Fig S1. Read depth per sample type in UZ Ghent cohort. Histogram of read depths for RP samples (left), PB samples (middle) and MLN samples (right).

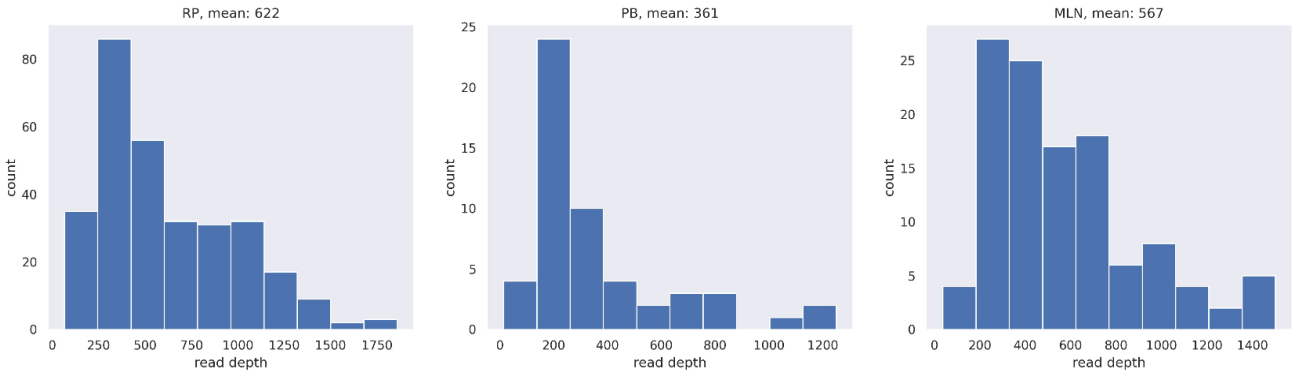


Fig S2. N times repeated cross-validation. The figure shows a simplified representation of our cross-validation strategy. Consider 12 patients (ID1-12) which are to be distributed in train/validation/test sets. Of these 12, four have a *TP53* mutation (TP53 bar). We first split off a held-out test set (bottom right corner) which is fixed and only used after all models are developed. The remaining 9 patients are to be split for three-fold cross validation. One possibility is shown on the left (cross validation config 1). In fold 1, patients with ID1-3 are used for validation; ID4-9 are used for training. In fold 2, ID4-6 are for validation (rest for training) and in fold 3, ID7-9 are for validation. Another possibility to make a three-fold cross validation is shown on the right (config N). Note that in all splits (train/val/test), at all times, the fraction of patients with mutation is consistent (1/3).

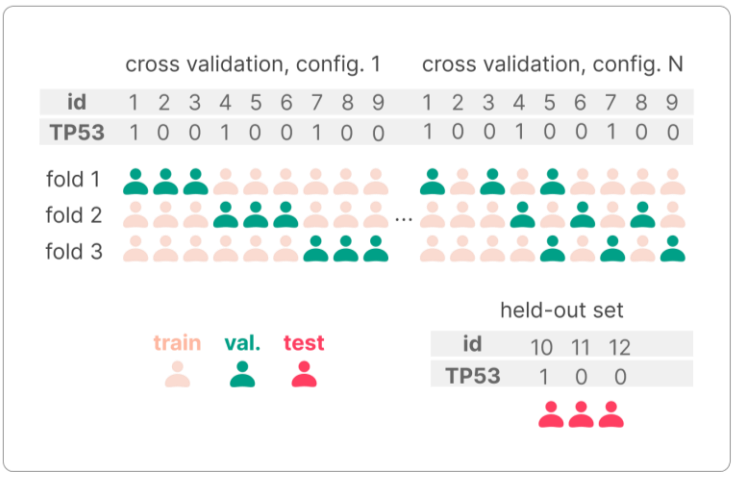


Fig S3. Receiver operating characteristic (ROC) curve for TCGA validation set with *BeTiDo*. The optimal threshold for predicted probability (0.437) is indicated in red. The optimal point was defined as the one with closest Euclidean distance to the optimal point (i.e. the point in the top left corner, where the TPR is 100% and FPR is 0%).

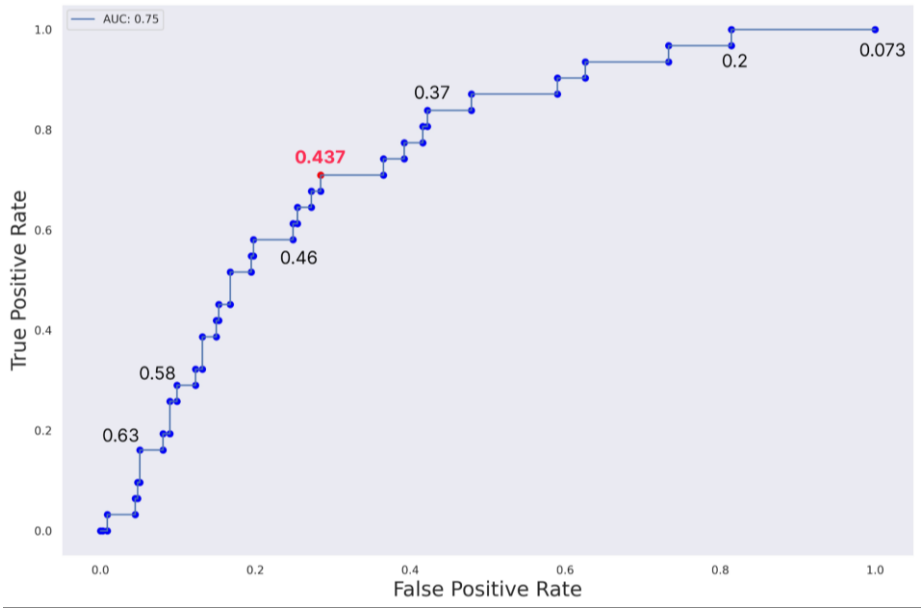


Fig S4. Confusion matrix on optimal prediction threshold for TCGA validation and test sets with *BeTiDo*. Left: normalized (across all patients). Right: unnormalized. The optimal prediction threshold was decided on the TCGA validation set based on the ROC curve (Suppl. Fig. 1)

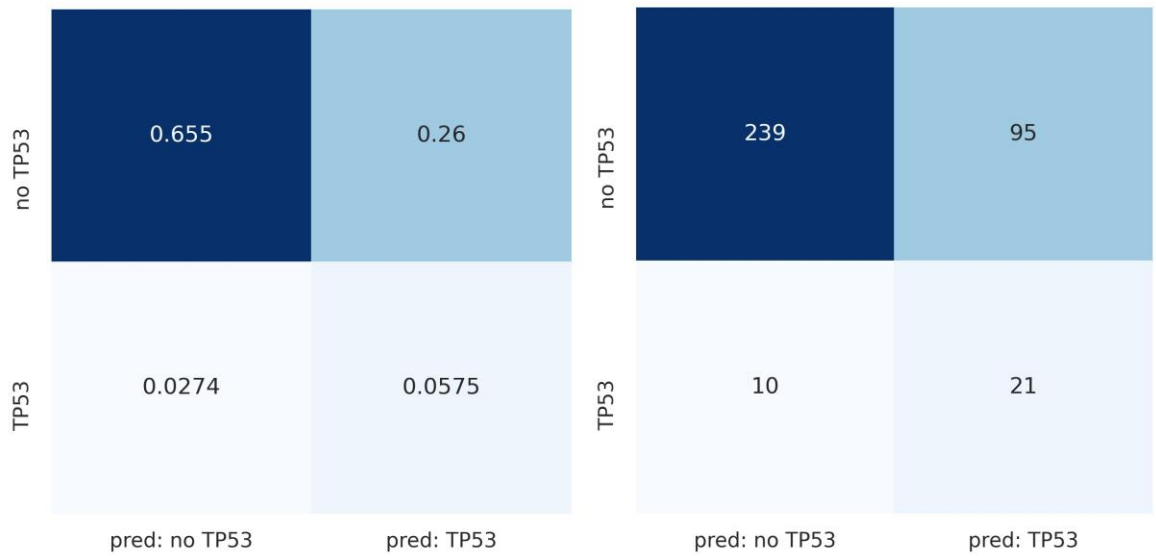


Fig S5. FN tiles for UZ Ghent cohort. Tiles where a *TP53* mutation label was found with sequencing, but where the model confidently predicted there is no *TP53* mutation (False Negatives). The predicted probability for a *TP53* mutation is shown on top of each tile.

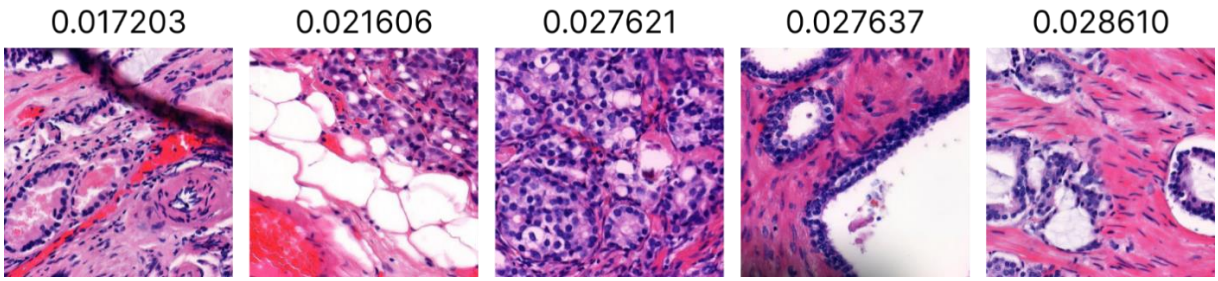
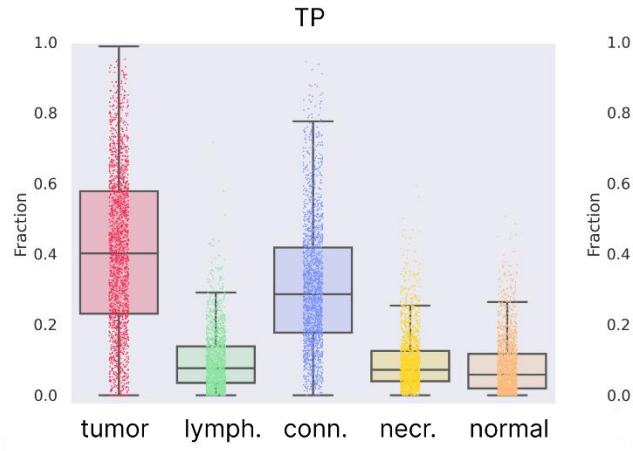


Fig S6. Predominant cell types in tiles predicted to (not) contain a *TP53* mutation in TCGA-PRAD. Top: 2000 tiles most confidently predicted to contain a mutation, for patients with mutation (left i.e. TP) and without mutation (right i.e. FP). Bottom: 2000 tiles most confidently predicted to not contain a mutation, for patients without mutation (i.e. TN). The table shows the average fraction of cell types for the three categories.



fraction of	TP	FP	TN
tumor cells	0.41 ↓	0.40 ↓	0.58 ↑
lymphocytes	0.10	0.08	0.05
conn. tissue	0.31 ↑	0.28 ↑	0.15 ↓
necrosis	0.09	0.10	0.12
normal cells	0.08	0.12	0.08

Fig S7. Association between model prediction and *TP53* CCF for the TCGA-PRAD cohort. a) high tumor purity (>=0.7); b) low tumor purity (<0.5).

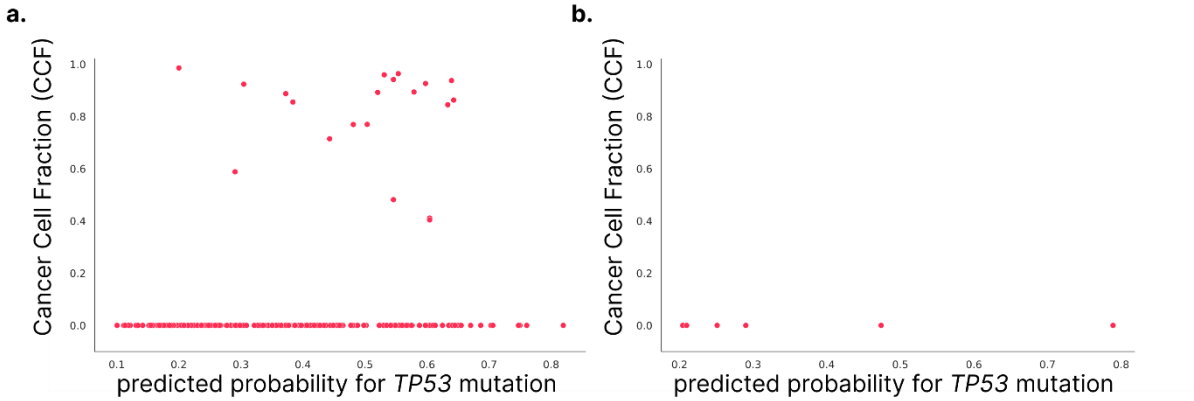


Fig S8. Cancer Cell Fraction (CCF) for different samples of the same patient in UZ Ghent cohort. Only shown for samples with tumor purity >= 0.8.

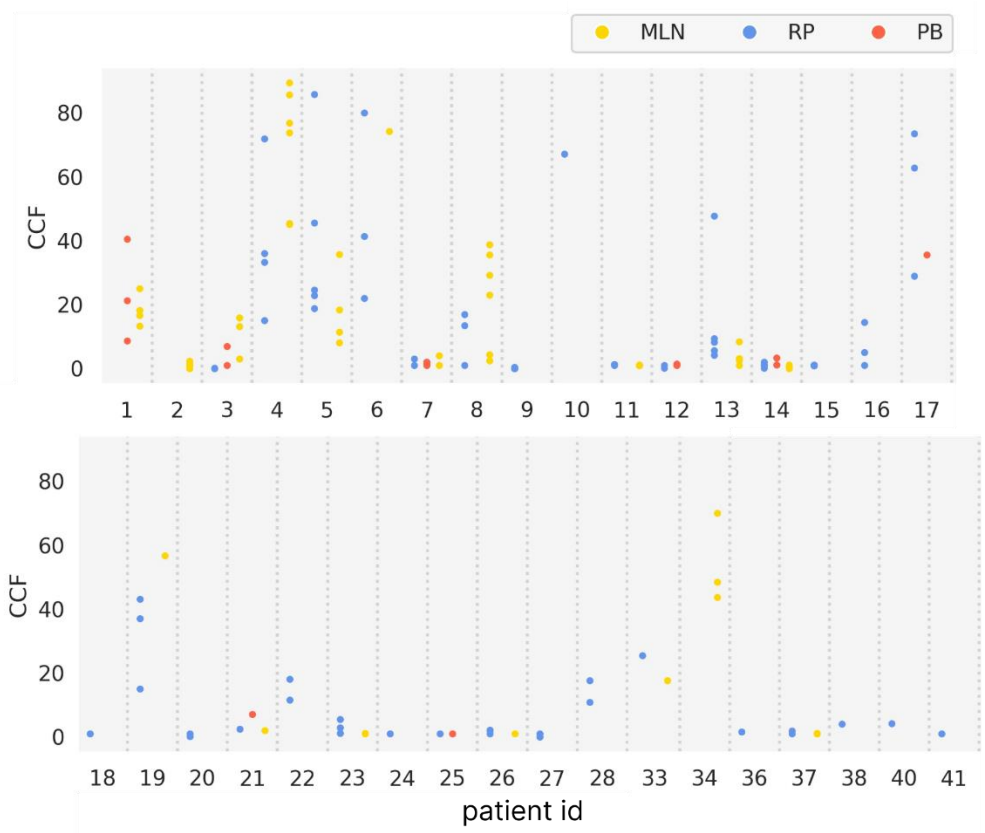


Fig S9. Median DMCCF for model predictions versus random baseline (UZ Ghent cohort).

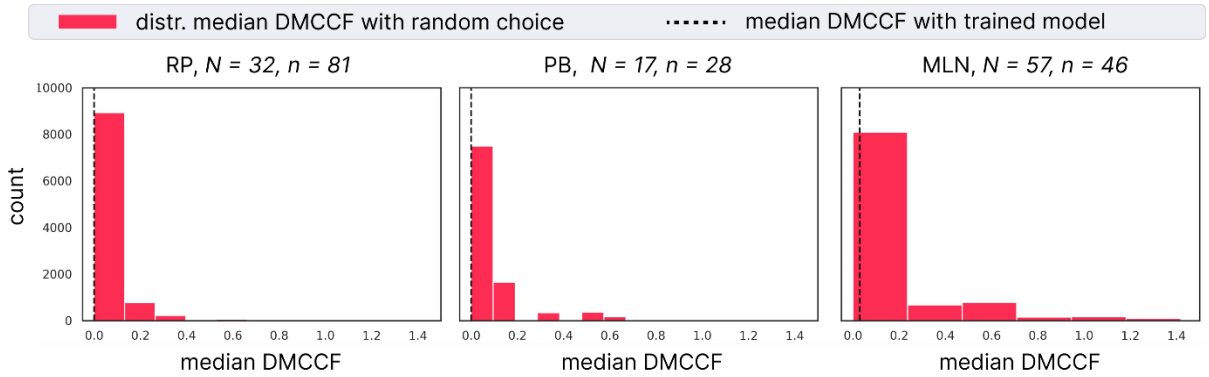


Fig S10. Prevalence of patients in TCGA-PRAD validation test sets with (red) and without (blue) lymph node metastasis. Shown for samples with lowest predicted probability for *TP53* mutation (left part, 0.25 quantile) and with highest predicted probability for *TP53* mutation (right part, 0.75 quantile). The number of patients whose lymph node metastasis status is unknown is shown next to '?' (those samples are not shown in the plot below).

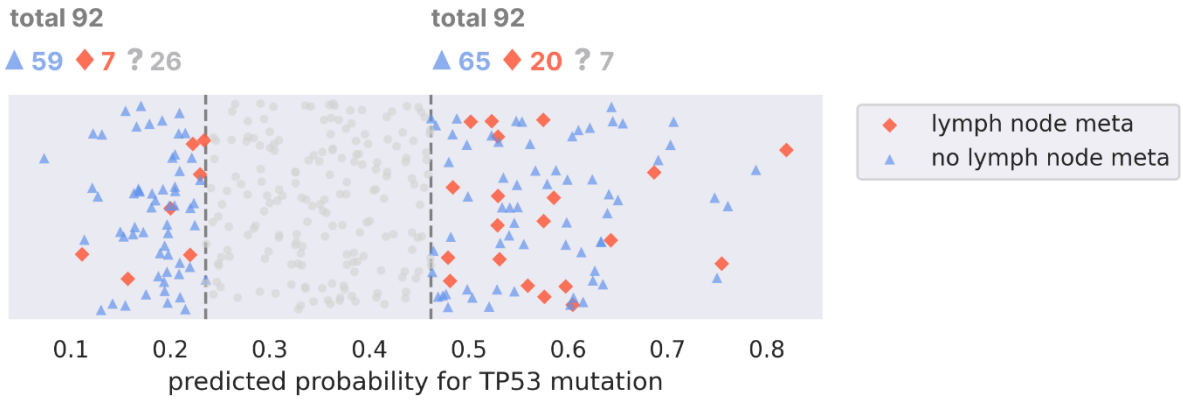


Fig S11. Prevalence of patients in TCGA-PRAD validation test sets with (red) and without (blue) biochemical recurrence. Shown for samples with lowest predicted probability for *TP53* mutation (left part, 0.25 quantile) and with highest predicted probability for *TP53* mutation (right part, 0.75 quantile). The number of patients whose biochemical recurrence status is unknown is shown next to '?' (those samples are not shown in the plot below).

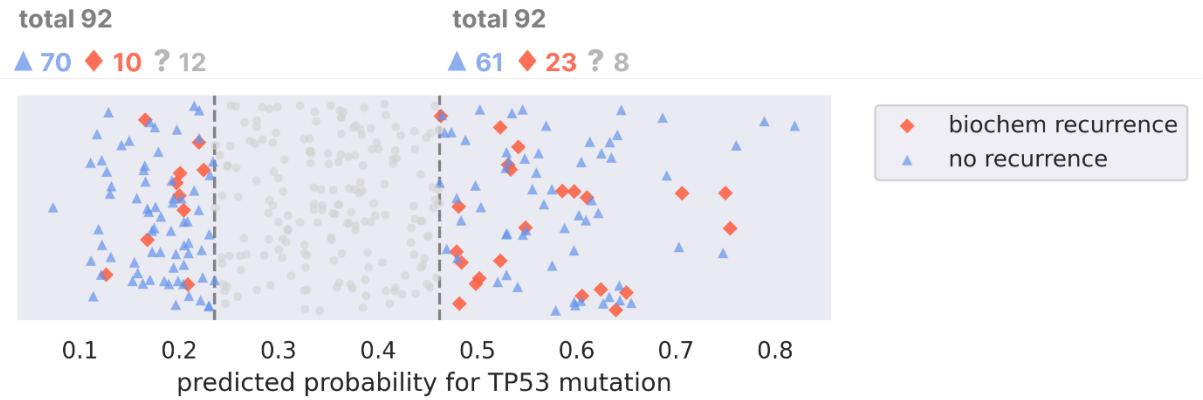


Fig S12. Tile-level predicted probability for mutation per Gleason Score, with lowest grade on the left, shown for the UZ Ghent cohort. Note that GS3+5 and GS5+3 are outliers since these grades are not frequently observed and only few samples are available (see Fig S12).

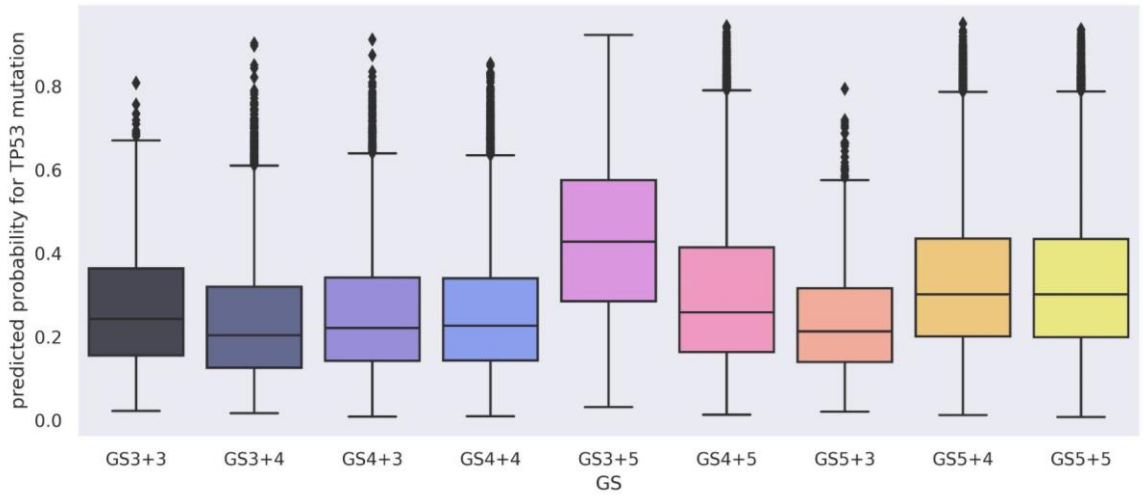


Fig S13. Number of tiles per Gleason Grade, shown for UZ Ghent cohort.

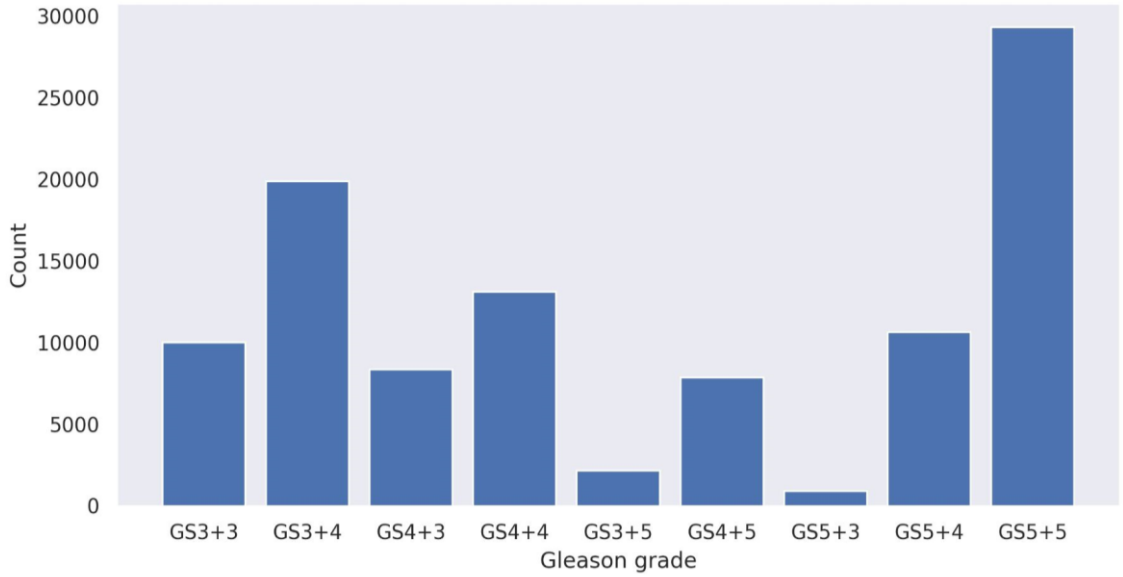


Fig S14. Performance (AUC) for lymph node status prediction on TCGA-PRAD, obtained with 100x repeated Monte-Carlo cross-validation. Significance calculated with paired t-test where ns: p > 0.05, *: 0.01 < p <= 0.05, **: 1e-3 < p <= 1e-2, ***: 1e-04 < p <= 1e-03, ****: p <= 1e-4. Label: indicates the true *TP53* status, Prob: probability of a *TP53* mutation being present as predicted by our model. Grade_nr: Gleason Grade of the lesion.

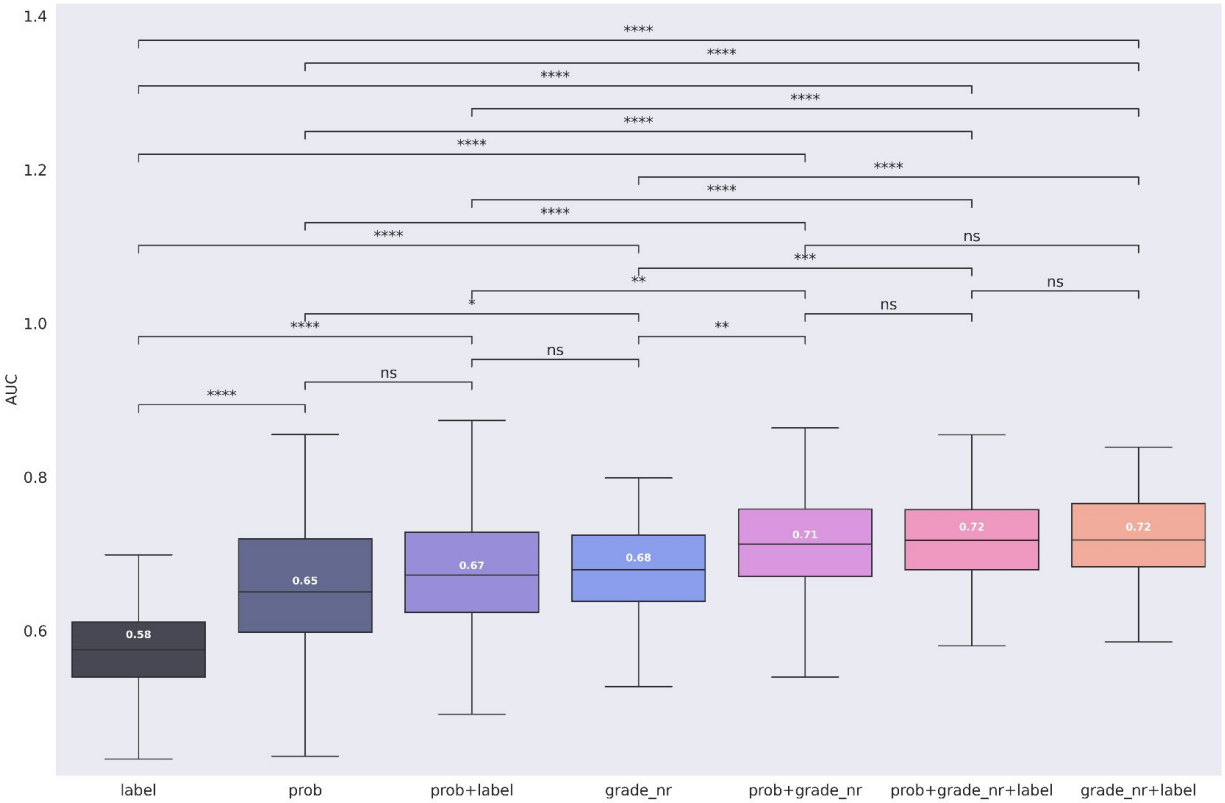


Fig S15. Performance (AUC) for biochemical recurrence prediction on TCGA-PRAD, obtained with 100x repeated Monte-Carlo cross-validation. Significance calculated with paired t-test where ns: p > 0.05, *: 0.01 < p <= 0.05, **: 1e-3 < p <= 1e-2, ***: 1e-04 < p <= 1e-03, ****: p <= 1e-4. Legend as in Fig. S10.

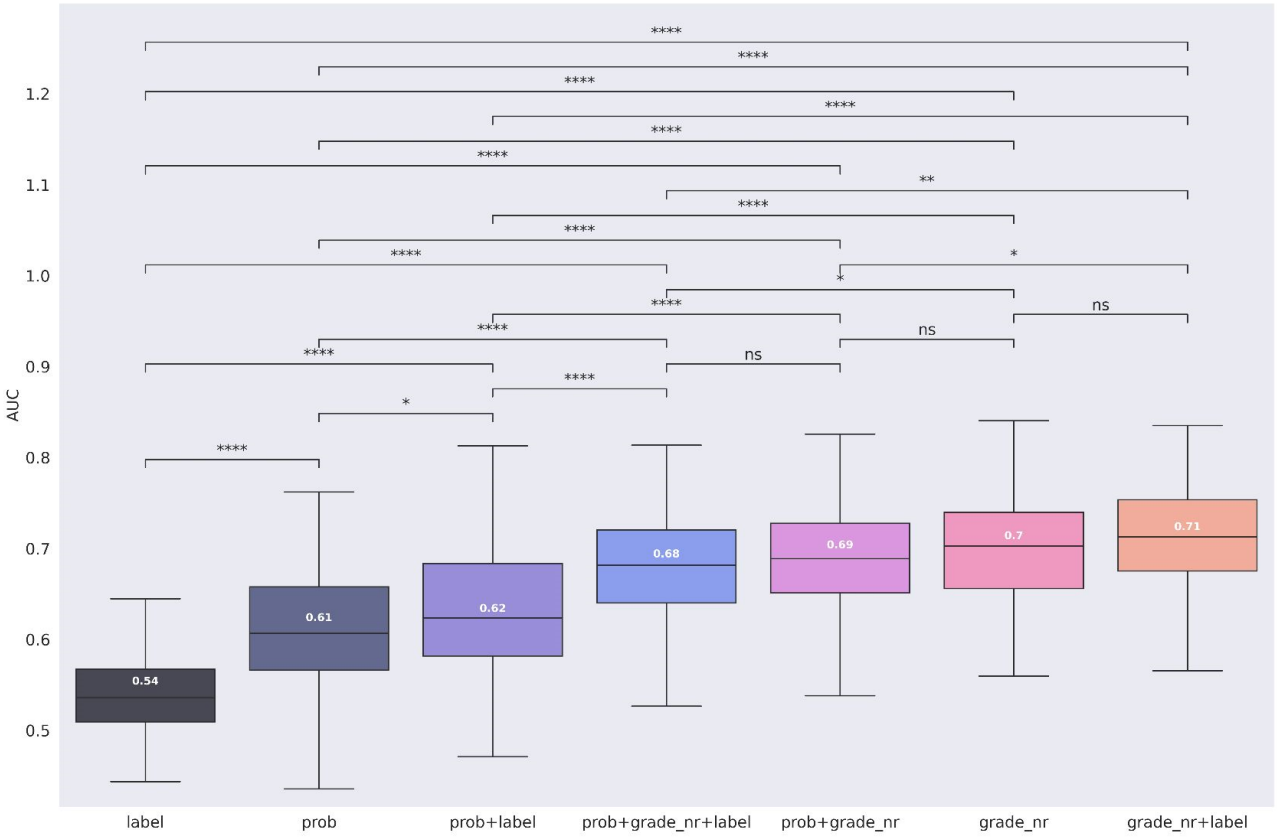


Table S1. Average number of patients and tiles in TCGA train/validation/test sets. Note that the number of patients/tiles in train and validation may vary slightly depending on specific folds and depending on specific cross-validation configurations (test is fixed).

	Train	**Validation**	**Test**
total nr of patients	194	97	74
nr of patients with *TP53* mutation	17	8	6
total nr of tiles in dominant region (with + without mutation)	112,182 (undersampled during training as described in methods)	56,091	42,058
total nr of tiles in all tumor regions (with + without mutation)	155,656 (undersampled during training as described in methods)	77,828	58,388
total nr of tiles in entire whole slide image (with + without mutation)	688,968 (undersampled during training as described in methods)	344,484	258,159

Table S2. Number of ("extreme") true/false positive/negative patients in TCGA-PRAD validation and test sets for *BeTiDo*.

	TP	**TN**	**FP**	**FN**
at optimal threshold	21	239	95	10
"extreme"	16 (0.5 quantile of positives)	84 (0.25 quantile of negatives)	84 (0.75 quantile of negatives)	/ (not used)

Table S3. Number of patients with biochemical recurrence and lymph node metastasis in TCGA-PRAD. Biochemical recurrence was defined for patients where *biochemical_recurrence* was indicated as 'YES' or where *days_to_first_biochemical_recurrence* was specified. Patients were defined not to have biochemical recurrence if their biochemical recurrence was indicated as 'NO' and if *days_to_first_biochemical_recurrence* was not specified.

	Yes	**No**	**Unknown**
Biochemical recurrence?	63	247	55
Lymph node metastasis?	48	257	60

Table S4. Number of patients, lesions, and tiles in UZ Ghent cohort. The first two rows show the total number of patients and tiles with/without mutations in the cohort. Note that some patients may have both lesions with and without mutation due to tumor heterogeneity, which is why the sum of patients with mutation (15) and without (34) is larger than the total number of patients (41). Then, the number of RP/PB/MLN lesions is shown along with the total number of tiles available of those regions as well as the unique number of patients that they originate from.

	With mutation	**Without mutation**
patients	15	34
tiles	44,446	104,747
RP lesions (nr patients/nr tiles)	64 (13 / 28,764)	167 (32 / 73,532)
PB lesions (nr patients/nr tiles)	21 (9 / 964)	30 (12 / 2,518)
MLN lesions (nr patients/nr tiles)	40 (12 / 14,718)	70 (20 / 28,697)

Table S5. Mean variability of CCF (UZ Ghent cohort). The mean standard deviation and mean absolute deviation of the Cancer Cell Fraction (CCF) for RP, PB and MLN samples of the same patient, and number of patients (*N*) for which the sample type is available. Shown for samples from Fig S6, i.e. samples with tumor purity >= 0.8.

mean variability of CCF	RP *N=32*	PB *N=8*	MLN *N=18*
mean standard deviation	4.45	2.21	3.85
mean absolute deviation	3.88	1.99	3.39

Supplementary Note 1. Model trained on *TP53* mutations + deletions.

We verified whether a model for the prediction of *TP53* mutation + deletion gives better performance than a model trained on *TP53* mutations (so considering a patient positive if they have mutation and/or deletion). As expected, the number of FP predictions is lower with this new model and the number of TP is higher (see confusion matrix below for mut+del model, and confusion matrix of the original model in Suppl. Fig. 3). However, now, the number of FN predictions is higher. The net result is that this new model gives similar performance as the original model trained on *TP53* mutations.

The reason for this similar performance (instead of an expected increase in performance) is likely because for cells carrying deletions additional genetic interactions might occur between the *TP53* alterations and other genes on the deleted regions that are not observed in samples carrying a mutation in *TP53*. So the sample set that becomes positively labeled is becoming more heterogenous than when only considering samples carrying *TP53* mutations (see also [1]). Indeed we observed in Figure 4b, that some patients with a deletion (9 patients) are assigned a low probability by our initial model. Hence, not all patients with deletion (in the absence of mutation) have the same histopathological phenotype as the patients where our model predicted a high probability. Therefore, we believe that including all patients with deletions as positive labels may have removed some wrong labels, but at the same time also added new noise, resulting in similar performance as before.

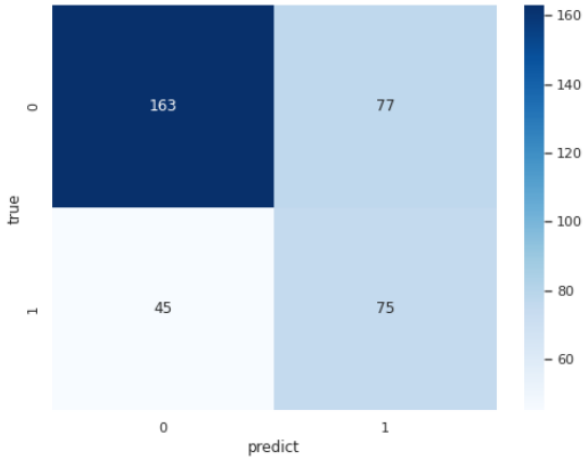


Figure: confusion matrix at optimal prediction threshold for *TP53* mut+del model. This confusion matrix can be compared to the one of the original *TP53* mutation model in Suppl. Fig. 3. The optimal prediction threshold was (in both cases) determined based on the ROC curve (see Suppl. Fig. 2 for more details).

[1] Liu Y, Chen C, Xu Z, et al. Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. Nature. 2016;531(7595):471-475