The results show a positive relationship between Coverage and TMB for 6 out of the 10 cancer types. OV (Spearman  $\rho = 0.15$ , p-value=0.00), LUAD (Spearman  $\rho = 0.04$ , p-value=0.32), BLCA (Spearman  $\rho = 0.08$ , p-value=0.07), BRCA (Spearman  $\rho = 0.13$ , p-value=0.00), COAD (Spearman  $\rho = 0.27$ , p-value=0.00), KIRC (Spearman  $\rho = 0.16$ , p-value=0.00), LGG (Spearman  $\rho = 0.15$ , p-value=0.00), LUSC (Spearman  $\rho = 0.13$ , p-value=0.00), PRAD (Spearman  $\rho = 0.02$ , p-value=0.57), SKCM (Spearman  $\rho = 0.07$ , p-value=0.47)

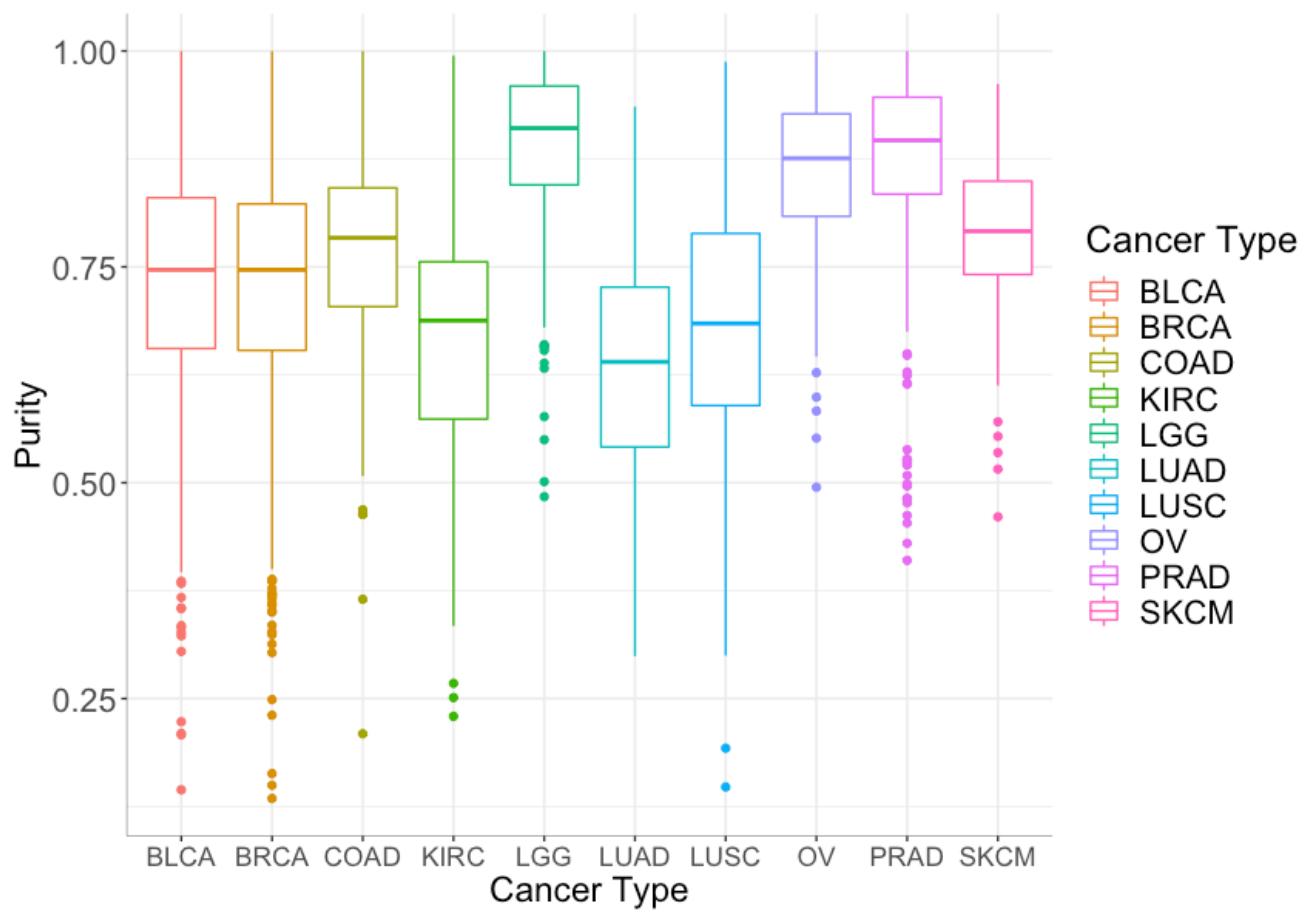


Fig. S5: Boxplot of sample purities across 10 cancer cohorts used in this study.

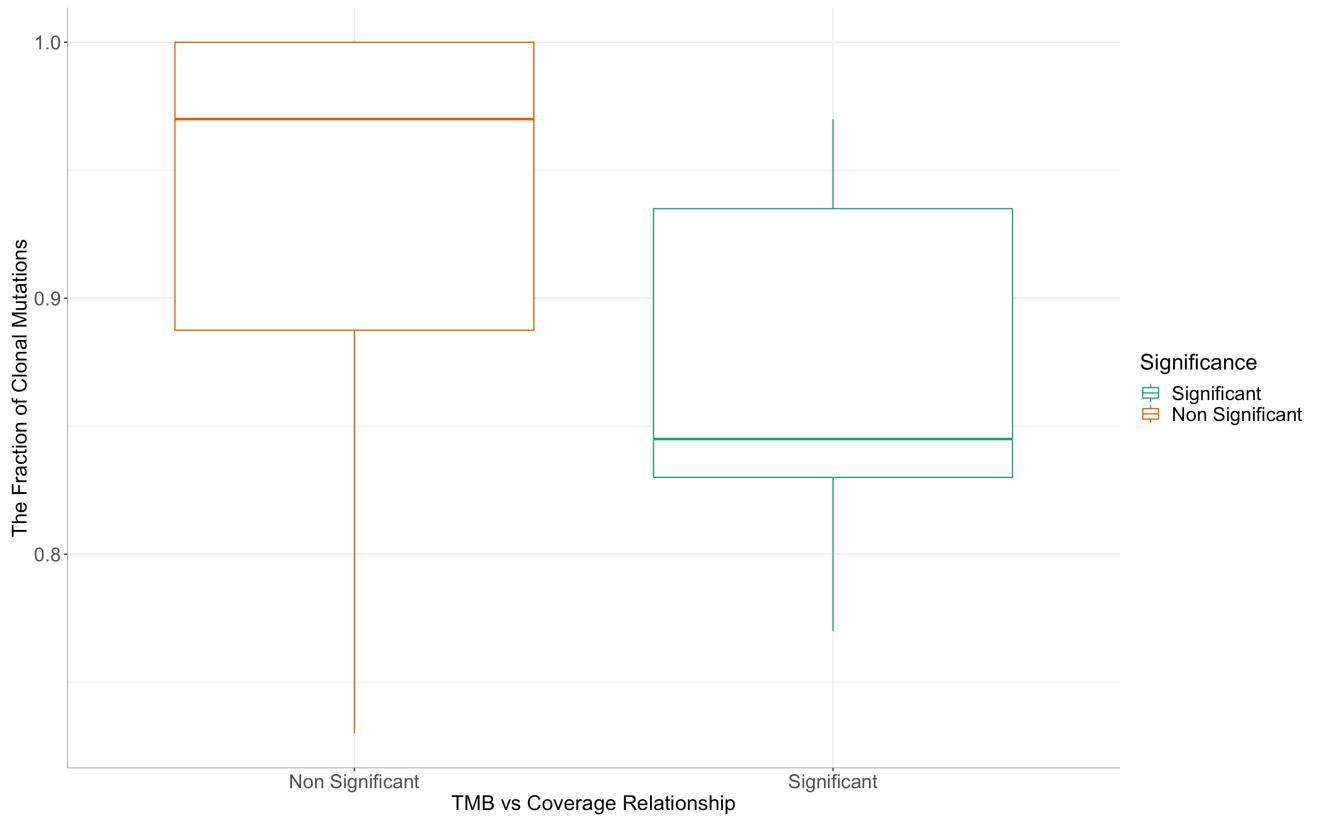


Fig. S6: In cancer types for which the correlation between TMB and coverage were found significant, the fraction of samples with high clonal proportion is generally lower. This suggest that the lack of correlation between TMB and mean coverage is due to the high clonal proportion.