

Screen Them All: High-Throughput Pan-Cancer Genetic and Phenotypic Biomarker Screening from H&E Whole Slide Images

Yi Kan Wang^{1,3}, Ludmila Tydlitatova^{1,3}, Jeremy D. Kunz^{1,3}, Gerard Oakley¹, Ran A. Godrich¹, Matthew C. H. Lee¹, Chad Vanderbilt², Razik Yousfi¹, Thomas Fuchs¹, David S. Klimstra¹, and Siqi Liu^{1,4}

¹Paige, NYC, NY United States

²Memorial Sloan Kettering Cancer Center, NYC, NY United States

³Equal contribution

⁴Corresponding author (siqi.liu@paige.ai)

Abstract

Many molecular alterations serve as clinically prognostic or therapy-predictive biomarkers, typically detected using single or multi-gene molecular assays. However, these assays are expensive, tissue destructive and often take weeks to complete. Using AI on routine H&E WSIs offers a fast and economical approach to screen for multiple molecular biomarkers. We present a high-throughput AI-based system leveraging Virchow2, a foundation model pre-trained on 3 million slides, to interrogate genomic features previously determined by next-generation sequencing (NGS) assay, using 47,960 scanned hematoxylin and eosin (H&E) whole slide images (WSIs) from 38,984 cancer patients. Unlike traditional methods that train individual models for each biomarker or cancer type, our system employs a unified model to simultaneously predict a wide range of clinically relevant molecular biomarkers across cancer types. By training the network to replicate the MSK-IMPACT targeted biomarker panel of 505 genes, it identified 80 high performing biomarkers with a mean AUROC of 0.89 in 15 most common cancer types. In addition, 40 biomarkers demonstrated strong associations with specific cancer histologic subtypes. Furthermore, 58 biomarkers were associated with targets frequently assayed clinically for therapy selection and response prediction. The model can also predict the activity of five canonical signaling pathways, identify defects in DNA repair mechanisms, and predict genomic instability measured by tumor mutation burden, microsatellite instability (MSI), and chromosomal instability (CIN). The proposed model can offer potential to guide therapy selection, improve treatment efficacy, accelerate patient

screening for clinical trials and provoke the interrogation of new therapeutic targets.

1 Introduction

Modern cancer treatment decisions rely on several important factors such as the patient age, life expectancy, and the specific type, grade, and stage of their cancer [1, 2]. Tools such as nomograms, which can inform clinical treatment guidelines, help doctors estimate patient prognosis and recommend the best treatment options [3]. Treatments can range from monitoring the cancer without active treatment to surgery, hormone therapy, chemotherapy, radiotherapy, targeted therapy, immunotherapy and combinations thereof.

In recent years, there has been a significant push towards developing personalized treatments for patients based on the genetic alterations within their cancers [1, 2]. This has led to rapid advances in molecular biomarker tests, including single and multi-gene assays [4], which analyze tissue, blood, and body fluid samples to identify targetable genomic alterations and guide doctors in making more informed treatment decisions. Based on improved analysis of cancer genomics, novel targets with potential clinical relevance are reported every year. For example, genomic alterations such as androgen receptor (AR) variants help predict endocrine versus chemotherapy resistance in metastatic castrate resistant prostate cancer (mCRPC) [5], or BRCA1/2 germline mutations predict poly(ADP-ribose) polymerase (PARP) inhibitor response in the treatment of high-risk early stage HER2-negative breast cancer [6].

Commercially available multi-gene tests in localized disease [7], such as Oncotype DX Breast Recurrence Score Test, MammaPrint test, Oncotype DX Genomics Prostate Score, ProMark, Decipher and Prolaris, provide significant prognostic guidance across specific cancers, complementing routine clinicopathological factors in clinical decision-making [8]. However, these tests are often costly, time-consuming, and require substantial tissue samples, posing challenges particularly in small core biopsies or those with limited tumor cells. To address these limitations, newer assays using minimally invasive approaches, such as “liquid biopsy” blood draws for circulating tumor cell or circulating nucleic acid analysis, have been developed. Yet, they also face challenges with sample quantity [9], standardization of cell collection and stabilization procedures, and suffer from sensitivity and specificity issues due to non-tumor mutated clones present even in the blood of patients without cancer [10]. Thus, there is a growing need for digital biomarkers derived from widely available digital H&E whole slide images to rapidly and cost-effectively screen patient samples for multiple genomic biomarkers in a robust and tissue sparing manner. This approach enables the swift identification of cases that require definitive genomic testing for clinical management and appropriate therapy selection while excluding cases where such testing is unlikely to be fruitful, thereby improving turnaround time and reducing testing costs without compromising clinical care.

Beyond their clinical impact, digital biomarkers offer substantial advantages

for the pharmaceutical industry [11, 12], particularly for drug development and clinical trials. Digital biomarkers facilitate novel target identification and drug discovery, leading to the development of more effective targeted therapies. They can enhance patient stratification, which may improve the cost efficiency and success rates of clinical trials. Additionally, digital biomarkers support the creation of companion and complementary diagnostics to help personalize therapy selection to the genetic profile of the patient's tumor. By optimizing resource use and accelerating decision-making processes, digital biomarkers may contribute to cost savings in drug discovery and improve access to novel targeted therapies by reducing the cost of necessary clinical screening for healthcare systems and payers.

Recent studies have demonstrated methods to identify morphological features associated with genomic abnormalities in routine H&E histopathology images across various cancer types, enabling the prediction of digital molecular biomarkers [13, 14, 15, 16]. However, nearly all previous studies focused on only one biomarker for a specific tissue or cancer type at a time, a method that is notably inefficient. This inefficiency stems from two main issues: (1) each biomarker suffers from limited training data, hindering generalization, even though they may share morphological phenotypes, and (2) the costs associated with developing individual models are significant. Therefore the feasibility of detecting many digital biomarkers simultaneously has not been previously reported.

A universal model is essential to identify all clinically relevant molecular biomarkers from tissue-agnostic H&E whole slide images for all cancer types, which would benefit both clinical applications and pharmaceutical research. Here, we describe an approach to high-throughput screening for genomic abnormalities applicable to all cancer types using routine H&E whole slide images (Figure 1). By leveraging image representations from Virchow2, a foundation model pre-trained on 3 million slides [17, 18], we develop a model to simultaneously predict 1,228 genomic biomarkers, representing genomic abnormalities in 70 human cancers (Figure 2), by training and testing on a cohort obtained from Memorial Sloan Kettering Cancer Center (MSKCC) of 47,960 WSIs from 38,984 patients. Patients in the cohort had associated known ground truth genomic abnormalities from the paired tumor-normal targeted sequencing using the Food and Drug Administration (FDA)-cleared Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Targets (MSK-IMPACT) assay [19]. Additional genomic features such as deficient mismatch repair (dMMR) status were confirmed by immunohistochemistry (IHC) assay. Our model identified 391 genomic alteration biomarkers with $AUC > 0.75$ in the 15 most common cancer types treated at MSK. Evaluating phenotype-genotype associations revealed 40 histologic biomarkers, while 58 treatment-associated biomarkers were identified as predictors of response to FDA-approved drugs. The genomic biomarkers were further validated using diagnostic slides and genomic data from the TCGA Pan-Cancer Atlas cohort [20]. The AI model significantly enhances the efficiency of biomarker screening across various cancer types, identifying not only clinically relevant genomic abnormalities but also histologies characterized by specific ge-

nomic alterations. The system can potentially be utilized in patient screening for definitive genome analysis, guiding treatment selection, and identifying new therapeutic targets.

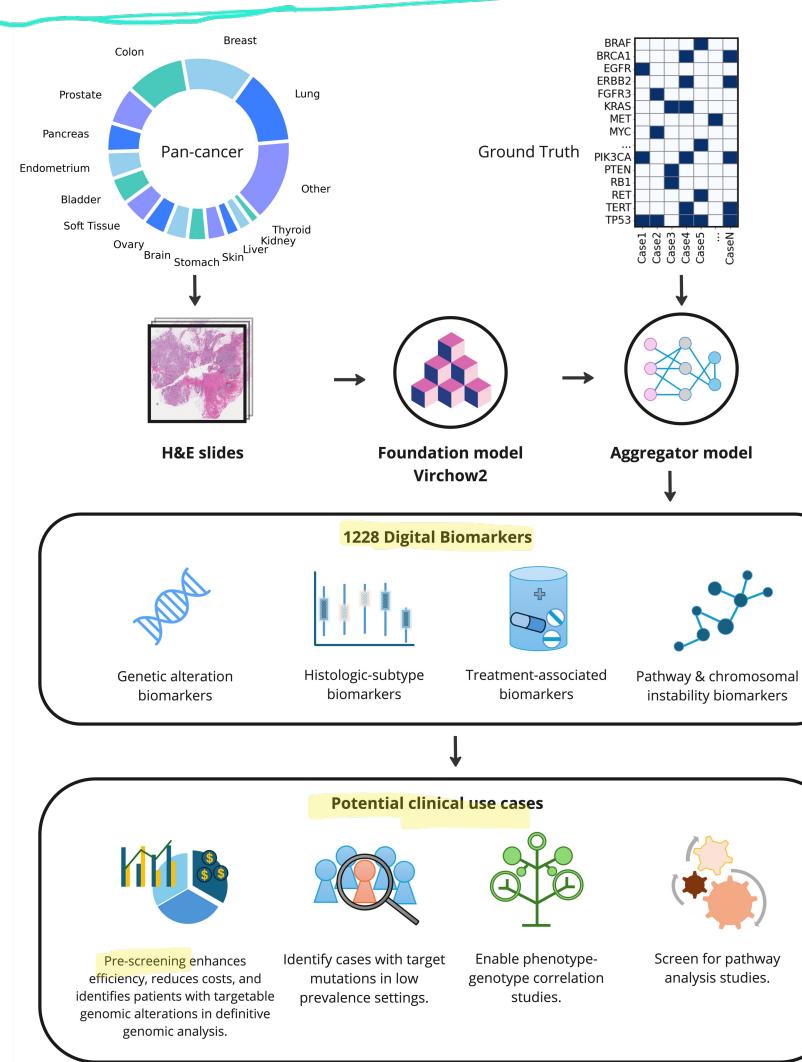


Figure 1: Overview of pan-cancer digital biomarker identification from H&E whole slide images (WSIs).

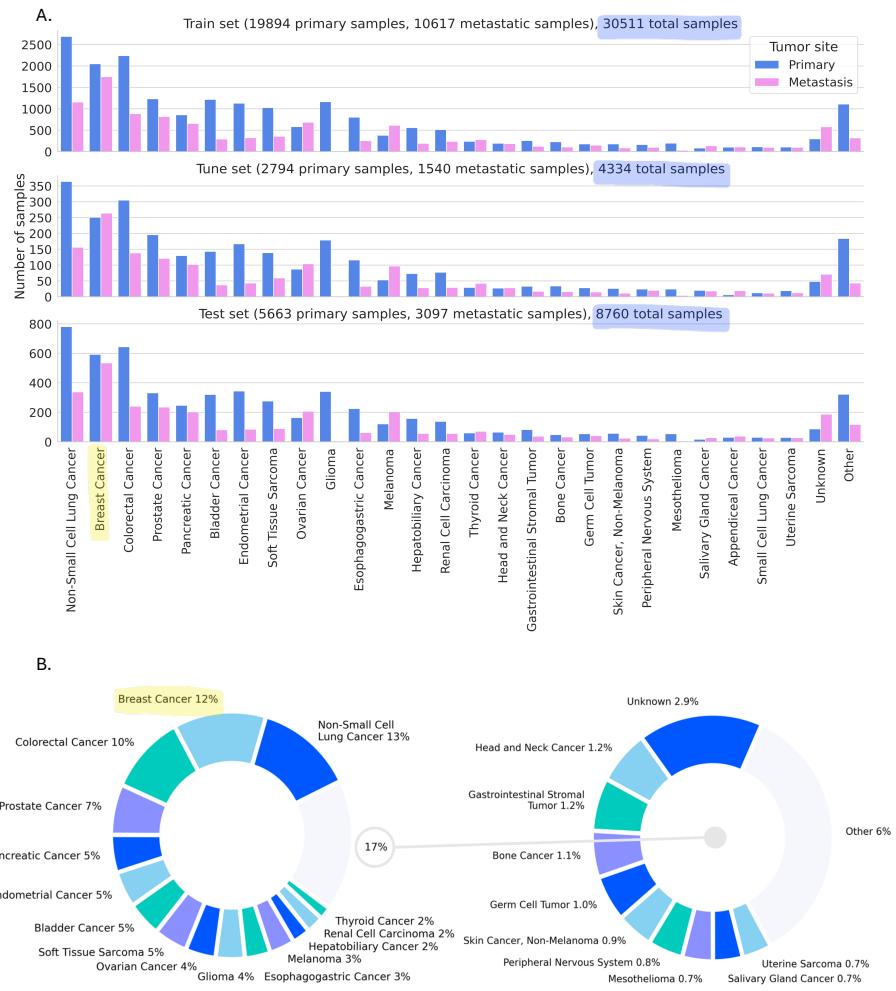


Figure 2: Distribution of MSK development dataset. A: Distribution of primary vs metastatic samples in 27 most common cancer types (including a 'Unknown' category, in which the cancer type of samples is not available) in the MSK dataset, split by train, tune and test. The 'Other' category comprises the remaining cancers, accounting for 6% of all samples. **B:** Donut charts showing the distribution of histologies in the MSK dataset. The left chart displays the 15 most common cancer types, representing 83% of all samples. The right chart shows the composition of cancer types in the remaining 17% of samples, including those in the 'Unknown' category and rare cancer samples in the 'Other' category.

2 Results

2.1 Digital biomarker panel enables high-throughput pan-cancer biomarker screening

Our model predicts the likelihood of the presence of each of 1,228 genomic features derived from the MSK-IMPACT dataset (see Section 4.2 for model development details). The model is intended to screen for any genomic alteration among 505 genes to prioritize or guide further confirmatory sequencing.

By computing the average precision (AP) and area under the receiver operating characteristic curve (AUC) for 1,228 biomarkers in the MSK test set, a positive correlation between AP and the ratio of positive samples of labels in the training set as expected (Figure 3A). In contrast, there was no strong correlation between AUC and the ratio of positive samples. We also evaluated the sensitivity and specificity for biomarkers showing $AUC > 0.5$ (Figure 3B). We identified 655 genetic alteration biomarkers, representing 320 genes, in 44 cancer types that were above the threshold expected by chance, with an $AUC > 0.5$, and at least 3 positive samples with a positive ratio $> 2\%$. We observed variability in the accuracy of biomarker prediction across different cancer types, which can be attributed to differences in sample size and prevalence. (Figure 3C).

By evaluating the performance of each biomarker in 7,208 samples (4,744 primary and 2,464 metastatic samples) of the 15 most common cancer types treated at MSK, the AI model we developed was able to detect 391 genetic alteration biomarkers in the test set, based on the filtering criteria of $AUC > 0.75$, sensitivity > 0.75 and specificity > 0.2 . This set of biomarkers, with a mean AUC of 0.84 (mean sensitivity = 0.92 and mean specificity = 0.55), represent 483 distinct gene and tumor histology associations, i.e., a signal in at least one cancer type for 222 genes (Table S3). A prediction signal for TP53 oncogenic mutations was identified in 14 of the most commonly treated cancer types at MSK, meeting or exceeding the filtering criteria used. The mean AUC was 0.86 ($n=13$, range 0.76-0.94) in primary samples and 0.84 ($n=8$, range 0.77-0.97) in metastatic samples (Figure S3A, Table S3). TP53 was also detected in esophagogastric cancer, achieving a high AP of 0.77 in primary cancers ($n=148$ positive and $n=77$ negative samples), with a sensitivity of 0.98 and a specificity of 0.34. However, the AUC of 0.7 fell below the filtering criterion of 0.75. Other commonly mutated oncogenes where high prediction AUCs for alteration were obtained in 5 or more of the most commonly treated cancer types at MSK included CDKN2A, CDKN2B, TERT, KRAS, CCND1, PTEN, RB1, PIK3CA, BRCA1, AGO2, ARID1A, CDK12, CTNNB1, ERBB2, KMT2B, MYC, NF1 (Figure S3A). Overall, the AUCs for prediction of a genomic alteration were higher for WSIs obtained from primary tumors versus WSIs obtained from samples of metastatic lesions (Figure S4).

Among the 15 most commonly treated cancer types at MSK, colorectal cancers had the highest number of genes with an AI model prediction $AUC > 0.75$ on the test set ($n=61$ genes in primary and $n=30$ genes in metastatic samples), followed by endometrial cancers ($n=61$ in primary and $n=19$ in metastatic sam-

ples) (Figure S3B). Breast cancers ($n=41$ genes in primary and $n=17$ genes in metastatic samples), bladder cancers ($n=26$ genes in primary and $n=21$ genes in metastatic samples), and ovarian cancers ($n=29$ genes in primary and $n=20$ genes in metastatic samples) also showed notable numbers of genes with $AUC > 0.75$. Since gliomas essentially never metastasize outside the central nervous system, no metastases were analyzed for this cancer type ($n=20$ genes in primary samples).

By assessing high-performing biomarkers (defined as biomarkers achieving $AUC > 0.85$) with inclusion criteria of sensitivity > 0.8 and specificity > 0.3 , 80 biomarkers (with at least 5% positive sample ratio), representing 47 genes, were identified in the 15 most common cancers with a mean AUC of 0.89 (AUC range 0.85-0.99, mean sensitivity=0.93, mean specificity=0.66) (Figures 3D, S5A). Top-performing genes with the best AUC scores in primary cancers included FGFR3 (bladder), CDH1 (breast), MSH3 (colorectal), TP53 and APC (endometrial), KMT2D (esophagogastric), IDH1 (glioma), SMAD4 (hepatobiliary), AGO2 (melanoma), STK11 and EGFR (NSCLC), TP53 and KRAS (ovarian), CDKN2B (pancreatic), BAP1 (renal cell carcinoma), CDKN2A (p16INK4a) (soft tissue sarcoma), and CDKN2A (p14ARF) (thyroid).

To assess the AI model's generalizability, further validation was conducted on an independent external cohort from The Cancer Genome Atlas (TCGA) dataset ($n=9,340$, Section 4.1.2). The optimal operating threshold determined from the tune set of the development cohort was used to convert the likelihood of a genomic alteration in each sample to binary predictions, indicating the presence or absence of the genetic variant in a gene. 27 of the top performing genes identified from the MSK test set were validated in the TCGA cohort, with a mean AUC of 0.87 (range 0.77-0.94, mean sensitivity=0.91, mean specificity=0.6), including FGFR3 (bladder), CDH1 and ERBB2 (breast), BRAF and RNF43 (colorectal), PTEN (endometrial), KMT2B (gastric), IDH and ATRX (glioma), STK11 and EGFR (non-small cell lung cancer), BRAF and NRAS (thyroid). TP53 was validated in both breast and endometrial cancers, and KMT2D was validated in both colorectal and gastric cancers (Tables 2, S3, S4).

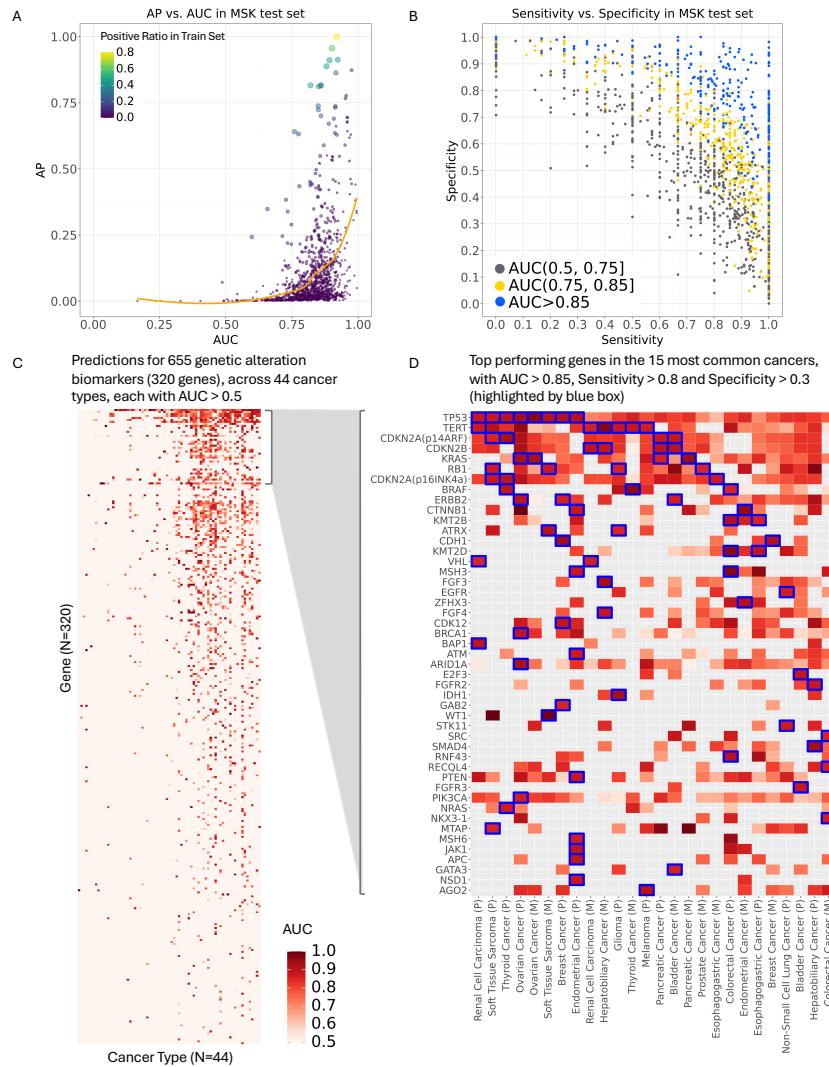


Figure 3: Overview of digital pan-cancer biomarker prediction performance. A: Scatter plot showing the average precision (AP) vs. AUC of biomarkers obtained in MSK test set. Color and size indicating the proportion of positive samples in the train set. **B:** Scatter plot showing sensitivity vs. specificity of biomarkers obtained in MSK test set. Color coded by the ranges of AUC: AUC > 0.5 and ≤ 0.75 , AUC > 0.75 and ≤ 0.85 , and AUC > 0.85. **C:** An overview heatmap showing the prediction of 655 genetic alteration biomarkers, with AUC > 0.5, representing 320 genes, across 44 cancers in the MSK test set. Biomarker labels included had at least 3 positive samples with a positive ratio > 2%. Color coded by AUC. **D:** Heatmap of gene biomarkers, with AUC > 0.5, identified from primary (P) and metastatic (M) cancers of the most common histologies. Blue boxes highlighted the top performing genes (N=47) in which the corresponding genetic alteration biomarkers (N=80) achieved an AUC > 0.85, sensitivity > 0.8 and specificity > 0.3.

2.2 Biomarkers associated with histologic subtypes of cancers

We next sought to determine whether the model trained to detect genomic alterations could be employed to diagnose specific histologic subtypes of cancers, particularly those in which certain genomic alterations are diagnostic or show high concordance with specific phenotypes. For each top performing genomic biomarker within a cancer type, we compared the test set inference probabilities of the biomarker between the histologic subtypes. Our analysis revealed 40 genomic alteration predictions that were highly associated with specific histologic subtypes (Kolmogorov-Smirnov (KS) test adjusted p-value < 0.01, and AUC > 0.85; Figure 4, Table 3).

We subsequently assessed the performance of these subtype-specific biomarkers in predicting the histologic subtype diagnosis within the withheld test set. Rather than using genomic alterations as the ground truth, we evaluated performance based on the actual cancer and subtype diagnosis assigned to each case. Specifically, we examined whether the model's prediction of subtype-associated genomic alterations could accurately identify the diagnosis of the corresponding subtype. For example, CDH1 alteration has a strong phenotype-genotype correlation with the invasive lobular carcinoma subtype of breast carcinoma. The model prediction of CDH1 oncogenic mutation presence was able to diagnose breast invasive lobular carcinoma with an AUC of 0.93, sensitivity of 0.94, and specificity of 0.77. This association was also validated in the TCGA cohort, showing an AUC of 0.95. Similarly, within thyroid carcinomas, prediction of a RET oncogenic mutation by the AI model achieved an AUC of 0.99 in diagnosing medullary thyroid cancer (MTC). Deleterious mutations of ARID1A, a member of the SWI/SNF chromatin remodeling genes, were common in clear cell and endometrioid ovarian cancers [21]. The prediction of oncogenic mutation/deletion in ARID1A displayed an AUC of 0.85 and 0.93, respectively, for identifying ovarian clear cell carcinoma (OCCC) ($n=22$, sensitivity=0.82, and specificity=0.79) and endometrioid ovarian cancers ($n=15$, sensitivity=0.93, specificity=0.76). KMT2D deficiency drives lung squamous cell carcinoma (LUSC) [22] and our evaluation showed that the diagnosis of LUSC by the prediction of KMT2D oncogenic mutation/deletion achieved an AUC of 0.90, with sensitivity of 0.96 and specificity of 0.54.

In soft tissue sarcomas, prediction of MDM2 amplification identified well-differentiated liposarcoma with AUC=0.93 (sensitivity=0.94, specificity=0.55), driven by the high prevalence of MDM2 amplification in this subtype [23]. When predicting well-differentiated liposarcoma ($n=17$) with atypical lipomatous tumor ($n=2$) and dedifferentiated liposarcoma ($n=35$), it showed a performance of 0.85 sensitivity and 0.6 specificity (AUC=0.85). The diagnosis of dedifferentiated liposarcoma via MDM2 amplification prediction was also validated in TCGA cohort, demonstrating an AUC of 0.84. Moreover, TERT oncogenic mutation prediction correctly predicted myxoid liposarcoma diagnosis with AUC=0.93, sensitivity=1 and specificity=0.29, while the prediction of RB1 loss showed an AUC of 0.86 for diagnosis of leiomyosarcoma with sensitiv-

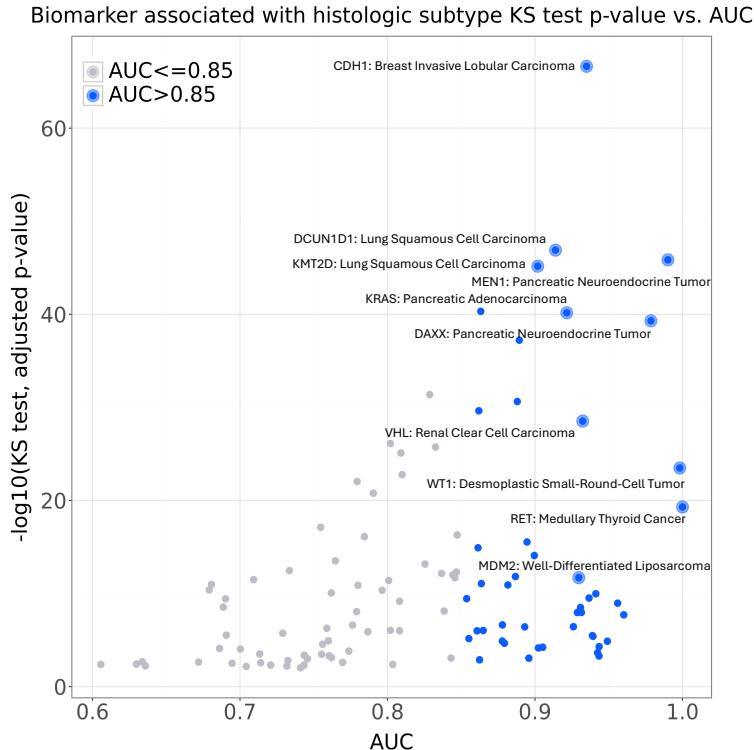


Figure 4: Biomarkers associated with histologic subtypes of cancers. Scatter plot showing $-\log_{10}(\text{KS test, adjusted P value})$ vs. AUC evaluated by using biomarker inference probabilities as prediction and histologic subtype as ground truth. Grey dots indicate $\text{AUC} \leq 0.85$, blue dots represent the biomarkers with $\text{AUC} > 0.85$. Text labels highlights a set of example biomarkers that are highly associated with histologic biomarkers, with $-\log_{10}(\text{KS test, adjusted P value}) > 15$ and $\text{AUC} > 0.9$

ity of 0.9 and specificity of 0.67. Finally, among sarcomas, WT1 fusion prediction diagnosed desmoplastic small round cell tumor which canonically carries a EWSR1-WT1 fusion [24] with $\text{AUC}=0.998$, sensitivity=0.76 and specificity=1.

In pancreatic cancers, MEN1 and DAXX were frequently mutated in pancreatic neuroendocrine tumor [25]. Prediction of oncogenic mutation/deletion in these two genes could diagnose pancreatic neuroendocrine tumor with $\text{AUC} > 0.97$ (sensitivity=0.79 and 0.55; specificity=0.99 and 1, respectively), while KRAS oncogenic mutation prediction correctly identified pancreatic adenocar-

cinoma diagnosis with $AUC=0.92$, sensitivity=0.87 and specificity=0.85. In renal cell carcinoma, loss of function in VHL is a hallmark of clear cell renal cell carcinoma (ccRCC) [26], and the prediction of VHL oncogenic mutation correctly identified ccRCC with an AUC of 0.93 (sensitivity=0.95 and specificity=0.83). Renal angiomyolipoma, though low prevalence ($n=5$) in the test set, was accurately detected by TSC2 oncogenic mutation prediction, achieving an AUC of 0.94, sensitivity of 1 and specificity 0.73. Both GNAQ and GNA11 oncogenic mutations were highly associated with uveal melanoma (KS adjusted p-value < 0.01), and prediction of genomic alteration in these genes in melanoma showed an AUC of 0.94 and 0.93, respectively, for diagnosis of uveal melanoma.

In glioma, oligodendrogloma is genetically defined by an IDH mutation and 1p19q codeletion [27]. The prediction of IDH1 oncogenic mutation diagnosed oligodendrogloma with an AUC of 0.89, sensitivity of 0.94 and specificity of 0.73. The performance remained comparable ($AUC=0.89$, sensitivity=0.89 and specificity=0.75) when combining oligodendrogloma ($n=17$) with anaplastic oligodendrogloma ($n=11$). Both ARTX and IDH1 were strongly associated with anaplastic astrocytoma (KS-test adjusted p-value < 0.01), and the prediction of oncogenic mutations in these genes showed an $AUC > 0.86$, for diagnosis of anaplastic astrocytoma ($n=22$).

2.3 Biomarkers associated with targeted therapeutic hotspots

We next evaluated the model's predictions of genomic alterations that are indicative of response to corresponding FDA-approved drugs across different cancers. We compiled a list of 54 treatment-associated genes with specific hotspot mutations reported in My Cancer Genome (MCG) [28, 29] and OncoKB [30, 31], focusing on the actionable targets that are FDA-recognized biomarkers (OncoKB therapeutic evidence level 1) or standard care biomarkers recommended by the National Comprehensive Cancer Network (NCCN) or other expert panels (OncoKB therapeutic evidence level 2). The therapeutic target ground truth was then established accordingly for each sample in the test set, based on the presence or absence of these specific mutations in the treatment-associated genes. By assessing the performance of the biomarkers trained by the AI model in predicting these therapeutic targets, our analysis identified 58 clinically relevant biomarkers in the 15 most common cancers with $AUC > 0.75$, sensitivity > 0.70 and specificity > 0.20 (Figures 5, S5B, Table 4). For example, BRAF V600E mutations are commonly found in melanoma, glioma, thyroid, lung and colorectal cancers, and are actionable targets as a standard of care in a subset of patients with these cancers [32]. The performance of BRAF oncogenic alterations in predicting BRAF V600E mutations showed an AUC of 0.93 in primary thyroid cancers, with sensitivity of 0.89 and specificity of 0.81, and an AUC of 0.96 in metastatic thyroid cancers, with sensitivity of 0.94 and specificity of 0.78. Though not all biomarkers had sufficient positive samples (mutation carriers) to be evaluated in metastatic cancers, the evaluation on primary samples showed that detection of BRAF oncogenic alterations achieved an AUC of 0.87 in primary melanoma, with sensitivity of 0.95 and specificity of 0.30, and an AUC

of 0.91 in primary colorectal cancers, with sensitivity of 0.98 and specificity of 0.48.

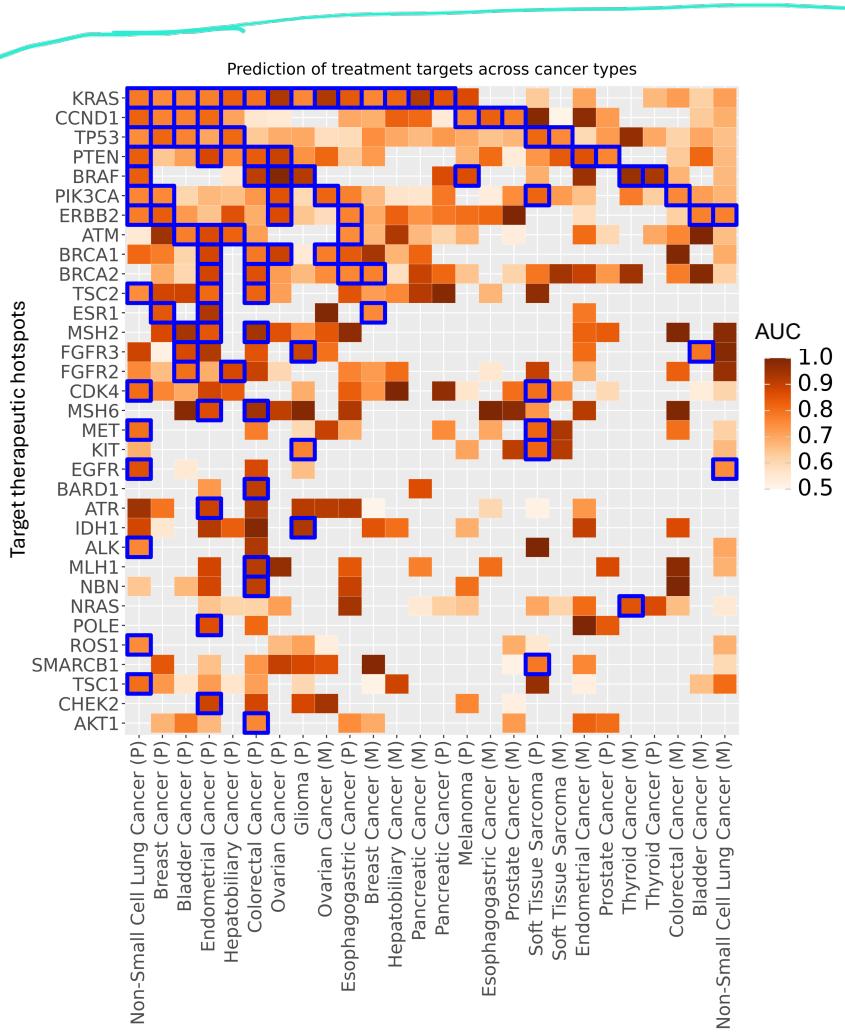


Figure 5: Heatmap showing the prediction of specific therapeutic hotspots in target genes (Y-axis) across common cancer types at primary (P) and metastatic (M) lesions, with $AUC > 0.5$. The AUCs were evaluated by using biomarker inference probabilities as prediction and presence or absence of specific hotspot mutations in a targeted gene as ground truth. Blue box highlighted the targeted genes harboring the hotspot mutations detected by our AI model, which achieved an $AUC > 0.75$, sensitivity > 0.7 , specificity > 0.2 , with at least 5 positive samples in the corresponding genetic alteration biomarkers ($N=58$) associated with 33 genes.

Trastuzumab is approved for the treatment of early-stage HER2+ (ERBB2-amplified) breast cancers [33]. Prediction of ERBB2 amplification showed an AUC of 0.84 (sensitivity=0.91 and specificity=0.54) in primary breast cancers and an AUC of 0.77 (sensitivity=0.89 and specificity=0.29) in primary gastric cancers. In addition, Fam-Trastuzumab Deruxtecan-nxki is approved for treatment of unresectable or metastatic non-small cell lung cancer (NSCLC), where activating mutations and amplification of ERBB2 are common mechanisms for upregulation of HER2 expression. Our evaluation showed that the prediction of ERBB2 oncogenic amplification or mutation in metastatic NSCLC (n=9 positive samples and n=329 negative samples) had a sensitivity of 0.89 and specificity of 0.27 (AUC=0.77), which showed a similar performance in primary NSCLC (n=30 positive samples and n=751 negative samples) with higher specificity (AUC=0.78, sensitivity=0.87 and specificity=0.46). The NPV of NSCLC in both primary and metastatic samples were 99%, which potentially could identify patients do not have ERBB2 amplification or mutation, hence would not benefit for HER2-targeted therapy. Similarly, an AUC of 0.77 with sensitivity of 0.93 and specificity of 0.34 was achieved in metastatic bladder cancers (n=14 positive samples and n=67 negative samples), though low prevalence, suggesting a subset of bladder cancer patients could be screened who might benefit from HER2-targeted therapy.

FDA-approved receptor tyrosine kinase inhibitors (TKIs) such as gefitinib, erlotinib, afatinib, and osimertinib are indicated for patients with specific oncogenic mutations in EGFR in the setting of advanced or metastatic NSCLC [34]. The AI model's performance predicting the TKI-targetable EGFR mutations (p.L858R, p.T790M, exon 19 deletion and exon 20 insertion) achieved an AUC of 0.86 (sensitivity=0.94 and specificity=0.45) in primary NSCLC and an AUC of 0.75 (sensitivity=0.86 and specificity=0.5) in metastatic NSCLC. Sotorasib and adagrasib are approved small molecule inhibitors targeting KRAS p.G12C carriers in NSCLC. The KRAS p.G12C is the most common KRAS variant found in NSCLC patients, and our AI model achieved an AUC of 0.78 (sensitivity=0.96 and specificity=0.30) for detection of this variant in primary NSCLC samples.

Tepotinib is another TKI approved in NSCLC and targets MET exon 14 skipping mutations, typically due to changes affecting RNA splicing [35, 36]. The performance of the MET oncogenic alteration AI model in predicting MET exon 14 deletion/splicing mutations showed an AUC of 0.80, sensitivity of 0.86 and specificity of 0.42 in primary NSCLC. Two other genes with targeted therapies available in NSCLC treatment were identified well by the AI model (AUC > 0.75). ALK fusion detection in primary NSCLC had AUC=0.76, sensitivity=0.78, specificity=0.52 while ROS1 fusion detection in primary NSCLC had AUC=0.75, sensitivity=0.75, specificity=0.52. Both ALK1 and ROS1 fusion detection results in primary NSCLC showed a high NPV of 0.99.

Elacestrant is approved by the FDA for patients with ER+, HER2-, and ESR1 mutated metastatic breast cancer [37]. Prediction of ESR1 hotspot mutations (including D538, E380, L536, S463P, Y537), using our AI model trained on ESR1 oncogenic mutations, achieved an AUC of 0.85 (sensitivity=0.86, speci-

ficity=0.62) in primary breast cancer, and an AUC of 0.76 (sensitivity=0.90, specificity=0.34) in metastatic breast cancer.

Other therapeutic targets for ER+/PR+ and HER2– patients with locally advanced or metastatic breast cancer include PIK3CA, AKT1, and PTEN alterations. Prediction of PIK3CA hotspot mutations showed a mean AUC of 0.8 (range 0.76-0.86) in five cancers with different prevalence of PIK3CA. Primary breast cancer had a high prevalence of PIK3CA mutation (n=182, 31%), and the AI model demonstrated an AUC of 0.76 (sensitivity=0.95, and specificity=0.32) predicting a PIK3CA mutation in this setting. Prediction of PIK3CA mutants in metastatic colorectal cancers (n=26, 11%) and primary NSCLC (n=35, 4%) had similar performance as breast cancer, showing an AUC of 0.76. The AI model showed better prediction of PIK3CA in ovarian cancers, with AUC=0.86 (n=13, 8%) for PIK3CA mutation detection in primary ovarian tumors and AUC=0.82 (n=12, 5.8%) for PIK3CA mutation detection in metastatic lesions. In primary soft tissue sarcoma, PIK3CA alterations were primarily PIK3CA amplification (n=8, 3%); however, the AI model showed an AUC of 0.82 for prediction of PIK3CA alterations in this setting. Compared to PIK3CA, PTEN loss had lower prevalence in our training and test sets in lung and ovarian cancers. Despite this, high performance by the AI model for PTEN loss prediction was seen in primary NSCLC (AUC=0.83, n=12, 1.5%) and ovarian cancers (AUC=0.9, n=8, 5%). In histologic settings with high prevalence of PTEN loss, such as endometrial cancer (n=192, 56%), high performance for PTEN loss prediction by the AI model was seen as well in primary and metastatic examples (AUC of 0.88 and 0.86, respectively). An AUC of 0.84 and 0.76 was achieved for detection of PTEN-mutant cases in primary colorectal and prostate cancers, respectively. Lastly, AKT1 (E17K) hotspot mutation was identified in primary colorectal cancers by the AI model with an AUC of 0.76, sensitivity of 0.89 and specificity of 0.37.

Erdafitinib was the first targeted therapy approved by FDA for the treatment of locally advanced or metastatic urothelial carcinoma with FGFR3 alterations [38]. Our AI model's prediction of FGFR3 hotspot mutations (R248C, S249C, G370C, Y373C) generated an AUC=0.88 (sensitivity=0.95 and specificity=0.48) in primary bladder cancer, slightly better than its performance in metastatic bladder cancers (AUC=0.79, sensitivity=0.73 and specificity=0.77).

Prediction of oncogenic mutation in mismatch repair (MMR) genes such as MLH1, MSH2, and MSH6 achieved a mean AUC of 0.85 (range 0.84-0.87) in primary endometrial cancer, a mean AUC of 0.94 (range 0.91-0.95) in primary colorectal cancer, and an AUC of 0.94 in primary bladder cancer, respectively. Other genes associated with “hypermutator” phenotypes, such as POLE mutations, could also be detected by the AI model. The AI model's prediction of POLE mutations in endometrial cancer achieved an AUC of 0.87, with a sensitivity of 0.86 and a specificity of 0.69.

Estimates of cost savings to enroll a 500-patient study, with all patients harboring a mutation in a particular gene by definitive NGS or polymerase chain reaction (PCR), after removing those cases unlikely to harbor a mutation in the gene based on pre-screening with our AI model (see Section 4.4), showed

substantial cost savings across all genes and cancer types investigated (Figure 7). Cost savings were highest for genes in which the targeted mutation type was lower in prevalence, due to reduced numbers of mutation negative patients sent on for definitive molecular screening. On average, a 14% and 22% cost saving could be achieved in PCR (range 3%-27%) and NGS (range 11%-34%) testing, respectively. This translates to an average cost reduction of approximately \$200K (range \$17K-670K) and \$800K (range \$140K-\$2,500K) for PCR and NGS testing.

2.4 Biomarkers associated with signaling pathways and genome instability

In addition to the genetic alterations in single genes, we evaluated the capabilities of the AI model to detect genetic alterations in any of a group of related genes canonically participating in a shared signaling pathway, hypothesizing that mutation in any member of the pathway may create an overlapping, shared phenotype due to the shared signaling cascade. In support of our hypothesis, the AI model predicted genomic alterations in any of the canonical tyrosine kinase receptor (RTK) MEK/ERK, mTOR, and TGF- β signaling pathways, as well as the HRD and DDR pathways (Figures 6, S5C, Table S3).

The MEK/ERK signaling pathway investigated comprised 34 genes including ALK, EGFR, ERBB2, FGFR1/2/3/4, RET and ROS1 (see Section 4) and alterations in any of the pathway members were identified with an AUC of 0.83 (sensitivity=0.88 and specificity=0.25) in primary thyroid cancer, and an AUC of 0.88 (sensitivity=1 and specificity=0.23) in metastatic thyroid cancer. Other canonical signaling pathways could be similarly detected. For example, in primary endometrial cancers, the AI model achieved a mean AUC of 0.80 (range 0.77-0.83) in predicting genomic alterations in canonical mTOR signaling (n=16 genes including AKT1/2/3, MTOR, PIK3CA, PTEN and TSC1/2, see Section 4), homologous recombination deficiency (HRD) (n=11 genes, including BRCA1/2, ATM and PALB2), the TGF- β canonical signaling pathway (SMAD2/3/4 and TGFB1/2), and the DNA damage response (DDR) (n=23 genes, including ATM, ATRX, BRCA1/2, MDM2/4 and PPP2R1A), with sensitivity and specificity in a range of 0.81-0.97 and 0.22-0.59, respectively. Phenotypes predictive of canonical signaling pathway alterations could be found in other histologies, and in both primary and metastatic settings as well. For example, predicting genetic alterations in any of the canonical TGF- β signaling pathway genes showed an AUC of 0.88 in primary hepatobiliary cancer, with sensitivity of 0.89 and specificity of 0.53, while prediction of genomic alterations in any of the canonical HRD associated genes achieved 0.81 AUC for metastatic soft tissue sarcoma.

Genomic instability is a hallmark of many cancers. We examined three measures of genomic instability: tumor mutation burden (TMB), microsatellite instability (MSI) and chromosomal instability (CIN), for model training and evaluation to see if phenotypes indicative of these measures of genomic instability could be detected.

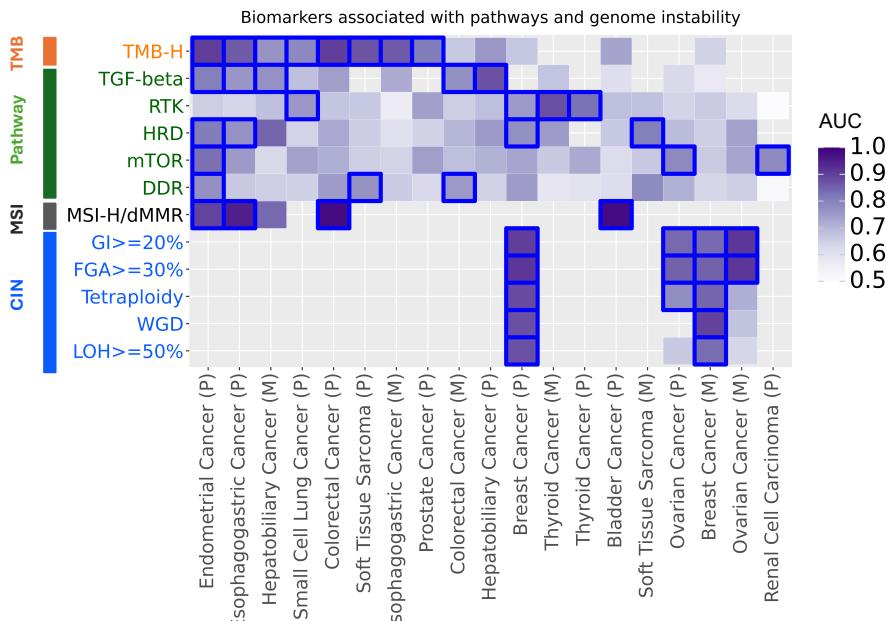


Figure 6: Heatmap showing the prediction of pathway and genome instability biomarkers, with $AUC > 0.5$ across cancer types at primary (P) or metastatic (M) lesions. 5 canonical signaling pathways include mTOR, Receptor tyrosine kinases (RTKs), TGF- β , homologous recombination deficiency (HRD) and DNA damage response (DDR) pathways. Genome instability include tumor mutation burden high (TMB-H), microsatellite instability high (MSI-H) or deficient mismatch repair (dMMR), and chromosomal instability (CIN) measures: fraction of genome altered (FGA) $\geq 30\%$, loss-of-heterozygosity (LOH) $\geq 50\%$, genome instability (GI) index ≥ 0.2 , tetraploidy and whole genome doubling (WGD). Blue boxes highlighted the biomarkers with positive sample ratio $> 2\%$, achieving an $AUC > 0.75$, sensitivity > 0.75 and specificity > 0.2 .

Prediction of tumor mutation burden-high (TMB-H) achieved a mean AUC of 0.85 (range 0.76-0.9) in 7 cancers, of which the prediction in primary endometrial and colorectal cancers showed the best performance with an AUC of 0.9 (sensitivity > 0.9 and specificity > 0.5), followed by esophagogastric cancer with an AUC of 0.86 in both primary and metastatic lesions. TMB-H prediction resulted in an AUC of 0.79 (sensitivity=0.89 and specificity=0.49) in primary NSCLC, where 150 samples harbored TMB-H and 624 had low TMB. TMB-H is a rare molecular subgroup in soft tissue sarcoma and prostate cancer. Though the majority of soft tissue sarcoma had low TMB, a small proportion of primary

samples harbouring TMB-H (n=8, out of 276, 2.9%), were identified, with sensitivity of 1 and specificity of 0.31 (AUC=0.87), suggesting potential clinical and therapeutic implications [39]. Similarly, prediction of TMB-H showed an AUC of 0.81 in primary prostate cancer (n=8, out of 326, 2.5%), with sensitivity of 0.88 and specificity of 0.5.

Microsatellite instability high (MSI-H) was found in 13.5% of primary colorectal cancer (n=84 MSI-H and n=538 MSS) and 18.6% (n=57 MSI-H and n=250 MSS) of primary endometrial cancer, from which prediction of MSI-H obtained an AUC of 0.98 (sensitivity=0.9 and specificity=0.95) and an AUC of 0.89 (sensitivity=0.89 and specificity=0.78) in primary colorectal and endometrial cancers , respectively. MSI-H is less frequent in bladder cancer (n=8 out of 291, 2.7%), and the model showed an AUC of 0.98, sensitivity of 1 and specificity of 0.91.

We also separately evaluated the performance of detecting dMMR defined as loss of IHC staining in MMR (MLH1, MSH2, MSH6 and PMS2) proteins and/or presence of genetic alterations in MMR genes. dMMR was found in 13.5% of primary colorectal cancers (n=36 out of 267), 18% of primary endometrial cancer (n=27 out of 149), and 17% of primary gastric cancers (n=12 out of 71). The model achieved an AUC of 0.997 (sensitivity=1 and specificity=0.93) for detection of dMMR among primary colorectal cancers, an AUC of 0.94 (sensitivity=0.96 and specificity=0.71) in primary endometrial cancers, and 0.999 (sensitivity=1 and specificity=0.98) in primary gastric cancers. In addition, individuals with Lynch syndrome (LS) are at increased hereditary risk of developing cancers with MSI-H/dMMR [40]. The diagnosis of LS is based on the detection of a germline pathogenic mutation in one of MMR genes or in EPCAM. Here, we derived a surrogate ground truth for Lynch Syndrome defined as the presence of germline mutation in any of MMR genes or EPCAM. Trained with this surrogate ground truth, our model showed a mean AUC of 0.87 (range 0.85-0.89) for prediction of LS in primary bladder, colorectal and endometrial cancers.

Chromosomal instability (CIN) was predicted by five metrics defined as presence of tetraploidy, whole genome doubling (WGD), fraction of genome altered (FGA) \geq 30%, loss-of-heterozygosity (LOH) \geq 50%, and genome instability index (GI index) \geq 0.2. CIN data was available for breast and high-grade serous ovarian cancer (HGSOC) only. When using tetraploidy as the ground truth for CIN, 178 samples were positive for tetraploidy, while 296 samples were negative for diploidy. From these, the model resulted in an AUC of 0.88 for CIN in primary breast cancers defined by identification of tetraploidy. WGD is highly associated with tetraploidy. Using WGD as the definition of CIN, 100 samples were positive for WGD while 284 samples were negative for lacking WGD, from which the model obtained an AUC of 0.88 CIN defined by WGD among primary breast cancers. Similarly, the model trained with FGA \geq 30% as ground truth resulted in an AUC of 0.91 from 313 samples positive for FGA \geq 30% and 192 samples with less than 30% of genome altered among primary breast cancers in evaluation, and LOH \geq 50% as ground truth resulted in an AUC of 0.87 in primary breast cancers where 27 samples had LOH \geq 50% and 478 samples were

negative, i.e. less than 50% of genome harboring LOH. GI index was derived by incorporating both FGA and LOH. When using GI index ≥ 0.2 as ground truth, the model obtained an AUC of 0.90 in primary breast cancers, of which 233 had GI index of ≥ 0.2 , and 272 samples had a GI index less than 0.2. Overall, the CIN measures in primary breast cancer achieved a mean AUC of 0.89 (range 0.87-0.91), with sensitivity > 0.9 and specificity > 0.52 , and the results in primary breast cancers were superior to AI model prediction in metastatic breast cancers overall for all definitions of CIN examined, with a mean AUC of 0.85 (range 0.83-0.9), sensitivity > 0.89 and specificity > 0.42 .

In HGSOC, the CIN measures showed a mean AUC of 0.78 (range 0.66-0.85) in primary cancers and a mean AUC of 0.77 (range 0.63-0.91) in metastatic cancers. In primary HGSOC, only three CIN measures, FGA $\geq 30\%$ (AUC=0.85, n=78 positive, and n=11 negative samples), GI index ≥ 0.2 (AUC=0.84, n=73 positive and n=16 negative samples) and tetraploidy (AUC=0.78, n=54 positive and n=29 negative samples) passed the baseline criteria of the model performance for AUC > 0.75 , showing sensitivity > 0.9 and specificity > 0.44 . In metastatic cancers, FGA $\geq 30\%$ and GI index ≥ 0.2 passed the baseline, showing an AUC of 0.91, sensitivity > 0.89 and specificity > 0.63 .

3 Discussion

A wide range of genomic abnormalities are documented in localized and advanced solid tumors via pan-cancer analysis, detailed in several publications and public data bases such as TCGA [41, 42, 43]. Genomic alterations in human cancers arise from mechanisms ranging from loss of heterozygosity, activating and inactivating point mutation, chromosomal loss or gain, gene amplification, insertions and/or deletions of small or large portions of genes, splice site alterations, to epigenetic mechanisms such as hypermethylation of promoter regions or the gene itself. Detection of genomic alterations plays an important role in diagnosis, therapy selection and response prediction. The mechanisms underlying carcinogenesis are not necessarily silenced by therapy either; point mutations and epigenetic alterations are common drivers of acquired tumor drug resistance. There has been extensive research on genomic biomarkers for prognosis and treatment prediction. The drive towards personalized medicine and delivery of targeted therapies requires robust biomarker assays to guide therapy selection in routine clinical care, and novel biomarker assays are often developed to guide inclusion in randomized clinical trials for investigational targeted therapies [44, 45]. Central to this paradigm is the ability to detect relevant genomic features efficiently and reliably. Specific genomic abnormalities may confer distinctive phenotypes, with particular biological characteristics and important clinical implications. The morphologic features of these phenotypes can be detectable by light microscopy by expert pathologists, such as the distinctively discohesive single file or single cell appearance, with intracytoplasmic lumens, of invasive lobular breast cancer that results from e-cadherin [CDH1] bi-allelic inactivation [46, 47]. However, morphological features of some

genomic alterations may be too subtle, even for expert pathologists, to be discovered and reported – or routinely identified. Machine learning on digitally scanned and rendered WSIs can leverage the recent and significant advantages of computer-assisted analysis of digital images to identify sometimes subtle, sometimes novel, features. Training to recognize relevant morphologic features for this kind of phenotype-genotype correlation modelling requires data from detailed molecular analyses to define specific alterations, in sufficient volume of cases, and in cases where scanned WSIs are available. Performing such training to generate, via machine learning, a new AI model that can rapidly and at scale in slide volume, scale in mutation targets, and scale in tumor histologies trained is possible and could lead to digital biomarkers to screen for clinically relevant genomic alterations in a way that saves time, tissue and costs for researchers and clinicians.

We demonstrate an effectively performing artificial intelligence (AI)-based model utilizing ground truth histological and molecular data from MSKCC and MSK-IMPACT. Our approach was to train a multi-label classifier for the prediction of 1,228 most clinically relevant genomic abnormalities in 505 genes from H&E WSIs in 70 human cancers. Our evaluation of the model focused on the 15 most common broad cancer types treated at MSKCC, and identified the most frequent mutations across histologies. We validated the performance of the model internally with a held out validation set of MSKCC cases, as well as externally with cases from the publicly available TCGA dataset. The performance of the model in TCGA validation dataset was slightly inferior to that observed in the internal MSK validation. Slight divergence in validation results on TCGA data and images versus other validation sets is not uncommon in pathology image analysis development, for a variety of reasons. The main advantage of the TCGA dataset for these applications is that the data is external, public, and multi-institution, making it convenient to evaluate digital pathology image applications; however, TCGA was not developed with image analysis in mind. The primary purpose was to create an atlas of genomic alterations across cancers, and the digital image submission requirement was a fortunate afterthought. This meant permitting a “representative image” in TCGA which can vary dramatically in quality, coupled with variation in the sequencing method and analysis for results appended to the case. The internal dataset was on withheld clinical grade material and WSIs, with prior quality control on both the clinical staining done on the slide as well as on the resulting scanned image, and single institution sequencing results using an FDA-approved NGS method. Thus, the differences between the validation sets in terms of purpose, quality control, and ground truth development most likely explain the minor divergence in validation testing results.

Internal and external validation, despite some of the limitations, did confirm high performance in many aspects for the AI model we developed, with clinical and future research implications. For example, the prediction of MSI-H/dMMR associated genes showed strong performance in colorectal and endometrial cancers, as well as bladder and gastric cancers. Tumors with MSI-H/dMMR often harboured high TMB and performance detecting TMB-H was best in colorectal cancer.

tal and endometrial cancers. Even in tumor histologies with low prevalence of TMB-H, such as soft tissue sarcoma and prostate cancer [48], our AI model was able to identify the rare tumors in low prevalence histologies with TMB-H, suggesting the potential to identify patients who might respond to immune checkpoint inhibitor (ICI) therapy based on TMB in these less frequently screened histologies. Moreover, our AI model demonstrated strong performance in genomic biomarkers already routinely screened clinically for therapy selection and response prediction, such as targeted therapy associated genes (EGFR, KRAS, MET, ALK, and ROS1) in NSCLC, ERBB2 amplification in breast and gastric cancers for HER2-targeted therapy, FGFR3 genomic alterations to target FGFR-altered urothelial carcinoma, and BRAF oncogenic mutation targeting BRAF V600E mutation carriers in metastatic thyroid cancer, melanoma and colorectal cancers, suggesting an AI assisted digital biomarker on the H&E stained WSI could be developed as a cost, time, and tissue sparing triage for definitive molecular testing for these targets. While some of the other genes investigated do not show a high AUC when set for a high sensitivity > 0.9 screening assay to identify targeted mutation carriers, our results for genomic alteration prediction in genes like ALK/ROS1 in NSCLC and ERBB2 alterations in NSCLC did show a high NPV of 0.99. This suggests that even these gene targets could be further developed as screening digital biomarkers to triage downstream definitive testing, functioning essentially as a highly accurate “rule out” assay, based on the NPV and thus ability to identify cases that almost certainly do not harbor one of these targeted mutations, and thus may not need definitive confirmatory testing.

As proof of this concept, we presented example projected cost savings using the performance characteristics results with this AI model to identify BRAF mutations in melanoma and colorectal carcinoma, KRAS mutations in NSCLC and colorectal carcinoma, EGFR and MET mutations in NSCLC, and FGFR3 mutations in bladder cancer (Section 4.4). We demonstrated the use of the AI model as a digital biomarker to triage downstream NGS and PCR testing in these settings could potentially save an average of \$200K (3%-27% cost reduction) for PCR testing and \$800K (11%-34% cost reduction) for NGS testing, assuming a planned enrollment of at least 500 patients with the targeted mutation in these settings, with published prevalence estimates (Figure 7). Cost savings were most pronounced for low prevalence targets, and show that triage with AI models for digital biomarkers may be sufficient to remove cost barriers to clinical studies seeking to use targeted therapies in settings where the targeted mutation is low prevalence. Use cases for digital screening and cost savings can be extrapolated for other targets we investigated, and beyond clinical trial and cost savings to cost reduction for routine clinical testing in particular laboratory and practice settings, when used to triage for more expensive definitive tests.

In further support of potential clinical trial applications of our work, our model detected many genomic alteration targets in multiple histologies. For example, phenotypes predictive of genomic alterations in PIK3CA and PTEN were identified in several tumor histologies. As of the time of publication, there are several active clinical trial programs for drugs targeting these genes, including

Estimated cost savings for AI model pre-screening for definitive PCR or NGS identification to enroll 500 patients with a mutation in a target gene and histology

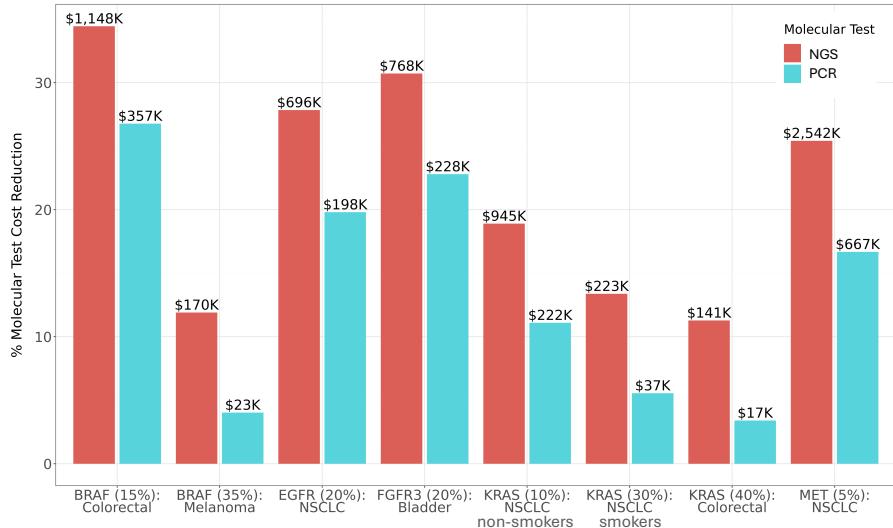


Figure 7: Estimated cost saving for PCR or NGS definitive screening test, using AI model pre-screening, to enroll 500 patients with a mutation in a targeted gene and histology. Y-axis showed the estimated cost reduction (%) for molecular testing with AI model pre-screening. The estimated amount (\$) of cost reduction is annotated on top of each bar. X-axis indicated targeted genes that are commonly assessed by NGS or PCR clinically, and their published prevalence of gene alterations (%) in each cancers: BRAF in colorectal cancer and melanoma, EGFR in non-small cell lung cancer (NSCLC), FGFR3 in bladder cancer, KRAS (a low estimate prevalence 10% for a population with little to no smoking, i.e., non-smokers, and a higher estimate prevalence 30% for smokers) in NSCLC and colorectal cancer, and MET in NSCLC.

pan-solid tumor trials or cohorts. Thus, the ability of our AI model to identify genomic alteration signals in multiple histologies suggests it may have use in research applications, ranging from cost-effective screening of tissues and tissue blocks to find uncommon or rare genomic alterations within histologies (i.e., TMB-high cases of prostate cancer), identification of phenotypes predicting the same forms of genomic alterations across several histologies for hypothesis generation, including models predicting shared phenotypes resulting from activating or inactivating mutations in members of canonical signaling pathways, as we demonstrated in our results. Screening for genomic alterations across histolo-

gies may also be utilized in clinical trial settings to rapidly and cost-effectively identify patients who could be candidates for a clinical study, before investing significant time, tissue, and resources into definitive molecular testing. This may not only speed up drug development, but better match patients to clinical trials which are appropriate for their particular tumor.

We would expect genomic alterations with known strong phenotype correlations in subtypes of certain histologies to be biased towards these subtypes in our results. As expected, MDM2 amplification detection in soft tissue sarcoma with AUC of 0.84 when all sarcomas were tested on subtype analysis showed primarily recognition of the distinctive liposarcoma phenotype [49] with AUC of 0.93. Similarly, strong results for VHL when all renal cell carcinomas were evaluated were biased to detection of clear cell renal cell carcinoma, a distinct histologic subtype that is known to be driven by VHL loss [50]. The top performing gene prediction in all gliomas was IDH1 (AUC of 0.93), which was driven by detection of oligodendrogloma, which is genetically defined by IDH mutation [27]. This demonstrates that our model and approach, when a strong correlation of phenotype with specific genomic alterations is expected in particular diagnoses and subtypes, finds the expected phenotype-genotype correlation, and our unsupervised model is training itself to known phenotypes in settings with known phenotype-genotype correlations. Identification of these expected biases in subtype diagnosis detection in settings where phenotype and genotype were already known to be correlated, across histologies we tested, strongly suggests that the model is not being spurious in identifying high confidence prediction of genomic alterations, even in genes where previous phenotypes have not been identified and described by human pathologists. Thus, coupled with high confidence detection of genomic alterations in settings where high prevalence actionable targets are also previously described, the finding of expected diagnostic phenotype-genotype correlations shows the method and model here is robust across genes and histologies.

Our AI model may thus be used to support pathologists and researchers to accurately predict gene mutations based on subtle morphological features reflecting these genomic alterations, and facilitate the diagnosis of histologies harboring disease-defining genetic alterations, or drive hypothesis about the novel phenotypic-genotypic correlation our model is discovering. Clinical application of such digital assays requires high performance, however, which could be further improved through re-training with larger patient numbers in focused histologic and clinical settings [51]. This could potentially eliminate the need to examine additional, precious tumor tissue for further analyses such as IHC, Fluorescence In Situ Hybridization (FISH) or NGS testing to assess genomic alterations, thereby saving time, costs and tissue. Moreover, identifying targets important for clinical trials could facilitate the rapid development of digital biomarkers for tissue-preserving screening, enabling quicker identification of patients eligible for enrollment and further optimizing resources in clinical trial screening.

Additionally, digital biomarkers could be used to pre-select cases most likely to harbor specific genomic alterations, justifying the costs associated with defin-

tive genetic analysis. Such digital biomarker molecular predictions may enable future fusion of clinical variables such as histologic grade/stage, ER/PR status in breast cancers, Gleason score in prostate cancers, prior/post treatment status with predicted genomic abnormalities for multi-modal refinement of treatment algorithms with future research.

The limitations of the current study include the following factors: data and analyses have been restricted to WSIs taken from an MSK-IMPACT cohort which included a small number of germline variants. Therefore, we have yet to determine if the observations in this study will generalize to other sequencing assays. For instance, other assays do not employ a germline control to whether variants are germline or somatic, which can affect the interpretation of the sequencing results.

The performance of our model may vary when applied to publicly available datasets such as TCGA, due to the fact that the sequencing for these cohorts was performed in a research setting with research approaches for library preparations and was not performed to a clinically meaningful sequencing depth.

Finally, some of the genomic alterations we studied were present in relatively small numbers of cases in our dataset, due to limited prevalence in the cases available to us. Adding additional cases with these alterations may reveal a stronger signal than was found in our study. As shown in Table S5 and Figure 2, our development cohort is enriched in primary samples compared to metastatic samples, which may explain why better performance was typically observed in primary samples vs metastatic samples across the histologies we tested. Collecting more samples from metastatic lesions may uncover more biomarkers linked to metastasis, and/or stronger performance of genomic alteration predictions in metastatic lesions.

In summary, we describe a fully automated method for rapid high-throughput AI-assisted screening of cancers not only to detect clinically meaningful genomic abnormalities across histologies, but also to identify the histologies characterized by genomic alterations, capable of directing selection of patients for definitive genomic analysis and/or clinical decisions. This pipeline is cost and time efficient, tissue sparing and capable of application to large patient populations with reliable performance metrics. This approach is suitable for application in research, including clinical trial settings, across different types of cancer.

4 Methods

4.1 Patient cohorts, histopathologic and genomic analyses

4.1.1 Development cohort: MSK IMPACT Dataset

A sample is defined as an MSK IMPACT assay paired with one or more H&E stained WSIs taken from the same formalin-fixed paraffin-embedded (FFPE) tissue block. The MSK IMPACT dataset consists of 43,605 samples (47,960 WSIs) from 71 cancer types (70 cancer types and a “Unknown” category for which the cancer type is not known). 28,351 (65%) of samples are primary cancer

sites while the remaining 15,254 (35%) are metastasis samples. All image data were retrieved from the hospital archives and verified to meet staining quality standards for histopathology review. All biopsy glass slides were scanned with Leica Aperio AT2 scanners (Leica Biosystems, Division of Leica Microsystems Inc, Buffalo Grove, IL, USA) at 20 \times (0.5 microns per pixel) or 40 \times (0.25 microns per pixel) magnification. See Table S5 for a detailed breakdown of the dataset.

The cohort underwent paired tumor-normal targeted sequencing using the MSK-IMPACT assay. MSK in-house bioinformatic pipelines were employed to analyze the sequencing results. The analysis determined the genomic alteration status, including mutations (single-nucleotide variants (SNVs) and insertion/deletions (indels)), copy number variations (amplification and deletions), and structural rearrangements (fusions) in 505 important cancer associated genes. The list of cancer associated genes were compiled from four IMPACT versions including v3, v5, v6 and v7 that respectively encompass 341, 410, 468 and 505 genes. Additional fusion events confirmed by MSK-Fusion panel via RNA sequencing were incorporated to improve the coverage of fusion detection for the genes that are available from MSK-IMPACT panel. The annotation of the oncogenicity and clinical implication of specific genetic alterations were determined using OncoKB [30, 31]. tumor mutation burden (TMB) score, microsatellite instability (MSI) score, LOH, WGD where available were analyzed and retrieved at Memorial Sloan Kettering (MSK). Additionally, dMMR status was confirmed by IHC assay.

Among MSK-IMPACT panel genes, five groups of genes that are canonically participated in shared pathways associated with DNA repair mechanisms were identified. The gene members of each pathways were retrieved from MCG [28, 29]. DNA damage response (DDR) is composed of 23 genes: ATM, ATR, ATRX, BRCA1, BRCA2, BRIP1, FANCA, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCL, MDM2, MDM4, MLH1, MUTYH, NPM1, PALB2, PPP2R1A, RAD50, RAD51, STAG2; receptor tyrosine kinase (RTK) pathway of 34 genes ALK, CBL, CSF1R, DDR2, EGFR, EPHA3, EPHA5, EPHB1, ERBB2, ERBB3, ERBB4, FGF19, FGF3, FGF4, FGFR1, FGFR2, FGFR3, FGFR4, FLT1, FLT3, FLT4, HGF, IGF1R, KIT, MET, NF1, NTRK1, NTRK2, NTRK3, PDGFRA, PDGFRB, PTPN11, RET, ROS1; homologous recombination deficiency (HRD) pathway of 11 genes: BRCA1, BRCA2, PALB2, ATM, CHEK1, CHEK2, RAD51, FANCA, CDK12, RAD51B, RAD51C; mTOR signaling pathway of 16 genes: AKT1, AKT2, AKT3, CRKL, IRS2, MTOR, PIK3CA, PIK3CG, PIK3R1, PIK3R2, PTEN, RICTOR, RNF43, RPTOR, TSC1, TSC2; and TGF- β signaling pathway of 5 genes: SMAD2, SMAD3, SMAD4, TGFBR1, TGFBR2.

4.1.2 Validation cohort: TCGA PanCancer Atlas

The external validation dataset consists of samples from TCGA PanCancer Atlas project [20]. A total of 11,406 diagnostic WSIs, corresponding to 32 cancer types, were retrieved together with the genomic alteration data, including mutation, copy number aberration, and fusion, from TCGA and the cBioPortal (<http://www.cbiopal.org>), respectively. 36 slides were excluded from the

validation cohort due to missing microns-per-pixel (mpp) information and a further 3 slides were excluded due to being out of focus and no foreground tiles being detected. The remaining 9,340 samples (11,367 WSIs) define the validation cohort (referred to as TCGA cohort, Table S6).

4.2 Development of an AI-based system for the detection of digital biomarkers in pan-cancer cancer using whole slide images

We developed a pan-cancer digital biomarker screening model to predict genomic abnormalities of interest in human cancers from H&E WSIs. The model was trained on 33,564 diagnostic clinical WSIs from a cohort of 27,290 patients treated at MSK. The training cohort covers 70 different cancer types with the 505 genes assessed by the MSK-IMPACT targeted sequencing oncology assay (Figures 2, S1, Table 1). The training ground truth labels include oncogenic point mutations, copy number variations (amplifications or deletions) and fusion events, or the presence of these types of genetic variations in any of a group of genes canonically participating in a shared signaling pathway associated with cancer, e.g., DNA damage responses, RTK pathway, and mTOR signaling pathway (Figure S2).

The task is framed as a multi-label binary classification task, where genomic features (biomarkers) are represented as binary labels. Each binary label indicates the presence or absence of genomic alterations in a single gene, or in any of a group genes that participate in a shared signaling pathway. The genomic feature binary labels derived from MSK-IMPACT results are oncogenic mutations, copy number amplification, copy number deletions, fusions, or the combination of oncogenic mutation/amplification if a gene is an oncogene and oncogenic mutation/deletion if a gene is a tumor suppressor gene (TSG). Additional genomic features included in training are TMB-H for TMB ≥ 10 mutations/megabase (mut/Mb), MSI-H for MSI score ≥ 10 , and microsatellite stable (MSS) for MSI score < 3 , dMMR for loss of IHC staining in MMR (MLH1, MSH2, MSH6, and PMS2) proteins and harbored genetic alterations in MMR genes, Lynch Syndrome for the presence of germline mutation in any of MLH1, MSH2, MSH6, PMS2, EPCAM, and CIN defined as presence of tetraploidy, whole genome doubling (WGD), loss-of-heterozygosity (LOH) in $\geq 50\%$ genome, fraction of genome altered (FGA) $\geq 30\%$, and genome instability index (GI index) ≥ 0.2 . The GI index is a metric derived from FGA and LOH ranging from 0 to 1.

The MSK development cohort was split into train, tune and test datasets with a ratio of 7:1:2. This partitioning resulted in sample sizes of 30,511 (33,564 WSIs) for the train set, 4,334 (4,762 WSIs) for the tune set, and 8,760 (9,634 WSIs) for the test set. The distribution of cancer types across the datasets included 70 different types in the train set, 58 in the tune set, and 62 in the test set (Figure 2, Table S5). Each dataset includes the category of an “Unknown” histology. A total of 1,228 biomarker labels with at least 8 positive and 8 negative samples in the train set and at least 4 positive and 4 negative samples

in the tune set were included in the training.

Each slide is split into image tiles of size 224×224 pixels. The tiles are filtered to only include those representing foreground (tissue) using a foreground detection model based on a Fully Convolutional Network (FCN). Each foreground tile is embedded with Virchow2 [17, 18] into a tile embedding of length 2560. Each slide is thus represented as a $N \times 2560$ tensor, where N is the number of foreground tiles.

The embeddings serve as input into a feed-forward network with an attention mechanism that aggregates the tile-level embeddings into a slide-level prediction [17]. See section 8.1 for an overview of model and training parameters. The model was trained on slide level. The final prediction per sample was determined as the maximum prediction over slides in the sample.

Checkpoint selection was done using the mean AUC and mean AP across all labels on the tune set. The operating threshold for each label was determined on the tune set by optimizing for 90% sensitivity in each cancer type present in the tune set. A threshold thus corresponds to a label and cancer type pair. These thresholds were then used to generate sample-level binary predictions from the inference probabilities in the MSK test set and the TCGA validation set, indicating the presence or absence of the genetic mutation. From 62 cancer types present in the MSK test set, 56 are also present in the tune set, thus we report results on the MSK test set for these 56 cancer types.

4.3 Phenotype-genotype correlation analysis

For each histology subtype in a cancer, we used the Kolmogorov-Smirnov (KS) one-side test to examine whether the inference probabilities of a biomarker label in the target subtype is greater than the inference probabilities in the other subtypes of cancer. The p-values were adjusted using the Benjamini-Hochberg method. P-values < 0.01 were considered as statistically significant. A histologic subtype ground truth is defined as a binary label indicating if the sample was annotated for a given histologic subtype. For a biomarker with significantly higher inference probabilities in a specific histologic subtype of cancer, the AUC of the biomarker in prediction of the corresponding subtype was evaluated by using inference probabilities as prediction and the binary label of histologic subtype as ground truth. The sensitivity and specificity was computed by using the binary prediction of the genomic alterations as prediction label and the binary label of histologic subtype as ground truth.

4.4 Cost saving analysis

To evaluate the cost-saving potential of using an AI model to triage for downstream molecular testing by excluding patients from downstream definitive NGS or PCR who are unlikely to harbor genetic mutations, we first assumed a typical Phase 3 study seeking to enroll 500 patients with a mutation in a target gene in a particular type of cancer. We selected BRAF (colorectal cancer and melanoma), EGFR (NSCLC), FGFR3 (bladder cancer), KRAS (NSCLC and

colorectal cancer), and MET (NSCLC), as these are commonly assessed by NGS or PCR clinically, allowing good estimates of the cost of these molecular assays along with a nominal cost of AI model screening. We assumed prevalence of these mutations in these forms of cancer at published rates, with KRAS in NSCLC given a low estimate for a population with little to no smoking and a higher estimate for a population with high smoking prevalence. From this, we calculated the estimated cost savings due to reduced numbers of definitive NGS or PCR tests needed after our AI model identified cases unlikely to harbor the targeted mutation based on the performance of our AI model for these genes. Our calculations were done according to the following method:

Assuming enrollment of a target number of patients all with tumors harboring a specific mutation, cost estimation for patient screening was calculated by 1) estimating cost to enroll the target number of patients using a conventional molecular test (NGS or PCR) only, and 2) NGS or PCR following pre-screening via an AI model to eliminate patients who were not likely to have the targeted mutation, hence would not benefit from the targeted therapy or respond to targeted drugs. The cost savings was then calculated as follows:

- Number of targeted enrolled patients: N_{target}
 - Prevalence of a genomic alteration in a cancer: $prevalence$
 - Sensitivity, i.e., the AI model algorithm's true positive rate: $sensitivity$
 - Specificity, i.e., the AI model algorithm's true negative rate: $specificity$
 - Cost of AI screening per patient: C_{AI}
 - Cost of molecular testing per patient: $C_{testing}$
1. Cost estimation for patient screening using a conventional molecular testing:
Number of patients to be screened by molecular testing without AI model:

$$N_{conventional} = \frac{N_{target}}{prevalence}$$

Total cost of patients to be screened by molecular testing only without engaging an AI model:

$$C_{conventional} = C_{testing} \times N_{conventional}$$

2. Cost estimation for patient screening using a molecular testing with an AI model for pre-screening:

Number of patients to be screened with an AI model:

$$N_{screened} = \frac{N_{target}}{prevalence \times sensitivity}$$

Number of true positives (TP):

$$TP = N_{screened} \times prevalence \times sensitivity$$

Number of false positives (FP):

$$FP = N_{screened} \times (1 - prevalence) \times (1 - specificity)$$

Number of patients sent for molecular testing:

$$N_{sent} = TP + FP$$

Cost of AI model for screening all patients:

$$C_{total(AI)} = C_{AI} \times N_{screened}$$

Cost of molecular testing for patients sent for testing:

$$C_{total(testing)} = C_{testing} \times N_{sent}$$

Total cost of patients' molecular testing with AI screening:

$$C_{with_AI} = C_{total(AI)} + C_{total(testing)}$$

Total cost saving by using an AI model for pre-screening:

$$C_{saving} = C_{conventional} - C_{with_AI}$$

Hence, the percentage of cost reduction by using an AI model for pre-screening:

$$C\% = \frac{C_{saving}}{C_{conventional}}$$

5 Acknowledgement

We would like to extend our sincere gratitude to several individuals who contributed significantly to the success of this project. Philippe Mathieu, Alexander van Eck, Wayne Hendricks, Sid Senthilnathan, Michael Singer and Eric Robert played a crucial role in developing the AI and data infrastructure, which enabled the large-scale training and evaluation essential to this work.

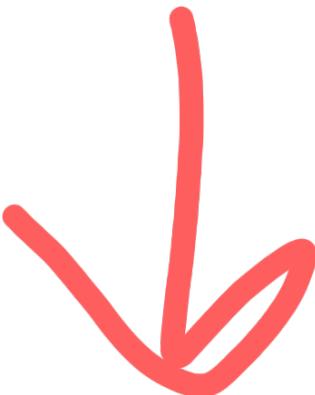
We also wish to acknowledge Jan Bernhard for his earlier contributions to the development of digital biomarker projects related to MSI detection, which laid the foundational work for this project.

Our thanks also go to Eugene Vorontsov, Julian Viret, Alican Bozkurt, Adam Casson, George Shaikovski, and Michal Zelechowski for their invaluable support in providing the Paige ML SDK and guidance on utilizing Virchow2 model embeddings.

Finally, we are deeply grateful to Christopher Kanan and Brandon Rothrock for their supervision and mentorship during the early stages of the biomarker related research, which significantly influenced the direction of this research.

6 Data protection

This project was governed by an Institutional Review Board approved retrospective research protocol, under which consent/authorization was waived before research was carried out. Data collection was conducted exclusively at MSK. The AI model was developed by Paige.



7 Tables

Table 1: Development and validation datasets - sample and slide counts

Dataset	Train		Tune		Test	
	Samples	WSIs	Samples	WSIs	Samples	WSIs
MSK	30,511	33,564	4,334	4,762	8,760	9,634
TCGA	-	-	-	-	9,340	11,367

Table 2: Genetic alteration biomarkers validated in both MSK and TCGA cohorts

Biomarker (MSK) - Predicted label	Cancer type	MSK			TCGA		
		AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
FGFR3 oncogenic mutation	Bladder Cancer	0.88	0.92	0.60	0.87	0.76	0.81
FGFR3 amplification/mutation	Bladder Cancer	0.87	0.95	0.47	0.87	0.82	0.73
CDH1 oncogenic mutation	Breast Cancer	0.94	0.89	0.84	0.94	0.95	0.82
CDH1 deletion/mutation	Breast Cancer	0.94	0.88	0.83	0.94	0.96	0.83
CDK12 amplification	Breast Cancer	0.89	1.00	0.48	0.78	0.87	0.41
ERBB2 amplification	Breast Cancer	0.86	0.96	0.48	0.81	0.95	0.39
TP53 oncogenic mutation	Breast Cancer	0.91	0.91	0.66	0.88	0.92	0.56
TP53 deletion/mutation	Breast Cancer	0.91	0.90	0.67	0.88	0.92	0.55
BRAF oncogenic mutation	Colorectal Cancer	0.85	0.91	0.50	0.84	0.93	0.43
BRAF amplification/mutation	Colorectal Cancer	0.85	0.91	0.48	0.84	0.98	0.23
KMT2D oncogenic mutation	Colorectal Cancer	0.96	1.00	0.67	0.92	1.00	0.48
KMT2D deletion/mutation	Colorectal Cancer	0.96	1.00	0.65	0.91	1.00	0.57
RNF43 oncogenic mutation	Colorectal Cancer	0.90	0.94	0.62	0.90	1.00	0.40
RNF43 deletion/mutation	Colorectal Cancer	0.90	0.89	0.71	0.89	0.94	0.49
CTNNB1 oncogenic mutation	Endometrial Cancer	0.88	0.94	0.47	0.77	0.84	0.51
CTNNB1 amplification/mutation	Endometrial Cancer	0.86	0.94	0.45	0.77	0.78	0.54
PTEN oncogenic mutation	Endometrial Cancer	0.88	0.94	0.55	0.87	0.79	0.77
PTEN deletion/mutation	Endometrial Cancer	0.86	0.94	0.52	0.87	0.81	0.79
TP53 oncogenic mutation	Endometrial Cancer	0.91	0.89	0.80	0.85	0.93	0.48
TP53 deletion/mutation	Endometrial Cancer	0.91	0.89	0.78	0.85	0.93	0.49
KMT2B oncogenic mutation	Esophagogastric Cancer	0.85	0.94	0.41	0.92	1.00	0.65
KMT2B deletion/mutation	Esophagogastric Cancer	0.86	0.94	0.44	0.92	1.00	0.68
KMT2D oncogenic mutation	Esophagogastric Cancer	0.89	1.00	0.62	0.91	0.95	0.71
KMT2D deletion/mutation	Esophagogastric Cancer	0.89	0.94	0.67	0.92	0.91	0.75
IDH1 oncogenic mutation	Glioma	0.93	0.83	0.89	0.80	0.87	0.55
IDH1 amplification/mutation	Glioma	0.93	0.84	0.88	0.81	0.91	0.54
ATRX deletion/mutation	Glioma	0.85	0.94	0.37	0.88	1.00	0.20
EGFR oncogenic mutation	Non-Small Cell Lung Cancer	0.87	0.93	0.52	0.84	0.92	0.59
EGFR amplification/mutation	Non-Small Cell Lung Cancer	0.85	0.93	0.46	0.83	0.92	0.50
STK11 oncogenic mutation	Non-Small Cell Lung Cancer	0.87	0.83	0.76	0.86	0.88	0.69
BRAF oncogenic mutation	Thyroid Cancer	0.89	0.83	0.77	0.90	0.96	0.61
BRAF amplification/mutation	Thyroid Cancer	0.89	0.83	0.80	0.90	0.96	0.60
NRAS amplification/mutation	Thyroid Cancer	0.87	1.00	0.72	0.86	0.77	0.81

Table 3: Genetic alteration biomarkers associated with histologic subtypes

Cancer Type	Histology	Biomarker	KS-test adj.P-value	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Breast Cancer	Breast Invasive Lobular Carcinoma	CDH1 deletion/mutation	2.28e-07	0.93	0.63	0.92	0.78	0.31	0.99	113	104	1127
Endometrial Cancer	Uterine Endometrioid Carcinoma	PTEN oncogenic mutation	4.8e-41	0.86	0.85	0.91	0.67	0.73	0.87	225	204	429
Endometrial Cancer	Uterine Serous Carcinoma/Uterine Papillary Serous Carcinoma	TP53 oncogenic mutation	1.23e-15	0.86	0.52	0.97	0.57	0.27	0.99	61	368	429
Esophageal Cancer	Esophageal Squamous Cell Carcinoma	FAT oncogenic mutation	1.28e-08	0.88	0.53	0.73	0.62	0.22	0.98	15	27	267
Glioma	Anaplastic Astrocytoma	ATRX oncogenic mutation	2.38e-07	0.88	0.46	0.95	0.41	0.22	0.99	22	320	342
Glioma	Anaplastic Oligodendrogloma	IDH1 oncogenic mutation	1.02e-06	0.86	0.28	0.77	0.73	0.17	0.98	22	320	342
Glioma	Anaplastic Oligodendrogloma	CIC deletion/mutation	1.98e-08	0.96	0.35	0.91	0.90	0.24	1.00	11	331	342
Glioma	Anaplastic Oligodendrogloma	FUBP1 deletion/mutation	3.88e-04	0.94	0.32	0.92	0.92	0.28	0.98	11	331	342
Glioma	Anaplastic Oligodendrogloma	NOTCH1 amplification/mutation	5.98e-05	0.91	0.38	1.00	0.48	0.06	1.00	11	331	342
Glioma	Anaplastic Oligodendrogloma	IDH1 amplification/mutation	1.33e-03	0.86	0.23	0.82	0.71	0.09	0.99	11	331	342
Glioma	Oligodendrogloma	FUBP1 deletion/mutation	9.74e-09	0.93	0.57	0.71	0.97	0.55	0.98	17	325	342
Glioma	Oligodendrogloma	CIC deletion/mutation	3.28e-03	0.93	0.27	0.76	0.87	0.09	1.00	17	325	342
Glioma	Oligodendrogloma	IDH1 oncogenic mutation	3.84e-07	0.89	0.27	0.94	0.73	0.15	1.00	17	325	342
Glioma	Oligodendrogloma	CTNNB1 oncogenic mutation	8.35e-15	0.98	0.81	0.56	0.95	0.92	0.91	39	175	214
Hepatobiliary Cancer	Hepatocellular Carcinoma	TERT oncogenic mutation	2.92e-16	0.89	0.77	0.72	0.95	0.76	0.94	39	175	214
Hepatobiliary Cancer	Hepatocellular Carcinoma	USP9X amplification/mutation	1.07e-10	0.94	0.65	0.75	0.98	0.08	1.00	20	305	325
Melanoma	Uveal Melanoma	SF3B1 amplification/mutation	3.09e-10	0.94	0.53	0.75	0.93	0.42	0.98	20	305	325
Melanoma	Uveal Melanoma	GNA11 oncogenic mutation	1.09e-08	0.95	0.65	0.70	0.95	0.47	0.98	20	305	325
Melanoma	Uveal Melanoma	RECQL4 amplification	6.98e-06	0.86	0.47	0.55	0.92	0.31	0.97	20	305	325
NSCLC	Lung Squamous Cell Carcinoma	DNMT3B amplification	1.28e-09	0.91	0.47	0.52	0.95	0.35	0.95	12	1007	1119
NSCLC	Lung Squamous Cell Carcinoma	KMT2D deletion/mutation	6.82e-46	0.90	0.52	0.96	0.54	0.19	0.99	12	1007	1119
NSCLC	Lung Squamous Cell Carcinoma	PIK3CA amplification	6.09e-38	0.89	0.61	0.63	0.95	0.50	0.96	12	1007	1119
NSCLC	Lung Squamous Cell Carcinoma	PRKCI amplification	2.35e-30	0.86	0.53	0.91	0.69	0.20	0.98	12	1007	1119
Ovary Cancer	Clear Cell Ovarian Cancer	ARID1A deletion/mutation	3.58e-10	0.93	0.59	0.92	0.97	0.29	0.99	23	327	371
Ovarian Cancer	Endometrioid Ovarian Cancer	PIK3R1 deletion/mutation	1.10e-09	0.96	0.63	1.00	0.66	0.11	1.00	15	356	371
Ovarian Cancer	Endometrioid Ovarian Cancer	ARID1A deletion/mutation	3.68e-07	0.95	0.39	0.93	0.76	0.14	1.00	15	356	371
Ovarian Cancer	Endometrioid Ovarian Cancer	PIK3CA oncogenic mutation	2.22e-05	0.88	0.28	1.00	0.58	0.06	1.00	15	356	371
Ovarian Cancer	Mucinous Ovarian Cancer	KRAS oncogenic mutation	9.41e-04	0.96	0.48	0.97	0.69	0.09	1.00	15	356	371
Ovarian Cancer	Ovarian Carcinosarcoma/Malignant Mixed Mesodermal Tumor	KRAS oncogenic mutation	8.91e-04	0.90	0.25	0.88	0.59	0.04	0.99	8	363	371
Ovarian Cancer	Ovarian Carcinosarcoma/Malignant Mixed Mesodermal Tumor	CDKN1B amplification	1.35e-05	0.95	0.52	1.00	0.37	0.04	1.00	9	362	371
Ovarian Cancer	Pancreatic Adenocarcinoma	ETV6 amplification	3.37e-06	0.94	0.50	0.72	0.27	0.03	1.00	9	362	371
Pancreatic Cancer	Pancreatic Adenocarcinoma	TP53 oncogenic mutation	6.82e-31	0.89	0.96	0.84	0.83	0.95	0.56	364	86	450
Pancreatic Cancer	Pancreatic Adenocarcinoma	MEN1 deletion/mutation	1.41e-46	0.95	0.93	0.79	0.57	0.87	0.97	52	398	450
Pancreatic Cancer	Pancreatic Neuroendocrine Tumor	DAXX deletion/mutation	5.03e-40	0.88	0.90	0.56	1.00	0.97	0.94	52	398	450
Renal Cell Carcinoma	Chromophobe Renal Carcinoma	TP53 oncogenic mutation	5.12e-05	0.94	0.44	1.00	0.50	0.09	1.00	7	187	194
Renal Cell Carcinoma	Renal Angiomyolipoma	TSC2 oncogenic mutation	5.2e-04	0.94	0.31	1.00	0.73	0.09	1.00	5	189	194
Renal Cell Carcinoma	Renal Clear Cell Carcinoma	VHL oncogenic mutation	3.01e-29	0.93	0.94	0.95	0.83	0.88	0.92	112	82	194
Soft Tissue Sarcoma	Desmoplastic Small-Round-Cell Tumor	WT1 fusion	3.26e-22	1.00	0.97	0.76	1.00	1.00	0.99	17	38	365
Soft Tissue Sarcoma	Leiomyosarcoma	FLCN amplification	1.51e-12	0.89	0.41	0.90	0.72	0.22	0.99	30	335	365
Soft Tissue Sarcoma	Leiomyosarcoma	MAP2K4 amplification	1.20e-11	0.88	0.65	0.50	0.58	0.71	0.96	30	335	365
Soft Tissue Sarcoma	Mesothelial Cell Liposarcoma	RBBR deletion/mutation	8.62e-12	0.86	0.31	0.90	0.67	0.20	0.99	30	335	365
Soft Tissue Sarcoma	Well-Differentiated Liposarcoma	TRAF3 oncogenic amplification	1.06e-08	0.93	0.60	1.00	0.29	0.08	1.00	16	349	365
Thyroid Cancer	Anaplastic Thyroid Cancer	MDM2 amplification	2e-12	0.93	0.40	0.94	0.59	0.09	1.00	17	343	365
Thyroid Cancer	Follicular Thyroid Cancer	TP53 deletion/mutation	7.05e-05	0.90	0.72	0.75	0.92	0.50	0.97	12	119	131
Thyroid Cancer	Medullary Thyroid Cancer	HRAS oncogenic mutation	2.3e-04	0.94	0.49	0.86	0.91	0.35	0.99	7	124	131
Thyroid Cancer	Medullary Thyroid Cancer	RET amplification/mutation	5.06e-20	1.00	1.00	1.00	1.00	0.94	1.00	15	115	131

8 Supplementary Information

8.1 Supplementary Methods

The aggregator model code can be accessed at <https://github.com/Paige-AI/paige-m1-sdk>. The aggregator model was trained with model parameters described in Table S1 and training parameters described in Table S2. The input embedding (generated by Virchow2) is a concatenation of the class token (dimension 1280) and mean patch tokens (dimension 1280). Training time took approximately 36 hours and was performed on 5 V100 nodes.

Table S1: Aggregator model parameters

Parameter description	Parameter name	Parameter value
Dimension of the input embedding	in_features	2560
Dimension of first linear layer	layer1_out_features	320
Dimension of second linear layer	layer2_out_features	640
Number of attention queries	n_attention_queries	16

Table S2: Aggregator model training parameters

Parameter name	Parameter value
Number of epochs	50
Batch size	2
Optimizer	AdamW [52]
Learning rate	0.0001
Weight decay	0.05
Training loss	Binary cross-entropy

8.2 Supplementary Tables

Table S3: Performance Metrics for Biomarker Prediction on the MSK Test Set. The performance of biomarkers in primary and metastatic lesions of the 15 most common cancer types are summarized in the following sub-tables S3.1 - S3.15. Metrics include Area Under the Curve (AUC), Average Precision (AP), Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV). Sample counts are provided for each biomarker, indicating the number of positive, negative, and total samples.

Table S3.1. Biomarkers detected in Bladder Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	MSI-H vs MSS	0.98	0.49	1.00	0.91	0.24	1.00	8	283	291
Primary	MSH2/MSH6 alterations	0.95	0.35	1.00	0.50	0.05	1.00	8	302	310
Primary	ASXL1 deletion/mutation	0.92	0.32	1.00	0.36	0.03	1.00	7	314	321
Primary	ASXL1 oncogenic mutation	0.91	0.23	1.00	0.28	0.03	1.00	7	314	321
Primary	E2F3 amplification	0.89	0.51	0.81	0.82	0.32	0.98	31	290	321
Primary	Lynch Syndrome	0.89	0.47	0.86	0.68	0.08	0.99	7	227	234
Primary	FGFR3 oncogenic mutation	0.88	0.74	0.92	0.60	0.41	0.96	74	247	321
Primary	FGFR3 amplification/mutation	0.87	0.73	0.95	0.47	0.35	0.97	75	246	321
Primary	MTAP deletion	0.83	0.60	0.80	0.69	0.36	0.94	20	91	111
Primary	TP53 oncogenic mutation	0.83	0.80	0.85	0.68	0.70	0.83	152	169	321
Primary	TP53 deletion/mutation	0.83	0.80	0.85	0.65	0.69	0.83	153	168	321
Primary	CDKN2A (p14ARF) deletion	0.83	0.56	0.92	0.52	0.32	0.96	64	257	321
Primary	RB1 deletion/mutation	0.82	0.55	0.92	0.52	0.30	0.96	59	262	321
Primary	CDKN2A (p16INK4a) deletion	0.82	0.54	0.92	0.56	0.34	0.97	64	257	321
Primary	CDKN2B deletion	0.82	0.54	0.94	0.52	0.33	0.97	64	257	321
Primary	MTAP deletion/mutation	0.82	0.56	0.80	0.65	0.33	0.94	20	91	111
Primary	CDKN2B deletion/mutation	0.81	0.50	0.91	0.54	0.33	0.96	64	257	321
Primary	CCNE1 amplification	0.81	0.12	0.86	0.60	0.09	0.99	14	307	321
Primary	FGFR2 amplification/mutation	0.80	0.11	0.88	0.63	0.06	0.99	8	313	321
Primary	CCNE1 amplification/mutation	0.80	0.11	0.86	0.64	0.10	0.99	14	307	321
Primary	AKT2 amplification	0.80	0.09	0.88	0.68	0.07	1.00	8	313	321
Primary	ERBB2 amplification	0.79	0.19	0.82	0.61	0.11	0.98	17	304	321
Primary	RB1 oncogenic mutation	0.79	0.48	0.87	0.53	0.26	0.95	52	269	321
Primary	ATM oncogenic mutation	0.78	0.25	1.00	0.41	0.07	1.00	13	308	321
Primary	MCL1 amplification/mutation	0.78	0.25	0.84	0.54	0.10	0.98	19	302	321
Primary	KMT2B deletion/mutation	0.78	0.19	0.80	0.52	0.06	0.99	10	262	272
Primary	ATM deletion/mutation	0.78	0.19	0.93	0.35	0.06	0.99	14	307	321
Primary	FGF3 amplification	0.78	0.36	0.91	0.48	0.16	0.98	32	289	321
Primary	TERT amplification/mutation	0.78	0.80	0.98	0.24	0.65	0.91	190	131	321
Primary	TERT oncogenic mutation	0.78	0.80	0.98	0.23	0.64	0.91	186	135	321
Primary	TERT amplification	0.77	0.09	0.89	0.30	0.04	0.99	9	312	321
Primary	CCND1 amplification	0.77	0.41	0.95	0.29	0.16	0.98	39	282	321
Primary	RAF1 amplification	0.77	0.11	1.00	0.23	0.06	1.00	14	307	321
Primary	FOXP1 amplification/mutation	0.77	0.21	0.82	0.33	0.04	0.98	11	310	321
Primary	MCL1 amplification	0.76	0.25	0.84	0.47	0.09	0.98	19	302	321
Primary	CCND1 amplification/mutation	0.76	0.40	0.90	0.41	0.17	0.97	39	282	321
Primary	SDHC amplification	0.76	0.17	1.00	0.38	0.11	1.00	23	298	321

Table S3.1. Biomarkers detected in Bladder Cancer (continued)

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	FGF19 amplification	0.76	0.33	0.94	0.32	0.15	0.98	36	285	321
Primary	KRAS amplification/mutation	0.76	0.47	0.80	0.38	0.06	0.97	15	306	321
Primary	MYCL amplification	0.76	0.13	0.86	0.46	0.07	0.99	14	307	321
Primary	KMT2B oncogenic mutation	0.76	0.21	0.80	0.51	0.06	0.99	10	262	272
Primary	FGF3 amplification/mutation	0.76	0.31	0.84	0.54	0.17	0.97	32	289	321
Primary	RAF1 amplification/mutation	0.76	0.11	0.93	0.28	0.06	0.99	14	307	321
Primary	MYCL amplification/mutation	0.75	0.10	0.93	0.39	0.07	0.99	14	307	321
Primary	CDH1 oncogenic mutation	0.75	0.13	0.80	0.68	0.08	0.99	10	311	321
Metastasis	ASXL1 amplification	0.93	0.47	1.00	0.86	0.27	1.00	4	77	81
Metastasis	NCOR1 deletion/mutation	0.91	0.25	1.00	0.58	0.08	1.00	3	78	81
Metastasis	CDKN2B deletion	0.90	0.69	1.00	0.59	0.26	1.00	10	71	81
Metastasis	IL7R amplification	0.89	0.20	1.00	0.67	0.10	1.00	3	78	81
Metastasis	CDKN2A (p14INK4a) deletion	0.89	0.62	0.89	0.58	0.21	0.98	9	72	81
Metastasis	SRC amplification	0.88	0.50	1.00	0.55	0.08	1.00	3	78	81
Metastasis	SRC amplification/mutation	0.88	0.47	1.00	0.60	0.09	1.00	3	78	81
Metastasis	CDKN2B deletion/mutation	0.88	0.57	1.00	0.61	0.26	1.00	10	71	81
Metastasis	MYCL amplification	0.88	0.17	1.00	0.60	0.09	1.00	3	78	81
Metastasis	CDKN2A (p16INK4a) deletion	0.88	0.61	0.78	0.61	0.20	0.96	9	72	81
Metastasis	ERBB2 amplification	0.86	0.51	0.83	0.55	0.13	0.98	6	75	81
Metastasis	MYCL amplification/mutation	0.86	0.17	1.00	0.54	0.08	1.00	3	78	81
Metastasis	GATA3 amplification	0.86	0.65	1.00	0.37	0.11	1.00	6	75	81
Metastasis	RICTOR amplification	0.85	0.16	1.00	0.59	0.09	1.00	3	78	81
Metastasis	TBX3 amplification	0.85	0.42	1.00	0.62	0.09	1.00	3	78	81
Metastasis	PPARG amplification	0.85	0.22	1.00	0.28	0.09	1.00	4	60	64
Metastasis	CREBBP deletion/mutation	0.83	0.36	1.00	0.66	0.13	1.00	4	77	81
Metastasis	CDK12 amplification	0.83	0.31	0.80	0.53	0.10	0.98	5	76	81
Metastasis	RAF1 amplification	0.81	0.20	1.00	0.32	0.07	1.00	4	77	81
Metastasis	TP53 deletion/mutation	0.81	0.80	0.86	0.68	0.76	0.81	44	37	81
Metastasis	RAF1 amplification/mutation	0.80	0.17	1.00	0.63	0.15	1.00	5	76	81
Metastasis	FGFR3 amplification/mutation	0.80	0.31	0.80	0.77	0.33	0.96	10	71	81
Metastasis	TP53 oncogenic mutation	0.79	0.77	0.89	0.65	0.75	0.83	44	37	81
Metastasis	TERT amplification/mutation	0.78	0.75	0.86	0.62	0.71	0.80	42	39	81
Metastasis	SDHC amplification	0.77	0.34	0.86	0.50	0.14	0.97	7	74	81
Metastasis	ERBB2 amplification/mutation	0.77	0.54	0.93	0.34	0.23	0.96	14	67	81
Metastasis	PIK3CA amplification/mutation	0.75	0.31	1.00	0.29	0.15	1.00	9	72	81

Table S3.2. Biomarkers detected in Breast Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	CDH1 oncogenic mutation	0.94	0.67	0.89	0.84	0.43	0.98	71	521	592
Primary	CDH1 deletion/mutation	0.94	0.68	0.88	0.83	0.42	0.98	72	520	592
Primary	Fragment of genome altered ≥ 30%	0.91	0.95	0.91	0.74	0.85	0.83	313	192	505
Primary	TP53 oncogenic mutation	0.91	0.89	0.91	0.66	0.65	0.91	243	349	592
Primary	TP53 deletion/mutation	0.91	0.89	0.90	0.67	0.66	0.91	247	345	592
Primary	Genome instability ≥ 20%	0.90	0.89	0.95	0.66	0.70	0.94	233	272	505
Primary	SRSF2 amplification	0.90	0.15	1.00	0.75	0.09	1.00	14	578	592
Primary	CDK12 amplification	0.89	0.46	1.00	0.48	0.13	1.00	43	549	592
Primary	Tetraploidy	0.88	0.84	0.90	0.61	0.58	0.91	178	296	474
Primary	AURKA amplification	0.88	0.12	1.00	0.41	0.04	1.00	15	577	592
Primary	Whole genome doubling	0.88	0.75	0.90	0.62	0.45	0.95	100	284	384
Primary	MSI2 amplification	0.87	0.19	0.94	0.56	0.07	1.00	17	493	510
Primary	Loss of heterozygosity ≥ 50%	0.87	0.33	0.96	0.53	0.10	1.00	27	478	505
Primary	ELF3 amplification	0.86	0.40	0.94	0.44	0.05	1.00	16	494	510
Primary	AURKA amplification/mutation	0.86	0.12	1.00	0.26	0.03	1.00	15	577	592
Primary	CCNE1 amplification	0.86	0.17	0.85	0.71	0.06	1.00	13	579	592
Primary	ERBB2 amplification	0.86	0.55	0.96	0.48	0.20	0.99	70	522	592
Primary	GAB2 amplification	0.85	0.23	1.00	0.64	0.14	1.00	4	70	74
Primary	CCNE1 amplification/mutation	0.85	0.14	0.77	0.79	0.08	0.99	13	579	592
Primary	RTELL1 amplification	0.85	0.18	0.94	0.51	0.06	1.00	16	494	510
Primary	RECQL4 amplification	0.85	0.23	0.96	0.66	0.11	1.00	26	566	592
Primary	HRD mutation signature	0.84	0.53	0.90	0.65	0.26	0.98	40	291	331
Primary	CD79B amplification	0.84	0.20	0.88	0.65	0.10	0.99	24	568	592
Primary	NBN amplification	0.84	0.29	0.79	0.73	0.13	0.99	28	564	592
Primary	BRIP1 amplification	0.84	0.18	0.91	0.64	0.09	0.99	22	570	592
Primary	ERBB2 amplification/mutation	0.84	0.55	0.91	0.54	0.25	0.97	85	507	592
Primary	AGO2 amplification	0.83	0.22	0.92	0.67	0.12	0.99	24	486	510
Primary	PPM1D amplification/mutation	0.83	0.23	0.97	0.51	0.11	1.00	36	556	592
Primary	RPS6KB2 amplification	0.83	0.18	0.83	0.61	0.04	0.99	12	580	592
Primary	ERBB2 oncogenic mutation	0.82	0.17	0.84	0.58	0.06	0.99	19	573	592
Primary	BRCA1 germline/somatic point mutation/deletion or fusion	0.82	0.23	0.82	0.57	0.10	0.98	17	282	299
Primary	PPM1D amplification	0.82	0.24	0.94	0.54	0.11	0.99	35	557	592
Primary	CBFB oncogenic mutation	0.82	0.23	0.95	0.45	0.05	1.00	19	573	592
Primary	AXIN2 amplification	0.81	0.14	0.87	0.58	0.08	0.99	23	569	592
Primary	RNF43 amplification	0.81	0.25	0.81	0.61	0.09	0.99	26	566	592
Primary	RB1 deletion/mutation	0.81	0.14	0.86	0.65	0.08	0.99	21	571	592
Primary	RAD51C amplification	0.81	0.20	0.83	0.58	0.09	0.98	29	563	592
Primary	GNAS amplification	0.80	0.15	0.92	0.43	0.03	1.00	12	580	592
Primary	GAB2 amplification/mutation	0.80	0.15	1.00	0.50	0.10	1.00	4	70	74
Primary	MYC amplification/mutation	0.80	0.29	0.82	0.58	0.16	0.97	51	541	592
Primary	HOXB13 amplification	0.80	0.12	1.00	0.40	0.05	1.00	17	575	592
Primary	RAD21 amplification	0.80	0.30	0.86	0.56	0.13	0.98	43	549	592
Primary	MCL1 amplification	0.80	0.26	0.80	0.50	0.05	0.99	20	572	592
Primary	MYC amplification	0.79	0.26	0.86	0.59	0.17	0.98	51	541	592
Primary	PRKARIA amplification	0.79	0.12	1.00	0.44	0.05	1.00	18	574	592
Primary	SMARCE1 amplification	0.79	0.12	1.00	0.37	0.06	1.00	3	71	74
Primary	CBFB deletion/mutation	0.78	0.19	0.95	0.33	0.04	0.99	19	573	592
Primary	SPOP amplification	0.78	0.14	0.83	0.60	0.08	0.99	24	568	592
Primary	FGF4 amplification	0.78	0.37	0.90	0.49	0.21	0.97	78	514	592
Primary	MAP3K1 oncogenic mutation	0.78	0.26	0.90	0.34	0.07	0.98	31	561	592
Primary	MDM2 amplification	0.78	0.06	0.77	0.73	0.06	0.99	13	579	592
Primary	PIK3CA oncogenic mutation	0.78	0.62	0.95	0.30	0.41	0.92	198	394	592
Primary	PIK3CA amplification/mutation	0.78	0.62	0.94	0.33	0.41	0.92	198	394	592
Primary	FGF3 amplification	0.78	0.34	0.90	0.46	0.20	0.97	77	515	592
Primary	MCL1 amplification/mutation	0.77	0.20	0.80	0.55	0.06	0.99	20	572	592
Primary	HRD pathway	0.77	0.41	0.93	0.46	0.25	0.97	96	496	592
Primary	MAP3K1 deletion/mutation	0.77	0.26	0.94	0.39	0.08	0.99	32	560	592
Primary	FGF19 amplification/mutation	0.77	0.37	0.89	0.50	0.22	0.97	80	512	592
Primary	FGF4 amplification/mutation	0.77	0.36	0.90	0.47	0.21	0.97	78	514	592
Primary	FGF3 amplification/mutation	0.77	0.34	0.88	0.47	0.20	0.96	77	515	592
Primary	FGF19 amplification	0.77	0.36	0.91	0.50	0.22	0.97	80	512	592
Primary	CCND1 amplification/mutation	0.77	0.38	0.89	0.45	0.21	0.96	85	507	592
Primary	CCND1 amplification	0.76	0.39	0.89	0.44	0.21	0.96	85	507	592
Primary	FGFR1 amplification/mutation	0.76	0.34	0.95	0.21	0.12	0.97	63	529	592
Primary	H3C14 amplification	0.76	0.08	0.93	0.24	0.03	0.99	14	578	592
Primary	RTK pathway	0.75	0.70	0.87	0.44	0.56	0.81	268	324	592

Table S3.2. Biomarkers detected in Breast Cancer (continued)

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Metastasis	CDH1 deletion/mutation	0.91	0.64	0.94	0.75	0.34	0.99	66	469	535
Metastasis	CDH1 oncogenic mutation	0.91	0.59	0.94	0.71	0.30	0.99	62	473	535
Metastasis	Whole genome doubling	0.90	0.71	0.90	0.68	0.44	0.96	63	225	288
Metastasis	Fragment of genome altered \geq 30%	0.85	0.93	0.94	0.46	0.82	0.74	316	123	439
Metastasis	Tetraploidy	0.85	0.78	0.89	0.60	0.63	0.88	171	227	398
Metastasis	Genome instability \geq 20%	0.84	0.84	0.93	0.51	0.67	0.88	225	214	439
Metastasis	Loss of heterozygosity \geq 50%	0.83	0.16	1.00	0.42	0.07	1.00	17	422	439
Metastasis	MTAP deletion/mutation	0.82	0.15	1.00	0.60	0.11	1.00	4	82	86
Metastasis	RECQL4 amplification	0.81	0.26	0.91	0.55	0.15	0.99	43	492	535
Metastasis	MYC amplification/mutation	0.79	0.36	0.89	0.48	0.19	0.97	65	470	535
Metastasis	MYC amplification	0.78	0.37	0.88	0.49	0.19	0.97	65	470	535
Metastasis	TP53 oncogenic mutation	0.78	0.70	0.84	0.54	0.54	0.84	211	324	535
Metastasis	TP53 deletion/mutation	0.78	0.70	0.86	0.56	0.57	0.85	216	319	535
Metastasis	CDKN2A (p14ARF) deletion	0.78	0.15	0.89	0.49	0.09	0.99	28	507	535
Metastasis	CDK12 amplification	0.77	0.32	0.81	0.43	0.10	0.97	37	498	535
Metastasis	IL10 amplification	0.77	0.22	0.93	0.29	0.04	0.99	15	520	535
Metastasis	NBN amplification	0.77	0.29	0.85	0.55	0.11	0.98	33	502	535
Metastasis	FOXA1 amplification/mutation	0.77	0.12	1.00	0.25	0.05	1.00	19	516	535
Metastasis	MTAP deletion	0.77	0.11	1.00	0.67	0.13	1.00	4	82	86
Metastasis	RAD21 amplification	0.77	0.30	0.88	0.40	0.13	0.97	50	485	535
Metastasis	AGO2 amplification	0.77	0.20	0.89	0.51	0.11	0.98	28	390	418
Metastasis	IKBKE amplification/mutation	0.77	0.17	0.94	0.24	0.04	0.99	16	519	535
Metastasis	CDKN2A (p16INK4a) deletion	0.76	0.12	0.90	0.55	0.10	0.99	29	506	535
Metastasis	IKBKE amplification	0.76	0.12	0.94	0.27	0.04	0.99	16	519	535
Metastasis	ELF3 amplification	0.75	0.14	0.94	0.39	0.06	0.99	17	401	418
Metastasis	MDM4 amplification/mutation	0.75	0.12	0.81	0.62	0.08	0.99	21	514	535
Metastasis	MCL1 amplification/mutation	0.75	0.09	0.94	0.33	0.04	0.99	17	518	535

Table S3.3. Biomarkers detected in Colorectal Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	MMR deficiency	1.00	0.98	1.00	0.93	0.69	1.00	36	231	267
Primary	MSI-H vs MSS	0.98	0.93	0.90	0.95	0.74	0.98	84	538	622
Primary	MLH1 deficient with BRAF V600E	0.98	0.80	0.90	0.94	0.51	0.99	20	286	306
Primary	MLH1 and PMS2 deficient	0.98	0.84	0.84	0.96	0.67	0.98	31	289	320
Primary	MSH3 oncogenic mutation	0.97	0.69	0.93	0.91	0.47	0.99	46	516	562
Primary	MSH3 deletion/mutation	0.97	0.64	0.98	0.89	0.44	1.00	46	516	562
Primary	KMT2D deletion/mutation	0.96	0.65	1.00	0.65	0.16	1.00	39	605	644
Primary	HNF1A oncogenic mutation	0.96	0.52	0.95	0.90	0.25	1.00	21	623	644
Primary	KMT2D oncogenic mutation	0.96	0.62	1.00	0.67	0.16	1.00	39	605	644
Primary	MSH2 and MSH6 deficient	0.96	0.33	0.89	0.88	0.17	1.00	9	305	314
Primary	HNF1A deletion/mutation	0.96	0.45	0.86	0.92	0.27	0.99	21	623	644
Primary	MSH2/MSH6 alterations	0.96	0.54	0.97	0.79	0.23	1.00	39	589	628
Primary	HLA-B deletion/mutation	0.95	0.28	1.00	0.77	0.09	1.00	13	549	562
Primary	MSH6 oncogenic mutation	0.95	0.32	0.92	0.88	0.24	1.00	26	618	644
Primary	MSH6 deletion/mutation	0.94	0.30	0.92	0.89	0.27	1.00	26	618	644
Primary	MSH2 deletion/mutation	0.94	0.29	1.00	0.73	0.09	1.00	17	627	644
Primary	MSH2 oncogenic mutation	0.94	0.33	1.00	0.72	0.09	1.00	17	627	644
Primary	BLM oncogenic mutation	0.93	0.25	1.00	0.48	0.04	1.00	13	631	644
Primary	BLM deletion/mutation	0.93	0.30	1.00	0.49	0.04	1.00	13	631	644
Primary	RECQL deletion/mutation	0.92	0.21	0.92	0.84	0.11	1.00	12	550	562
Primary	KMT2B deletion/mutation	0.92	0.48	0.85	0.89	0.38	0.99	41	521	562
Primary	GNAS oncogenic mutation	0.92	0.21	0.94	0.79	0.11	1.00	17	627	644
Primary	KMT2B oncogenic mutation	0.92	0.47	0.90	0.86	0.34	0.99	41	521	562
Primary	EP300 deletion/mutation	0.92	0.19	1.00	0.46	0.05	1.00	17	627	644
Primary	MSH2/MSH6 deficient	0.91	0.42	0.93	0.78	0.17	1.00	15	305	320
Primary	MLH1 oncogenic mutation	0.91	0.20	0.77	0.87	0.11	0.99	13	631	644
Primary	MLH1 deletion/mutation	0.91	0.15	0.77	0.86	0.10	0.99	13	631	644
Primary	CIC deletion/mutation	0.91	0.21	0.94	0.55	0.06	1.00	18	626	644
Primary	PTCH1 oncogenic mutation	0.91	0.16	0.81	0.84	0.15	0.99	21	623	644
Primary	PTPRS oncogenic mutation	0.90	0.25	0.85	0.90	0.21	0.99	20	624	644
Primary	EP300 oncogenic mutation	0.90	0.16	1.00	0.45	0.05	1.00	17	627	644
Primary	RNF43 oncogenic mutation	0.90	0.61	0.94	0.62	0.21	0.99	62	582	644
Primary	JAK1 deletion/mutation	0.90	0.20	0.90	0.83	0.14	1.00	20	624	644
Primary	TERT oncogenic mutation	0.90	0.16	1.00	0.25	0.03	1.00	15	629	644
Primary	TMB-H	0.90	0.82	0.94	0.51	0.32	0.97	125	513	638
Primary	JAK1 amplification/mutation	0.90	0.21	0.85	0.88	0.18	0.99	20	624	644
Primary	RNF43 deletion/mutation	0.90	0.60	0.89	0.71	0.25	0.98	62	582	644
Primary	GNAS amplification/mutation	0.90	0.20	0.88	0.76	0.09	1.00	17	627	644
Primary	PTCH1 deletion/mutation	0.90	0.16	0.81	0.87	0.17	0.99	21	623	644
Primary	FAT1 oncogenic mutation	0.90	0.11	1.00	0.56	0.05	1.00	14	630	644
Primary	PTPRS deletion/mutation	0.90	0.20	0.76	0.93	0.26	0.99	21	623	644
Primary	CIC oncogenic mutation	0.89	0.17	0.94	0.37	0.04	1.00	18	626	644
Primary	Lynch Syndrome	0.88	0.39	0.96	0.43	0.10	0.99	25	381	406
Primary	INPPL1 deletion/mutation	0.88	0.29	0.82	0.91	0.22	0.99	17	545	562
Primary	SRC amplification/mutation	0.87	0.23	0.85	0.75	0.12	0.99	26	618	644
Primary	INPPL1 oncogenic mutation	0.87	0.26	0.88	0.86	0.17	1.00	17	545	562
Primary	RAD50 deletion/mutation	0.87	0.14	0.92	0.59	0.04	1.00	13	631	644
Primary	KMT2C deletion/mutation	0.86	0.13	0.95	0.52	0.06	1.00	20	624	644
Primary	BRCA2 oncogenic mutation	0.86	0.26	0.91	0.70	0.10	1.00	22	622	644
Primary	BRCA2 deletion/mutation	0.86	0.27	0.82	0.80	0.12	0.99	22	622	644
Primary	NKX3-1 deletion/mutation	0.86	0.13	0.77	0.79	0.07	0.99	13	631	644
Primary	KMT2C oncogenic mutation	0.86	0.12	1.00	0.39	0.05	1.00	20	624	644
Primary	RAD50 oncogenic mutation	0.86	0.12	0.92	0.56	0.04	1.00	13	631	644
Primary	BRCA1/2 carriers	0.86	0.35	0.89	0.69	0.12	0.99	27	589	616
Primary	NFI deletion/mutation	0.86	0.12	0.78	0.71	0.07	0.99	18	626	644
Primary	CREBBP deletion/mutation	0.85	0.24	1.00	0.28	0.05	1.00	22	622	644
Primary	BRAF amplification/mutation	0.85	0.44	0.91	0.48	0.15	0.98	58	586	644
Primary	SRC amplification	0.85	0.20	0.77	0.74	0.11	0.99	26	618	644
Primary	RASA1 oncogenic mutation	0.85	0.13	0.78	0.82	0.11	0.99	18	626	644
Primary	BRAF oncogenic mutation	0.85	0.41	0.91	0.50	0.15	0.98	58	586	644
Primary	CTCF deletion/mutation	0.85	0.19	0.79	0.84	0.13	0.99	19	625	644
Primary	MGA oncogenic mutation	0.85	0.22	0.83	0.71	0.08	0.99	18	626	644
Primary	BRCA2 germline/somatic mutation, deep deletion or fusion	0.85	0.29	0.91	0.66	0.13	0.99	22	404	426
Primary	RTE11 amplification	0.85	0.32	0.97	0.40	0.09	1.00	34	528	562
Primary	GNAS amplification	0.85	0.17	0.95	0.61	0.07	1.00	20	624	644
Primary	MGA deletion/mutation	0.84	0.12	0.83	0.77	0.09	0.99	18	626	644
Primary	B2M oncogenic mutation	0.84	0.27	0.76	0.85	0.19	0.99	29	615	644
Primary	SETD2 oncogenic mutation	0.84	0.22	0.84	0.58	0.07	0.99	25	619	644
Primary	CREBBP oncogenic mutation	0.84	0.24	0.95	0.33	0.05	1.00	22	622	644
Primary	NCOA3 amplification	0.84	0.16	0.80	0.73	0.09	0.99	20	624	644
Primary	NFI oncogenic mutation	0.83	0.16	0.94	0.64	0.07	1.00	18	626	644
Primary	B2M deletion/mutation	0.83	0.27	0.86	0.67	0.11	0.99	29	615	644
Primary	BRCA1/2 somatic/germline mutant	0.83	0.30	0.90	0.57	0.13	0.99	29	394	423
Primary	SETD2 deletion/mutation	0.83	0.24	0.76	0.80	0.13	0.99	25	619	644
Primary	ASXL1 amplification	0.83	0.17	0.85	0.70	0.11	0.99	27	617	644
Primary	PTEN oncogenic mutation	0.83	0.20	0.97	0.43	0.09	1.00	36	608	644
Primary	DNMT3B amplification	0.83	0.19	0.92	0.67	0.10	1.00	25	619	644

Table S3.3. Biomarkers detected in Colorectal Cancer (continued)

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	AXIN2 deletion/mutation	0.82	0.21	0.92	0.26	0.05	0.99	25	619	644
Primary	NSD3 amplification	0.82	0.10	0.77	0.67	0.05	0.99	13	549	562
Primary	AXIN2 oncogenic mutation	0.82	0.22	0.88	0.45	0.06	0.99	25	619	644
Primary	ARID1A deletion/mutation	0.82	0.42	0.90	0.46	0.16	0.97	68	576	644
Primary	ARID1A oncogenic mutation	0.82	0.41	0.93	0.36	0.14	0.98	67	577	644
Primary	FLCN oncogenic mutation	0.82	0.14	0.86	0.72	0.06	1.00	14	630	644
Primary	PTEN deletion/mutation	0.82	0.21	0.98	0.27	0.08	0.99	41	603	644
Primary	SMARCA4 oncogenic mutation	0.81	0.13	0.95	0.32	0.05	1.00	22	622	644
Primary	ARID1B deletion/mutation	0.81	0.12	1.00	0.24	0.03	1.00	16	628	644
Primary	TP53 oncogenic mutation	0.81	0.91	0.86	0.58	0.85	0.61	471	173	644
Primary	TP53 deletion/mutation	0.81	0.91	0.87	0.57	0.85	0.61	474	170	644
Primary	LATS2 amplification	0.81	0.18	0.78	0.69	0.07	0.99	18	626	644
Primary	SMAD3 deletion/mutation	0.80	0.31	0.89	0.56	0.06	0.99	19	625	644
Primary	MYC amplification	0.80	0.15	0.81	0.61	0.05	0.99	16	628	644
Primary	PTPRT deletion/mutation	0.80	0.15	0.91	0.34	0.05	0.99	22	622	644
Primary	SMARCA4 deletion/mutation	0.80	0.12	0.91	0.36	0.05	0.99	22	622	644
Primary	MYC amplification/mutation	0.80	0.12	0.82	0.66	0.06	0.99	17	627	644
Primary	BRCA1 germline/somatic point mutation/deletion or fusion	0.79	0.15	0.77	0.68	0.11	0.98	22	425	447
Primary	PTPRT oncogenic mutation	0.79	0.14	0.90	0.41	0.05	0.99	21	623	644
Primary	KRAS oncogenic mutation	0.79	0.72	0.93	0.37	0.50	0.89	260	384	644
Primary	FLCN deletion/mutation	0.79	0.16	0.88	0.57	0.05	0.99	17	627	644
Primary	EPHA3 oncogenic mutation	0.79	0.12	0.94	0.38	0.04	1.00	18	626	644
Primary	TCF7L2 oncogenic mutation	0.79	0.39	0.97	0.27	0.14	0.99	71	573	644
Primary	TCF7L2 deletion/mutation	0.79	0.40	0.94	0.34	0.15	0.98	71	573	644
Primary	KRAS amplification/mutation	0.78	0.73	0.93	0.33	0.49	0.87	265	379	644
Primary	TERT amplification/mutation	0.78	0.16	0.95	0.33	0.04	1.00	20	624	644
Primary	APC oncogenic mutation	0.78	0.88	0.95	0.39	0.79	0.75	458	186	644
Primary	CDK8 amplification	0.78	0.19	0.84	0.59	0.08	0.99	25	619	644
Primary	PIK3R1 deletion/mutation	0.78	0.20	0.95	0.30	0.08	0.99	38	606	644
Primary	APC deletion/mutation	0.78	0.87	0.95	0.39	0.79	0.76	459	185	644
Primary	ARID1B oncogenic mutation	0.78	0.11	0.94	0.31	0.03	0.99	16	628	644
Primary	ARID2 oncogenic mutation	0.78	0.06	0.86	0.63	0.05	0.99	14	630	644
Primary	FLT1 amplification	0.77	0.22	0.97	0.42	0.08	1.00	33	611	644
Primary	EPHA3 deletion/mutation	0.77	0.18	0.89	0.30	0.04	0.99	18	626	644
Primary	CDK8 amplification/mutation	0.77	0.18	0.96	0.44	0.07	1.00	25	619	644
Primary	FLT1 amplification/mutation	0.77	0.17	0.94	0.42	0.08	0.99	33	611	644
Primary	FLT3 amplification	0.76	0.17	0.93	0.46	0.11	0.99	41	603	644
Primary	ATRX deletion/mutation	0.76	0.09	0.85	0.50	0.03	0.99	13	631	644
Primary	SMAD2 deletion/mutation	0.75	0.13	0.77	0.63	0.09	0.98	30	614	644
Metastasis	ELF3 amplification	0.92	0.17	1.00	0.29	0.03	1.00	4	170	174
Metastasis	PREX2 amplification	0.91	0.31	1.00	0.64	0.08	1.00	5	169	174
Metastasis	ELOC amplification	0.90	0.24	0.86	0.83	0.13	0.99	7	234	241
Metastasis	NBN amplification	0.89	0.34	0.86	0.78	0.10	0.99	7	234	241
Metastasis	RAD21 amplification	0.88	0.24	0.80	0.77	0.13	0.99	10	231	241
Metastasis	AGO2 amplification	0.87	0.52	1.00	0.28	0.11	1.00	14	160	174
Metastasis	NKK3-1	0.87	0.28	0.92	0.64	0.13	0.99	13	228	241
Metastasis	RECQL4 amplification	0.87	0.46	0.94	0.31	0.09	0.99	16	225	241
Metastasis	SOX17 amplification	0.86	0.15	0.83	0.62	0.05	0.99	6	235	241
Metastasis	SOX9 oncogenic mutation	0.86	0.34	0.90	0.49	0.07	0.99	10	231	241
Metastasis	SRC amplification	0.85	0.28	0.94	0.74	0.22	0.99	18	223	241
Metastasis	PRDM14 amplification	0.85	0.15	0.80	0.80	0.11	0.99	5	169	174
Metastasis	SOX9 amplification/mutation	0.85	0.35	0.90	0.45	0.07	0.99	10	231	241
Metastasis	DNNMT3B amplification	0.85	0.18	0.92	0.64	0.12	0.99	12	229	241
Metastasis	SRC amplification	0.85	0.35	0.89	0.67	0.18	0.99	18	223	241
Metastasis	IIR2 amplification/mutation	0.84	0.10	1.00	0.43	0.04	1.00	5	236	241
Metastasis	NCOA3 amplification	0.84	0.20	0.85	0.71	0.14	0.99	13	228	241
Metastasis	BRCA2 germline/somatic point mutation/deletion or fusion	0.84	0.47	1.00	0.41	0.09	1.00	6	100	106
Metastasis	RTET1 amplification	0.83	0.40	0.95	0.41	0.19	0.98	22	152	174
Metastasis	DUSP4 deletion	0.83	0.19	0.91	0.62	0.14	0.99	11	163	174
Metastasis	BCL2L1 amplification	0.83	0.23	0.81	0.73	0.18	0.98	16	225	241
Metastasis	MYC amplification	0.82	0.37	0.94	0.44	0.12	0.99	18	223	241
Metastasis	SMAD2 deletion/mutation	0.82	0.35	0.80	0.66	0.09	0.99	10	231	241
Metastasis	CNCND2 amplification/mutation	0.82	0.15	1.00	0.34	0.04	1.00	6	235	241
Metastasis	MYC amplification/mutation	0.82	0.33	0.94	0.48	0.13	0.99	18	223	241
Metastasis	CNCND2 amplification	0.80	0.23	1.00	0.41	0.04	1.00	6	235	241
Metastasis	GNAS amplification	0.80	0.24	0.92	0.50	0.10	0.99	13	228	241
Metastasis	SOX9 deletion/mutation	0.80	0.27	0.91	0.40	0.07	0.99	11	230	241
Metastasis	KDM5A amplification	0.80	0.09	1.00	0.43	0.05	1.00	7	234	241
Metastasis	IIR2 amplification	0.79	0.08	1.00	0.51	0.04	1.00	5	236	241
Metastasis	SMAD3 deletion/mutation	0.78	0.08	0.83	0.69	0.06	0.99	6	235	241
Metastasis	KDM5A amplification/mutation	0.78	0.08	1.00	0.47	0.05	1.00	7	234	241
Metastasis	TGF beta pathway	0.77	0.59	0.90	0.38	0.32	0.92	58	183	241
Metastasis	TP53 deletion/mutation	0.77	0.91	0.86	0.50	0.86	0.49	189	52	241
Metastasis	CDK12 amplification	0.77	0.08	1.00	0.42	0.06	1.00	8	233	241
Metastasis	TP53 oncogenic mutation	0.76	0.90	0.89	0.47	0.86	0.56	188	53	241
Metastasis	BRCA1/2 germline/somatic point mutation/deletion or fusion	0.76	0.47	0.90	0.22	0.11	0.95	10	97	107
Metastasis	SMAD4 deletion/mutation	0.76	0.48	0.91	0.32	0.23	0.94	44	197	241
Metastasis	FAT1 deletion/mutation	0.76	0.40	0.80	0.31	0.02	0.99	5	236	241
Metastasis	DNA damage response	0.76	0.41	0.85	0.42	0.16	0.96	27	214	241
Metastasis	APC deletion/mutation	0.75	0.83	0.87	0.58	0.83	0.66	169	72	241
Metastasis	APC oncogenic mutation	0.75	0.83	0.85	0.58	0.83	0.63	169	72	241

Table S3.4. Biomarkers detected in Endometrial Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	FLT3 amplification/mutation	0.96	0.29	1.00	0.61	0.06	1.00	9	335	344
Primary	FLT3 oncogenic mutation	0.95	0.42	1.00	0.72	0.09	1.00	9	335	344
Primary	RAD50 oncogenic mutation	0.95	0.20	1.00	0.75	0.08	1.00	7	337	344
Primary	GRIN2A deletion/mutation	0.94	0.28	1.00	0.35	0.04	1.00	8	336	344
Primary	MMR deficiency	0.94	0.76	0.96	0.71	0.43	0.99	27	122	149
Primary	RAD50 deletion/mutation	0.93	0.19	1.00	0.72	0.07	1.00	7	337	344
Primary	POLD1 oncogenic mutation	0.93	0.16	1.00	0.26	0.03	1.00	7	337	344
Primary	MLH1 deficient with BRAF V600E	0.92	0.63	0.94	0.74	0.27	0.99	16	160	176
Primary	SOX17 amplification	0.92	0.18	1.00	0.68	0.07	1.00	8	336	344
Primary	SETD2 oncogenic mutation	0.91	0.22	0.78	0.81	0.10	0.99	9	335	344
Primary	TP53 deletion/mutation	0.91	0.88	0.89	0.78	0.72	0.92	133	211	344
Primary	POLD1 deletion/mutation	0.91	0.11	1.00	0.51	0.04	1.00	7	337	344
Primary	SETD2 deletion/mutation	0.91	0.25	1.00	0.77	0.10	1.00	9	335	344
Primary	TP53 oncogenic mutation	0.91	0.87	0.89	0.80	0.74	0.92	133	211	344
Primary	ESR1 oncogenic mutation	0.91	0.28	1.00	0.40	0.04	1.00	8	336	344
Primary	XPO1 oncogenic mutation	0.91	0.31	1.00	0.20	0.04	1.00	12	332	344
Primary	BRCA1 germline/somatic mutation, deep deletion or fusion	0.90	0.29	0.89	0.78	0.12	0.99	9	249	258
Primary	TMB-H	0.90	0.81	0.93	0.67	0.58	0.95	113	230	343
Primary	BRCA1/2 carriers	0.90	0.45	0.95	0.65	0.15	0.99	19	300	319
Primary	CCNE1 amplification/mutation	0.90	0.25	0.78	0.84	0.21	0.99	18	326	344
Primary	APC oncogenic mutation	0.90	0.33	0.91	0.73	0.19	0.99	22	322	344
Primary	JAK1 amplification/mutation	0.90	0.51	0.85	0.78	0.34	0.98	40	304	344
Primary	ATR deletion/mutation	0.89	0.24	1.00	0.48	0.05	1.00	10	334	344
Primary	MSI-H vs MSS	0.89	0.56	0.89	0.78	0.49	0.97	57	250	307
Primary	ATR oncogenic mutation	0.89	0.19	1.00	0.35	0.04	1.00	10	334	344
Primary	JAK1 oncogenic mutation	0.89	0.49	0.78	0.80	0.33	0.96	40	304	344
Primary	MSH2/MSH6 deficient	0.89	0.38	1.00	0.57	0.14	1.00	13	181	194
Primary	CHEK2 deletion/mutation	0.89	0.09	1.00	0.47	0.04	1.00	7	337	344
Primary	XPO1 amplification/mutation	0.89	0.32	1.00	0.23	0.04	1.00	12	332	344
Primary	JAK1 deletion/mutation	0.88	0.47	0.80	0.80	0.34	0.97	40	304	344
Primary	MSH3 oncogenic mutation	0.88	0.24	0.83	0.80	0.20	0.99	18	296	314
Primary	APC deletion/mutation	0.88	0.30	0.82	0.78	0.20	0.98	22	322	344
Primary	BRCNA2 oncogenic mutation	0.88	0.26	0.94	0.70	0.13	1.00	16	328	344
Primary	CTNNB1 oncogenic mutation	0.88	0.64	0.94	0.47	0.24	0.98	53	291	344
Primary	NSD1 deletion/mutation	0.88	0.36	0.95	0.65	0.15	1.00	21	323	344
Primary	PTEN oncogenic mutation	0.88	0.89	0.94	0.55	0.72	0.88	192	152	344
Primary	BRCNA2 deletion/mutation	0.88	0.24	0.94	0.69	0.13	1.00	16	328	344
Primary	ATM oncogenic mutation	0.87	0.38	1.00	0.54	0.18	1.00	31	313	344
Primary	MSH6 deletion/mutation	0.87	0.31	0.83	0.77	0.20	0.98	23	321	344
Primary	PTEN deletion/mutation	0.86	0.87	0.94	0.52	0.72	0.88	194	150	344
Primary	BRCNA1 deletion/mutation	0.86	0.11	0.86	0.84	0.10	1.00	7	337	344
Primary	NSD1 amplification/mutation	0.86	0.36	0.86	0.67	0.14	0.99	21	323	344
Primary	CTNNB1 amplification/mutation	0.86	0.63	0.94	0.45	0.24	0.98	53	291	344
Primary	ESR1 amplification/mutation	0.86	0.34	0.89	0.38	0.04	0.99	9	335	344
Primary	INPP4B deletion/mutation	0.86	0.32	1.00	0.65	0.06	1.00	8	336	344
Primary	SPEN deletion/mutation	0.86	0.14	1.00	0.52	0.06	1.00	10	334	344
Primary	MSH3 deletion/mutation	0.86	0.25	0.89	0.74	0.18	0.99	19	295	314
Primary	ATM deletion/mutation	0.86	0.34	0.94	0.58	0.18	0.99	31	313	344
Primary	NSD1 oncogenic mutation	0.86	0.33	0.90	0.61	0.13	0.99	21	323	344
Primary	INPP4B oncogenic mutation	0.86	0.16	1.00	0.62	0.06	1.00	8	336	344
Primary	SPEN oncogenic mutation	0.85	0.14	0.90	0.71	0.09	1.00	10	334	344
Primary	Lynch Syndrome	0.85	0.28	1.00	0.42	0.05	1.00	8	263	271
Primary	POLE oncogenic mutation	0.85	0.38	0.86	0.69	0.20	0.98	29	315	344
Primary	MSH2/MSH6 alterations	0.85	0.39	0.89	0.59	0.21	0.98	37	296	333
Primary	MSH2 and MSH6 deficient	0.85	0.14	1.00	0.70	0.11	1.00	7	181	188
Primary	KMT2B oncogenic mutation	0.84	0.44	0.96	0.54	0.26	0.99	45	269	314
Primary	MSH6 oncogenic mutation	0.84	0.24	0.83	0.73	0.18	0.98	23	321	344
Primary	ARHGAP35 amplification/mutation	0.84	0.42	1.00	0.25	0.12	1.00	9	89	98
Primary	PTPRD deletion/mutation	0.84	0.09	0.86	0.79	0.08	1.00	7	337	344
Primary	ACVR1 amplification/mutation	0.84	0.22	1.00	0.60	0.06	1.00	8	336	344
Primary	LATS1 deletion/mutation	0.84	0.15	0.90	0.39	0.04	0.99	10	334	344
Primary	BRCA1/2 somatic/germline mutant	0.84	0.53	0.87	0.53	0.16	0.98	23	225	248
Primary	KMT2B deletion/mutation	0.84	0.43	0.89	0.59	0.27	0.97	45	269	314
Primary	PTCH1 deletion/mutation	0.83	0.20	1.00	0.64	0.11	1.00	15	329	344
Primary	ARID1A oncogenic mutation	0.83	0.76	0.97	0.37	0.55	0.95	153	191	344
Primary	mTOR pathway	0.83	0.95	0.97	0.22	0.85	0.61	281	63	344
Primary	BRCA2 germline/somatic point mutation/deletion or fusion	0.83	0.57	0.83	0.54	0.24	0.95	41	232	273
Primary	ZFHX3 oncogenic mutation	0.83	0.38	0.98	0.51	0.25	0.99	49	295	344
Primary	ACVR1 oncogenic mutation	0.83	0.12	0.88	0.70	0.06	1.00	8	336	344
Primary	STAG2 oncogenic mutation	0.83	0.14	1.00	0.41	0.05	1.00	10	334	344
Primary	TERT oncogenic mutation	0.83	0.43	0.88	0.61	0.15	0.98	25	319	344
Primary	INPL1 deletion/mutation	0.82	0.38	0.84	0.64	0.24	0.97	38	276	314
Primary	ARID1A deletion/mutation	0.82	0.76	0.96	0.38	0.56	0.92	154	190	344
Primary	ZFHX3 deletion/mutation	0.82	0.39	1.00	0.36	0.21	1.00	49	295	344
Primary	BRCA1/2 germline/somatic mutation/deletion or fusion	0.82	0.44	0.87	0.48	0.14	0.97	23	225	248
Primary	FAT1 deletion/mutation	0.82	0.25	0.92	0.46	0.12	0.99	26	318	344

Table S3.4. Biomarkers detected in Endometrial Cancer (continued)

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	MLH1 and PMS2 deficient	0.82	0.59	0.89	0.54	0.36	0.94	46	158	204
Primary	NF1 oncogenic mutation	0.82	0.40	0.85	0.55	0.17	0.97	34	310	344
Primary	FUBP1 deletion/mutation	0.82	0.13	0.91	0.50	0.06	0.99	11	333	344
Primary	EP300 deletion/mutation	0.82	0.08	1.00	0.58	0.07	1.00	10	334	344
Primary	TET2 deletion/mutation	0.81	0.11	1.00	0.41	0.04	1.00	8	336	344
Primary	TET2 oncogenic mutation	0.81	0.19	1.00	0.34	0.03	1.00	8	336	344
Primary	CCND1 oncogenic mutation	0.81	0.22	0.94	0.53	0.09	0.99	16	328	344
Primary	TSC2 oncogenic mutation	0.81	0.14	1.00	0.28	0.03	1.00	7	337	344
Primary	INPPL1 oncogenic mutation	0.81	0.37	0.82	0.67	0.26	0.96	38	276	314
Primary	HRD pathway	0.81	0.46	0.95	0.40	0.23	0.97	55	289	344
Primary	NF1 deletion/mutation	0.81	0.39	0.85	0.54	0.17	0.97	34	310	344
Primary	ANKRD11 deletion/mutation	0.81	0.18	0.90	0.66	0.07	1.00	10	334	344
Primary	MSH2 deletion/mutation	0.81	0.17	0.76	0.66	0.10	0.98	17	327	344
Primary	TGF beta pathway	0.81	0.26	0.81	0.59	0.09	0.98	16	328	344
Primary	LATS1 oncogenic mutation	0.81	0.20	1.00	0.24	0.04	1.00	10	334	344
Primary	FAT1 oncogenic mutation	0.80	0.27	0.88	0.62	0.16	0.98	26	318	344
Primary	KMT2D oncogenic mutation	0.80	0.31	0.90	0.51	0.19	0.97	39	305	344
Primary	SOX17 oncogenic mutation	0.80	0.29	0.96	0.41	0.10	0.99	23	321	344
Primary	EP300 oncogenic mutation	0.80	0.07	1.00	0.63	0.08	1.00	10	334	344
Primary	CTCF oncogenic mutation	0.80	0.42	0.94	0.45	0.23	0.98	51	293	344
Primary	FUBP1 oncogenic mutation	0.80	0.13	0.91	0.37	0.05	0.99	11	333	344
Primary	MSH2 oncogenic mutation	0.80	0.17	0.76	0.65	0.10	0.98	17	327	344
Primary	NOTCH2 deletion/mutation	0.80	0.11	0.78	0.54	0.04	0.99	9	335	344
Primary	PIK3CB amplification/mutation	0.80	0.13	1.00	0.21	0.03	1.00	8	336	344
Primary	MYC amplification	0.80	0.11	0.85	0.76	0.12	0.99	13	331	344
Primary	ATRX deletion/mutation	0.80	0.19	1.00	0.21	0.07	1.00	18	326	344
Primary	SOX17 deletion/mutation	0.79	0.28	0.96	0.37	0.10	0.99	23	321	344
Primary	KMT2D deletion/mutation	0.79	0.31	0.92	0.50	0.19	0.98	39	305	344
Primary	MAP3K1 deletion/mutation	0.79	0.15	0.94	0.21	0.06	0.99	17	327	344
Primary	STAG2 deletion/mutation	0.79	0.12	1.00	0.28	0.04	1.00	10	334	344
Primary	ATRX oncogenic mutation	0.79	0.20	0.89	0.47	0.08	0.99	18	326	344
Primary	SOX9 amplification/mutation	0.79	0.07	0.86	0.62	0.04	1.00	7	337	344
Primary	RASA1 oncogenic mutation	0.79	0.30	0.86	0.36	0.08	0.97	22	322	344
Primary	PAX5 deletion/mutation	0.79	0.20	0.86	0.61	0.04	1.00	7	337	344
Primary	KDM6A deletion/mutation	0.79	0.21	0.88	0.57	0.05	0.99	8	336	344
Primary	BCOR oncogenic mutation	0.79	0.21	0.97	0.35	0.13	0.99	31	313	344
Primary	NOTCH2 amplification/mutation	0.79	0.11	0.89	0.54	0.05	0.99	9	335	344
Primary	CTCF deletion/mutation	0.79	0.38	0.92	0.50	0.24	0.97	51	293	344
Primary	KMT2C oncogenic mutation	0.78	0.24	0.80	0.59	0.11	0.98	20	324	344
Primary	NOTCH2 oncogenic mutation	0.78	0.09	0.89	0.58	0.05	0.99	9	335	344
Primary	RNF43 oncogenic mutation	0.78	0.26	0.94	0.44	0.16	0.99	35	309	344
Primary	TERT amplification/mutation	0.78	0.36	0.82	0.47	0.12	0.97	28	316	344
Primary	KRAS oncogenic mutation	0.77	0.47	0.92	0.31	0.23	0.95	62	282	344
Primary	RNF43 deletion/mutation	0.77	0.25	0.94	0.39	0.15	0.98	35	309	344
Primary	DNA damage response	0.77	0.58	0.92	0.34	0.34	0.92	93	251	344
Primary	CIC deletion/mutation	0.77	0.10	0.92	0.39	0.06	0.99	13	331	344

Table S3.4. Biomarkers detected in Endometrial Cancer (continued)

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	MYC amplification/mutation	0.76	0.10	0.87	0.67	0.11	0.99	15	329	344
Primary	RASA1 deletion/mutation	0.76	0.31	0.78	0.39	0.08	0.96	23	321	344
Primary	BCOR deletion/mutation	0.76	0.21	0.94	0.38	0.13	0.98	31	313	344
Primary	KRAS amplification/mutation	0.76	0.45	0.97	0.25	0.22	0.97	62	282	344
Primary	CCND1 amplification/mutation	0.75	0.21	1.00	0.42	0.08	1.00	16	328	344
Primary	MAP3K1 oncogenic mutation	0.75	0.12	1.00	0.34	0.07	1.00	16	328	344
Metastasis	CTNNB1 amplification/mutation	0.96	0.79	1.00	0.66	0.31	1.00	11	74	85
Metastasis	MCL1 amplification/mutation	0.95	0.40	1.00	0.22	0.04	1.00	3	82	85
Metastasis	CTNNB1 oncogenic mutation	0.94	0.79	0.91	0.73	0.33	0.98	11	74	85
Metastasis	KMT2A deletion/mutation	0.93	0.27	1.00	0.85	0.20	1.00	3	82	85
Metastasis	ZFHX3 oncogenic mutation	0.92	0.33	1.00	0.70	0.20	1.00	6	79	85
Metastasis	TCF3 deletion/mutation	0.91	0.21	1.00	0.77	0.14	1.00	3	82	85
Metastasis	NF1 deletion/mutation	0.89	0.47	1.00	0.62	0.09	1.00	3	82	85
Metastasis	MSI-H	0.88	0.52	0.80	0.83	0.40	0.97	10	71	81
Metastasis	ZFHX3 deletion/mutation	0.88	0.26	1.00	0.58	0.15	1.00	6	79	85
Metastasis	H3C14 amplification	0.88	0.33	1.00	0.43	0.06	1.00	3	82	85
Metastasis	ARID1B oncogenic mutation	0.88	0.17	1.00	0.61	0.09	1.00	3	82	85
Metastasis	KMT2B deletion/mutation	0.86	0.30	0.80	0.80	0.24	0.98	5	66	71
Metastasis	PTEN deletion/mutation	0.85	0.81	0.79	0.82	0.69	0.89	28	57	85
Metastasis	H3C13 amplification	0.85	0.33	1.00	0.30	0.05	1.00	3	82	85
Metastasis	TP53 deletion/mutation	0.84	0.86	0.94	0.66	0.77	0.89	47	38	85
Metastasis	ARID1A deletion/mutation	0.84	0.74	0.77	0.78	0.61	0.88	26	59	85
Metastasis	ARID1A oncogenic mutation	0.84	0.72	0.85	0.71	0.56	0.91	26	59	85
Metastasis	KMT2D deletion/mutation	0.84	0.34	1.00	0.62	0.17	1.00	6	79	85
Metastasis	TP53 oncogenic mutation	0.83	0.84	0.93	0.59	0.73	0.88	46	39	85
Metastasis	KMT2B oncogenic mutation	0.83	0.24	0.80	0.76	0.20	0.98	5	66	71
Metastasis	ARID1B deletion/mutation	0.82	0.13	1.00	0.43	0.06	1.00	3	82	85
Metastasis	ERBB3 amplification/mutation	0.81	0.23	1.00	0.39	0.06	1.00	3	82	85
Metastasis	KMT2C oncogenic mutation	0.81	0.16	1.00	0.61	0.09	1.00	3	82	85
Metastasis	AGO2 amplification	0.81	0.20	1.00	0.49	0.08	1.00	3	68	71
Metastasis	KMT2D oncogenic mutation	0.80	0.35	1.00	0.59	0.16	1.00	6	79	85
Metastasis	KMT2C deletion/mutation	0.79	0.23	1.00	0.56	0.08	1.00	3	82	85
Metastasis	BCOR oncogenic mutation	0.78	0.25	0.86	0.64	0.18	0.98	7	78	85
Metastasis	SOX17 amplification	0.78	0.20	0.80	0.44	0.08	0.97	5	80	85
Metastasis	BRCA1/2 carriers	0.76	0.32	0.80	0.61	0.11	0.98	5	79	84

Table S3.5. Biomarkers detected in Esophagogastric Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	MMR deficiency	1.00	0.99	1.00	0.98	0.92	1.00	12	59	71
Primary	HNF1A deletion/mutation	0.94	0.41	0.86	0.94	0.33	1.00	7	218	225
Primary	KMT2A deletion/mutation	0.94	0.21	1.00	0.75	0.08	1.00	5	220	225
Primary	KMT2A oncogenic mutation	0.92	0.17	1.00	0.62	0.06	1.00	5	220	225
Primary	RASA1 oncogenic mutation	0.92	0.30	1.00	0.51	0.05	1.00	6	219	225
Primary	SMARCE1 amplification	0.89	0.43	1.00	0.35	0.07	1.00	3	66	69
Primary	KMT2D deletion/mutation	0.89	0.38	0.94	0.67	0.19	0.99	17	208	225
Primary	KMT2D oncogenic mutation	0.89	0.40	1.00	0.62	0.18	1.00	17	208	225
Primary	NF1 oncogenic mutation	0.88	0.33	1.00	0.47	0.04	1.00	5	220	225
Primary	RHOA oncogenic mutation	0.88	0.30	0.86	0.84	0.15	0.99	7	218	225
Primary	NF1 deletion/mutation	0.87	0.17	0.80	0.54	0.04	0.99	5	220	225
Primary	RASA1 deletion/mutation	0.87	0.18	1.00	0.45	0.05	1.00	6	219	225
Primary	CIC deletion/mutation	0.86	0.29	0.90	0.57	0.09	0.99	10	215	225
Primary	TMB-H	0.86	0.66	0.80	0.77	0.43	0.95	40	183	223
Primary	ZFHX3 oncogenic mutation	0.86	0.15	1.00	0.52	0.05	1.00	6	219	225
Primary	CTCF oncogenic mutation	0.86	0.09	0.80	0.87	0.12	0.99	5	220	225
Primary	KMT2B deletion/mutation	0.86	0.42	0.94	0.44	0.12	0.99	16	189	205
Primary	KMT2B oncogenic mutation	0.85	0.44	0.94	0.41	0.12	0.99	16	189	205
Primary	KRAS oncogenic mutation	0.84	0.27	1.00	0.22	0.07	1.00	12	213	225
Primary	BRCA1 germline/somatic point mutation/deletion or fusion	0.84	0.17	1.00	0.68	0.15	1.00	7	128	135
Primary	ZFHX3 deletion/mutation	0.84	0.19	1.00	0.44	0.05	1.00	6	219	225
Primary	GNAS amplification/mutation	0.83	0.41	0.83	0.69	0.07	0.99	6	219	225
Primary	CIC oncogenic mutation	0.83	0.20	0.80	0.62	0.09	0.99	10	215	225
Primary	CDK12 amplification	0.82	0.34	1.00	0.53	0.13	1.00	15	210	225
Primary	FBXW7 oncogenic mutation	0.82	0.15	0.89	0.69	0.11	0.99	9	216	225
Primary	FBXW7 deletion/mutation	0.82	0.17	0.80	0.71	0.11	0.99	10	215	225
Primary	ERBB2 amplification	0.81	0.47	0.93	0.53	0.22	0.98	28	197	225
Primary	BRCA1/2 carriers	0.80	0.35	0.80	0.65	0.10	0.99	10	209	219
Primary	MDM2 amplification	0.79	0.17	0.82	0.73	0.14	0.99	11	214	225
Primary	BRCA1/2 germline/somatic point mutation/deletion or fusion	0.79	0.52	0.86	0.38	0.21	0.94	22	118	140
Primary	ARID1B deletion/mutation	0.79	0.10	1.00	0.44	0.04	1.00	5	220	225
Primary	SMAD4 oncogenic mutation	0.78	0.15	1.00	0.47	0.08	1.00	10	215	225
Primary	HRD pathway	0.77	0.43	0.92	0.46	0.26	0.97	39	186	225
Primary	ERBB2 amplification/mutation	0.77	0.52	0.89	0.29	0.20	0.93	37	188	225
Primary	MYC amplification/mutation	0.76	0.25	0.80	0.60	0.12	0.98	15	210	225
Primary	CASP8 deletion/mutation	0.76	0.09	1.00	0.51	0.05	1.00	6	219	225
Primary	MYC amplification	0.76	0.23	0.87	0.59	0.13	0.98	15	210	225
Primary	TGF beta pathway	0.76	0.29	0.87	0.40	0.14	0.96	23	202	225
Primary	BRCA2 germline/somatic point mutation/deletion or fusion	0.76	0.32	0.81	0.49	0.17	0.95	16	123	139
Primary	ATM deletion/mutation	0.75	0.11	0.78	0.55	0.07	0.98	9	216	225
Primary	RARA amplification	0.75	0.24	0.77	0.62	0.11	0.98	13	212	225
Metastasis	CDKN2A (p16INK4a) oncogenic mutation	0.87	0.25	1.00	0.52	0.12	1.00	4	58	62
Metastasis	TMB-H	0.86	0.66	0.82	0.71	0.38	0.95	11	51	62
Metastasis	PTPRD deletion/mutation	0.85	0.20	1.00	0.32	0.07	1.00	3	59	62
Metastasis	MYC amplification	0.85	0.53	0.83	0.55	0.17	0.97	6	56	62
Metastasis	CCNE1 amplification/mutation	0.84	0.46	0.80	0.67	0.17	0.97	5	57	62
Metastasis	EGFR amplification/mutation	0.83	0.45	1.00	0.44	0.08	1.00	3	59	62
Metastasis	CCND1 amplification/mutation	0.83	0.27	1.00	0.54	0.19	1.00	6	56	62
Metastasis	CDK12 amplification	0.83	0.42	1.00	0.59	0.27	1.00	8	54	62
Metastasis	MYC amplification/mutation	0.82	0.49	0.83	0.59	0.18	0.97	6	56	62
Metastasis	CCNE1 amplification	0.82	0.45	0.80	0.60	0.15	0.97	5	57	62
Metastasis	SMAD4 deletion/mutation	0.81	0.26	1.00	0.60	0.15	1.00	4	58	62
Metastasis	CCND1 amplification	0.80	0.22	0.80	0.63	0.16	0.97	5	57	62
Metastasis	KDM5A amplification	0.80	0.21	1.00	0.51	0.09	1.00	3	59	62
Metastasis	KDM5A amplification/mutation	0.78	0.17	1.00	0.47	0.09	1.00	3	59	62

Table S3.6. Biomarkers detected in Glioma

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	FUBP1 oncogenic mutation	0.96	0.47	0.82	0.92	0.26	0.99	11	330	341
Primary	FUBP1 deletion/mutation	0.95	0.48	0.82	0.96	0.41	0.99	11	330	341
Primary	IDH1 oncogenic mutation	0.93	0.82	0.83	0.89	0.73	0.94	90	251	341
Primary	IDH1 amplification/mutation	0.93	0.83	0.84	0.88	0.72	0.94	90	251	341
Primary	RET deletion	0.89	0.23	0.79	0.84	0.18	0.99	14	327	341
Primary	TERT oncogenic mutation	0.88	0.83	0.79	0.83	0.79	0.84	150	191	341
Primary	TERT amplification/mutation	0.87	0.82	0.80	0.78	0.75	0.83	152	189	341
Primary	RB1 oncogenic mutation	0.86	0.24	0.96	0.62	0.17	0.99	25	316	341
Primary	ATRX deletion/mutation	0.85	0.65	0.94	0.37	0.22	0.97	54	287	341
Primary	PIK3CD deletion	0.85	0.26	1.00	0.54	0.06	1.00	9	332	341
Primary	EGFR amplification/mutation	0.84	0.70	0.86	0.55	0.40	0.92	88	253	341
Primary	CDKN2A (p16INK4a) deletion	0.84	0.79	0.89	0.56	0.57	0.89	134	207	341
Primary	RB1 deletion/mutation	0.84	0.30	0.93	0.59	0.17	0.99	28	313	341
Primary	EGFR amplification	0.84	0.69	0.84	0.58	0.39	0.92	82	259	341
Primary	ATRX oncogenic mutation	0.84	0.57	0.92	0.44	0.23	0.97	52	289	341
Primary	CDKN2A (p14ARF) deletion	0.83	0.76	0.89	0.59	0.59	0.89	135	206	341
Primary	NOTCH1 oncogenic mutation	0.83	0.32	1.00	0.47	0.07	1.00	13	328	341
Primary	CDKN2B deletion/mutation	0.83	0.76	0.86	0.56	0.55	0.87	132	209	341
Primary	MYCN amplification/mutation	0.82	0.24	0.91	0.46	0.05	0.99	11	330	341
Primary	TP53 deletion/mutation	0.82	0.72	0.89	0.52	0.56	0.88	139	202	341
Primary	CDKN2B deletion	0.82	0.74	0.83	0.66	0.60	0.86	132	209	341
Primary	TP53 oncogenic mutation	0.82	0.70	0.91	0.50	0.54	0.89	135	206	341
Primary	NOTCH1 deletion/mutation	0.82	0.24	0.85	0.51	0.06	0.99	13	328	341
Primary	NOTCH1 amplification/mutation	0.81	0.27	0.92	0.48	0.07	0.99	13	328	341
Primary	PTEN deletion	0.80	0.16	0.93	0.41	0.06	0.99	14	327	341
Primary	EGFR oncogenic mutation	0.79	0.38	0.83	0.50	0.16	0.96	35	306	341
Primary	EGFR fusion	0.79	0.36	0.80	0.57	0.22	0.95	46	295	341
Primary	PTEN deletion/mutation	0.77	0.50	0.85	0.58	0.40	0.92	85	256	341
Primary	KIT amplification	0.77	0.11	0.94	0.39	0.07	0.99	16	325	341
Primary	TEK deletion	0.77	0.24	0.86	0.55	0.10	0.98	14	240	254
Primary	PDGFRA amplification	0.76	0.22	0.96	0.38	0.11	0.99	26	315	341
Primary	SETD2 deletion/mutation	0.76	0.11	1.00	0.32	0.06	1.00	13	328	341
Primary	KDR amplification	0.76	0.11	0.86	0.41	0.06	0.99	14	327	341

Table S3.7. Biomarkers detected in Hepatobiliary Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	RB1 deletion	0.97	0.74	1.00	0.58	0.06	1.00	4	154	158
Primary	ELOC amplification	0.96	0.49	1.00	0.90	0.21	1.00	4	154	158
Primary	SMAD4 deletion/mutation	0.94	0.65	0.92	0.64	0.17	0.99	12	146	158
Primary	SMAD4 oncogenic mutation	0.92	0.48	0.90	0.65	0.15	0.99	10	148	158
Primary	ERBB3 amplification/mutation	0.88	0.20	0.83	0.78	0.13	0.99	6	152	158
Primary	FGFR2 fusion	0.88	0.36	1.00	0.51	0.17	1.00	14	144	158
Primary	TGF beta pathway	0.88	0.67	0.89	0.53	0.20	0.97	19	139	158
Primary	BAP1 deletion/mutation	0.85	0.25	1.00	0.61	0.12	1.00	8	150	158
Primary	TERT amplification/mutation	0.84	0.53	0.95	0.39	0.18	0.98	20	138	158
Primary	YES1 amplification	0.83	0.32	1.00	0.47	0.05	1.00	4	154	158
Primary	CDKN2B deletion	0.83	0.20	0.92	0.64	0.17	0.99	12	146	158
Primary	KRAS oncogenic mutation	0.83	0.47	0.78	0.64	0.22	0.96	18	140	158
Primary	ATM deletion/mutation	0.82	0.14	0.80	0.66	0.07	0.99	5	153	158
Primary	CDKN2A (p14ARF) deletion	0.82	0.23	0.85	0.63	0.17	0.98	13	145	158
Primary	CDKN2B deletion/mutation	0.82	0.21	0.92	0.63	0.17	0.99	12	146	158
Primary	CDKN2A (p16INK4a) oncogenic mutation	0.82	0.32	0.88	0.38	0.07	0.98	8	150	158
Primary	CDKN2A (p16INK4a) deletion	0.82	0.20	0.85	0.71	0.21	0.98	13	145	158
Primary	TP53 oncogenic mutation	0.82	0.73	0.88	0.46	0.43	0.89	50	108	158
Primary	TP53 deletion/mutation	0.81	0.72	0.90	0.44	0.42	0.90	50	108	158
Primary	BAP1 oncogenic mutation	0.80	0.14	1.00	0.62	0.11	1.00	7	151	158
Primary	YES1 amplification/mutation	0.77	0.37	1.00	0.40	0.04	1.00	4	154	158
Primary	ATM oncogenic mutation	0.76	0.11	0.80	0.53	0.05	0.99	5	153	158
Metastasis	TERT amplification/mutation	0.96	0.75	1.00	0.45	0.15	1.00	5	51	56
Metastasis	FGF3 amplification	0.94	0.41	1.00	0.53	0.11	1.00	3	53	56
Metastasis	FGF3 amplification/mutation	0.90	0.29	1.00	0.38	0.08	1.00	3	53	56
Metastasis	CDKN2B deletion	0.89	0.63	0.86	0.57	0.22	0.97	7	49	56
Metastasis	FGF4 amplification/mutation	0.88	0.31	1.00	0.43	0.09	1.00	3	53	56
Metastasis	FGF4 amplification	0.86	0.24	1.00	0.51	0.10	1.00	3	53	56
Metastasis	CDKN2B deletion/mutation	0.85	0.55	1.00	0.59	0.26	1.00	7	49	56
Metastasis	STK11 deletion/mutation	0.84	0.28	1.00	0.64	0.14	1.00	3	53	56
Metastasis	CCND1 amplification	0.83	0.25	1.00	0.31	0.10	1.00	4	52	56
Metastasis	KRAS oncogenic mutation	0.82	0.54	1.00	0.39	0.31	1.00	12	44	56
Metastasis	TGFBR2 deletion/mutation	0.82	0.17	1.00	0.64	0.14	1.00	3	53	56
Metastasis	CDKN2A (p14ARF) deletion	0.81	0.50	1.00	0.60	0.23	1.00	6	50	56
Metastasis	CTCF oncogenic mutation	0.81	0.43	1.00	0.55	0.11	1.00	3	53	56
Metastasis	TGFBR2 oncogenic mutation	0.81	0.21	1.00	0.58	0.12	1.00	3	53	56
Metastasis	ARID1A oncogenic mutation	0.80	0.56	1.00	0.22	0.24	1.00	11	45	56
Metastasis	CTCF deletion/mutation	0.80	0.43	1.00	0.57	0.12	1.00	3	53	56
Metastasis	KRAS amplification/mutation	0.80	0.49	1.00	0.30	0.28	1.00	12	44	56
Metastasis	CDKN2A (p16INK4a) deletion	0.79	0.44	0.83	0.60	0.20	0.97	6	50	56
Metastasis	TGF beta pathway	0.77	0.48	0.93	0.48	0.37	0.95	14	42	56
Metastasis	TMB-H	0.76	0.37	1.00	0.43	0.20	1.00	7	49	56

Table S3.8. Biomarkers detected in Melanoma

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	SPRED1 deletion	1.00	1.00	1.00	0.79	0.12	1.00	3	100	103
Primary	GNA11 oncogenic mutation	0.99	0.92	1.00	0.93	0.43	1.00	6	115	121
Primary	GNA11 amplification/mutation	0.99	0.86	1.00	0.96	0.55	1.00	6	115	121
Primary	CRLF2 deletion	0.97	0.55	1.00	0.66	0.07	1.00	3	118	121
Primary	ZRSR2 deletion	0.97	0.32	1.00	0.71	0.08	1.00	3	118	121
Primary	RAF1 amplification	0.95	0.24	1.00	0.47	0.05	1.00	3	118	121
Primary	AGO2 amplification	0.91	0.46	0.83	0.89	0.31	0.99	6	97	103
Primary	ARID1A oncogenic mutation	0.91	0.30	1.00	0.47	0.06	1.00	4	117	121
Primary	SPRED1 deletion/mutation	0.89	0.44	1.00	0.32	0.06	1.00	4	99	103
Primary	ARID1A deletion/mutation	0.88	0.25	1.00	0.44	0.06	1.00	4	117	121
Primary	KRAS oncogenic mutation	0.87	0.23	1.00	0.25	0.03	1.00	3	118	121
Primary	TERT oncogenic mutation	0.87	0.75	0.98	0.51	0.53	0.98	43	78	121
Primary	MDC1 amplification	0.87	0.15	1.00	0.44	0.06	1.00	4	117	121
Primary	DAXX amplification	0.87	0.15	1.00	0.58	0.08	1.00	4	117	121
Primary	KMT2A deletion/mutation	0.86	0.36	1.00	0.50	0.06	1.00	4	117	121
Primary	TERT amplification/mutation	0.85	0.73	0.95	0.51	0.53	0.95	44	77	121
Primary	BRAF oncogenic mutation	0.84	0.63	0.92	0.31	0.25	0.94	24	97	121
Primary	E2F3 amplification	0.82	0.13	1.00	0.47	0.05	1.00	3	118	121
Primary	BRAF amplification/mutation	0.82	0.62	0.92	0.36	0.26	0.95	24	97	121
Primary	KRAS amplification	0.82	0.12	1.00	0.64	0.07	1.00	3	118	121
Primary	CDKN2A (p16INK4a) oncogenic mutation	0.81	0.35	1.00	0.39	0.16	1.00	13	108	121
Primary	KMT2A oncogenic mutation	0.81	0.32	1.00	0.34	0.05	1.00	4	117	121
Primary	CDKN2B deletion	0.81	0.40	0.89	0.52	0.25	0.96	18	103	121
Primary	CDKN2A (p14ARF) deletion	0.81	0.44	0.96	0.49	0.32	0.98	24	97	121
Primary	ARID2 deletion/mutation	0.81	0.32	1.00	0.28	0.06	1.00	5	116	121
Primary	MGA deletion/mutation	0.81	0.08	1.00	0.32	0.04	1.00	3	118	121
Primary	RAF1 amplification/mutation	0.80	0.09	1.00	0.29	0.03	1.00	3	118	121
Primary	CDKN2A (p16INK4a) deletion	0.80	0.48	0.83	0.54	0.31	0.93	24	97	121
Primary	ARID2 oncogenic mutation	0.79	0.31	1.00	0.24	0.05	1.00	5	116	121
Primary	PAK1 amplification/mutation	0.78	0.11	1.00	0.32	0.05	1.00	4	117	121
Primary	NF1 oncogenic mutation	0.78	0.36	1.00	0.32	0.20	1.00	18	103	121
Primary	CDKN2B deletion/mutation	0.78	0.35	0.94	0.38	0.21	0.97	18	103	121
Primary	CCND1 amplification	0.77	0.11	1.00	0.52	0.08	1.00	5	116	121
Primary	CDKN2A (p14ARF) oncogenic mutation	0.76	0.13	0.83	0.31	0.06	0.97	6	115	121
Primary	FGF4 amplification	0.76	0.11	0.80	0.54	0.07	0.98	5	116	121
Primary	KRAS amplification/mutation	0.76	0.22	0.80	0.47	0.06	0.98	5	116	121
Metastasis	ELOC amplification	0.98	0.47	1.00	0.96	0.38	1.00	5	199	204
Metastasis	CTNNB1 oncogenic mutation	0.90	0.20	0.86	0.74	0.11	0.99	7	197	204
Metastasis	CTNNB1 amplification/mutation	0.89	0.21	1.00	0.71	0.11	1.00	7	197	204
Metastasis	PAK1 amplification/mutation	0.87	0.27	1.00	0.42	0.07	1.00	9	195	204
Metastasis	MGA deletion	0.83	0.09	1.00	0.43	0.04	1.00	5	199	204
Metastasis	RAD51 deletion/mutation	0.82	0.09	1.00	0.50	0.05	1.00	5	199	204
Metastasis	PAK1 amplification	0.82	0.23	1.00	0.44	0.08	1.00	9	195	204
Metastasis	FBXW7 deletion/mutation	0.81	0.22	0.89	0.29	0.05	0.98	9	195	204
Metastasis	TP53BP1 deletion	0.81	0.10	1.00	0.55	0.07	1.00	5	154	159
Metastasis	TERT oncogenic mutation	0.81	0.75	0.87	0.54	0.63	0.81	98	106	204
Metastasis	B2M oncogenic mutation	0.80	0.11	1.00	0.32	0.06	1.00	8	196	204
Metastasis	RTELI amplification	0.79	0.08	1.00	0.23	0.03	1.00	4	155	159
Metastasis	TERT amplification/mutation	0.79	0.78	0.87	0.49	0.65	0.77	107	97	204
Metastasis	TP53 oncogenic mutation	0.79	0.54	0.77	0.66	0.38	0.91	43	161	204
Metastasis	CDKN2A (p14ARF) deletion	0.75	0.52	0.95	0.28	0.33	0.93	55	149	204
Metastasis	CDKN2A (p16INK4a) deletion	0.75	0.07	1.00	0.64	0.08	1.00	5	154	159

Table S3.9. Biomarkers detected in Non-Small Cell Lung Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	SOX2 amplification	0.98	0.50	0.84	0.93	0.24	1.00	19	762	781
Primary	DCUN1D1 amplification	0.98	0.42	0.94	0.93	0.23	1.00	16	765	781
Primary	MET oncogenic mutation	0.90	0.19	0.94	0.47	0.04	1.00	17	764	781
Primary	KMT2D oncogenic mutation	0.89	0.22	0.95	0.66	0.07	1.00	20	761	781
Primary	KMT2D deletion/mutation	0.88	0.21	0.95	0.60	0.06	1.00	20	761	781
Primary	RB1 oncogenic mutation	0.87	0.27	0.94	0.57	0.09	1.00	33	748	781
Primary	STK11 deletion/mutation	0.87	0.53	0.79	0.74	0.28	0.97	86	695	781
Primary	STK11 oncogenic mutation	0.87	0.51	0.83	0.76	0.29	0.97	81	700	781
Primary	RB1 deletion/mutation	0.87	0.31	0.97	0.40	0.07	1.00	34	747	781
Primary	EGFR oncogenic mutation	0.87	0.69	0.93	0.52	0.37	0.96	185	596	781
Primary	PTEN deletion/mutation	0.86	0.12	0.81	0.73	0.06	0.99	16	765	781
Primary	KEAP1 oncogenic mutation	0.86	0.25	0.94	0.53	0.08	0.99	33	748	781
Primary	EGFR amplification/mutation	0.86	0.68	0.93	0.46	0.36	0.95	192	589	781
Primary	FGF4 amplification	0.85	0.18	0.94	0.61	0.05	1.00	17	764	781
Primary	FGF4 amplification/mutation	0.84	0.19	0.94	0.65	0.06	1.00	17	764	781
Primary	KEAP1 deletion/mutation	0.83	0.19	0.85	0.64	0.10	0.99	34	747	781
Primary	FGF19 amplification/mutation	0.83	0.16	0.89	0.66	0.06	1.00	18	763	781
Primary	FGF3 amplification/mutation	0.83	0.09	0.88	0.61	0.04	1.00	16	765	781
Primary	TP53 deletion/mutation	0.83	0.77	0.90	0.58	0.64	0.88	352	429	781
Primary	CCND1 amplification/mutation	0.83	0.15	0.95	0.56	0.05	1.00	19	762	781
Primary	TP53 oncogenic mutation	0.82	0.76	0.88	0.61	0.65	0.86	351	430	781
Primary	FGF19 amplification	0.82	0.15	0.89	0.64	0.06	1.00	18	763	781
Primary	FGF3 amplification	0.81	0.10	0.88	0.63	0.05	1.00	16	765	781
Primary	CDK4 amplification	0.81	0.09	0.96	0.45	0.05	1.00	24	757	781
Primary	BRCA1 germline/somatic mutation, deep deletion or fusion	0.81	0.25	1.00	0.23	0.06	1.00	3	60	63
Primary	CDKN1 amplification	0.80	0.16	0.95	0.58	0.05	1.00	19	762	781
Primary	MET amplification/mutation	0.80	0.23	0.85	0.42	0.05	0.99	26	755	781
Primary	CDK4 amplification/mutation	0.80	0.09	0.88	0.49	0.05	0.99	25	756	781
Primary	MTAP deletion	0.79	0.10	1.00	0.48	0.07	1.00	5	136	141
Primary	TMB-H	0.79	0.43	0.89	0.49	0.30	0.95	150	624	774
Primary	KRAS oncogenic mutation	0.78	0.60	0.96	0.30	0.36	0.94	229	552	781
Primary	FOXA1 amplification	0.78	0.18	0.76	0.68	0.05	0.99	17	764	781
Primary	KRAS amplification/mutation	0.78	0.59	0.94	0.30	0.36	0.93	232	549	781
Primary	ERBB2 amplification/mutation	0.78	0.13	0.87	0.46	0.06	0.99	30	751	781
Primary	CDKN2A (p14ARF) deletion	0.78	0.21	0.75	0.67	0.14	0.98	50	731	781
Primary	MDM2 amplification/mutation	0.77	0.13	0.97	0.39	0.07	1.00	34	747	781
Primary	FOXA1 amplification/mutation	0.77	0.10	0.76	0.67	0.05	0.99	17	764	781
Primary	MDM2 amplification	0.77	0.13	0.88	0.43	0.07	0.99	34	747	781
Primary	PIK3CA amplification/mutation	0.77	0.28	0.86	0.42	0.10	0.97	56	725	781
Primary	SMARCA4 deletion/mutation	0.77	0.14	0.77	0.55	0.05	0.99	22	759	781
Primary	CDKN2B deletion/mutation	0.77	0.18	0.78	0.66	0.13	0.98	49	732	781
Primary	SMARCA4 oncogenic mutation	0.77	0.10	0.82	0.52	0.05	0.99	22	759	781
Primary	ALK fusion	0.76	0.16	0.78	0.52	0.05	0.99	23	758	781
Primary	SETD2 deletion/mutation	0.76	0.08	0.95	0.20	0.03	0.99	19	762	781
Primary	RTK pathway	0.76	0.76	0.92	0.30	0.54	0.82	365	416	781
Primary	MTAP deletion/mutation	0.76	0.08	1.00	0.63	0.09	1.00	5	136	141
Metastasis	DCUN1D1 amplification	0.97	0.42	0.86	0.96	0.30	1.00	7	331	338
Metastasis	RAC1 amplification/mutation	0.88	0.20	1.00	0.26	0.04	1.00	9	329	338
Metastasis	RAC1 amplification	0.87	0.19	1.00	0.32	0.04	1.00	9	329	338
Metastasis	MCL1 amplification	0.86	0.32	0.89	0.81	0.11	1.00	9	329	338
Metastasis	STK11 fusion	0.84	0.23	0.88	0.63	0.05	1.00	8	330	338
Metastasis	RB1 deletion/mutation	0.83	0.34	1.00	0.22	0.09	1.00	23	315	338
Metastasis	KEAP1 deletion/mutation	0.82	0.28	0.86	0.60	0.16	0.98	28	310	338
Metastasis	RECQL4 amplification	0.82	0.13	0.91	0.37	0.05	0.99	11	327	338
Metastasis	CDKN2B deletion/mutation	0.81	0.33	0.95	0.37	0.17	0.98	41	297	338
Metastasis	KMT2D deletion/mutation	0.81	0.27	1.00	0.28	0.03	1.00	8	330	338
Metastasis	PRKCI amplification	0.80	0.27	1.00	0.32	0.07	1.00	12	245	257
Metastasis	CDKN2B deletion	0.80	0.33	0.93	0.39	0.17	0.97	41	297	338
Metastasis	AKT1 amplification	0.80	0.20	0.86	0.62	0.05	1.00	7	331	338
Metastasis	RB1 oncogenic mutation	0.80	0.27	0.95	0.35	0.08	0.99	20	318	338
Metastasis	KEAP1 oncogenic mutation	0.80	0.24	1.00	0.50	0.15	1.00	27	311	338
Metastasis	ILTR amplification	0.79	0.06	1.00	0.37	0.03	1.00	7	331	338
Metastasis	CDKN2A (p16INK4a) deletion	0.79	0.31	0.88	0.38	0.16	0.96	41	297	338
Metastasis	CDKN2A (p16INK4a) deletion	0.79	0.33	0.91	0.39	0.18	0.97	43	295	338
Metastasis	FAT1 deletion/mutation	0.79	0.09	0.89	0.44	0.04	0.99	9	329	338
Metastasis	AGO2 amplification	0.77	0.21	0.90	0.51	0.07	0.99	10	247	257
Metastasis	ERBB2 amplification/mutation	0.77	0.24	0.89	0.27	0.03	0.99	9	329	338
Metastasis	EGFR amplification/mutation	0.76	0.57	0.84	0.51	0.37	0.90	86	252	338

Table S3.10. Biomarkers detected in Ovarian Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	CTNNB1 oncogenic mutation	1.00	0.94	1.00	0.98	0.67	1.00	6	158	164
Primary	CTNNB1 amplification/mutation	0.99	0.87	1.00	0.97	0.55	1.00	6	158	164
Primary	TERT amplification	0.95	0.43	1.00	0.74	0.09	1.00	4	160	164
Primary	PIK3R1 oncogenic mutation	0.95	0.54	1.00	0.69	0.11	1.00	6	158	164
Primary	PIK3R1 deletion/mutation	0.95	0.41	1.00	0.58	0.08	1.00	6	158	164
Primary	TP53 oncogenic mutation	0.94	0.97	0.89	0.95	0.98	0.76	120	44	164
Primary	TP53 deletion/mutation	0.94	0.97	0.88	0.93	0.97	0.75	120	44	164
Primary	KRAS oncogenic mutation	0.93	0.71	0.94	0.67	0.25	0.99	17	147	164
Primary	ARID1A deletion/mutation	0.93	0.63	0.95	0.74	0.33	0.99	20	144	164
Primary	ARID1A oncogenic mutation	0.93	0.64	0.89	0.75	0.32	0.98	19	145	164
Primary	CDKN2B deletion/mutation	0.92	0.48	0.80	0.73	0.09	0.99	5	159	164
Primary	CDKN2B deletion	0.92	0.36	1.00	0.65	0.08	1.00	5	159	164
Primary	CDKN2A (p14ARF) deletion	0.91	0.45	0.80	0.72	0.08	0.99	5	159	164
Primary	CDKN2A (p16INK4a) deletion	0.91	0.48	0.80	0.69	0.08	0.99	5	159	164
Primary	NKK3-1 deletion/mutation	0.91	0.46	1.00	0.63	0.08	1.00	5	159	164
Primary	AKT2 amplification/mutation	0.90	0.25	1.00	0.44	0.06	1.00	6	158	164
Primary	PTEN oncogenic mutation	0.90	0.51	0.88	0.88	0.27	0.99	8	156	164
Primary	BRCA1 deletion/mutation	0.89	0.34	1.00	0.55	0.08	1.00	6	158	164
Primary	ARAF deletion	0.89	0.31	0.80	0.82	0.12	0.99	5	159	164
Primary	AKT2 amplification	0.89	0.23	1.00	0.53	0.07	1.00	6	158	164
Primary	KDM5A amplification	0.89	0.18	1.00	0.31	0.03	1.00	4	160	164
Primary	KDM5C deletion/mutation	0.89	0.30	1.00	0.75	0.13	1.00	6	158	164
Primary	BRCA1 germline/somatic mutation, deep deletion or fusion	0.88	0.32	1.00	0.50	0.16	1.00	11	116	127
Primary	PTEN deletion/mutation	0.88	0.50	0.88	0.72	0.14	0.99	8	156	164
Primary	CDK12 amplification	0.87	0.12	1.00	0.64	0.07	1.00	4	160	164
Primary	ERBB2 amplification/mutation	0.87	0.29	0.82	0.78	0.21	0.98	11	153	164
Primary	PIK3CA oncogenic mutation	0.86	0.39	1.00	0.39	0.17	1.00	18	146	164
Primary	BRCA1/2 carriers	0.85	0.17	1.00	0.70	0.13	1.00	7	151	158
Primary	Fragment of genome altered ≥ 30%	0.85	0.97	0.95	0.45	0.93	0.56	78	11	89
Primary	BCOR deletion/mutation	0.85	0.32	0.83	0.66	0.09	0.99	6	158	164
Primary	CDKN1B amplification	0.85	0.20	1.00	0.35	0.05	1.00	5	159	164
Primary	PPP2R1A oncogenic mutation	0.85	0.30	1.00	0.29	0.05	1.00	6	158	164
Primary	ETV6 amplification	0.84	0.17	1.00	0.26	0.03	1.00	4	160	164
Primary	AGO2 amplification	0.84	0.15	1.00	0.62	0.11	1.00	7	147	154
Primary	Genome instability ≥ 20%	0.84	0.95	0.90	0.62	0.92	0.59	73	16	89
Primary	KDM5A amplification/mutation	0.84	0.19	1.00	0.33	0.04	1.00	4	160	164
Primary	KRAS amplification/mutation	0.83	0.67	0.77	0.63	0.24	0.95	22	142	164
Primary	TERT amplification/mutation	0.83	0.28	0.89	0.38	0.08	0.98	9	155	164
Primary	PPP2R1A deletion/mutation	0.82	0.16	1.00	0.33	0.05	1.00	6	158	164
Primary	TERT oncogenic mutation	0.81	0.16	0.80	0.37	0.04	0.98	5	159	164
Primary	RAD52 amplification	0.81	0.16	1.00	0.34	0.04	1.00	4	160	164
Primary	KDM5C deletion	0.81	0.27	0.80	0.72	0.08	0.99	5	159	164
Primary	BRCA1/2 somatic/germline mutant	0.81	0.38	0.95	0.50	0.29	0.98	22	104	126
Primary	PIK3CA amplification/mutation	0.80	0.35	0.89	0.47	0.18	0.97	19	145	164
Primary	BRCA1 oncogenic mutation	0.80	0.08	1.00	0.60	0.06	1.00	4	160	164
Primary	MYC amplification/mutation	0.79	0.21	0.91	0.46	0.11	0.99	11	153	164
Primary	mTOR pathway	0.78	0.65	0.89	0.38	0.36	0.90	46	118	164
Primary	RECQL4 amplification	0.78	0.13	0.86	0.46	0.07	0.99	7	157	164
Primary	Tetraploidy	0.78	0.85	0.93	0.45	0.76	0.76	54	29	83
Primary	KMT2B amplification	0.78	0.09	1.00	0.52	0.05	1.00	4	150	154
Primary	IGF1R amplification	0.77	0.07	1.00	0.39	0.04	1.00	4	160	164
Primary	KRAS amplification	0.77	0.12	1.00	0.22	0.05	1.00	7	157	164
Metastasis	TP53 deletion/mutation	0.97	0.99	0.91	0.91	0.98	0.67	172	35	207
Metastasis	TP53 oncogenic mutation	0.97	0.99	0.91	0.86	0.97	0.65	172	35	207
Metastasis	KRAS oncogenic mutation	0.93	0.48	1.00	0.60	0.15	1.00	14	193	207
Metastasis	Fragment of genome altered ≥ 30%	0.91	0.99	0.91	0.75	0.98	0.41	147	12	159
Metastasis	Genome instability ≥ 20%	0.91	0.99	0.90	0.64	0.94	0.50	137	22	159
Metastasis	MYCL amplification/mutation	0.88	0.14	0.83	0.78	0.10	0.99	6	201	207
Metastasis	PRDM14 amplification	0.87	0.31	0.83	0.68	0.08	0.99	6	185	191
Metastasis	STK11 deletion/mutation	0.86	0.13	1.00	0.31	0.03	1.00	5	202	207
Metastasis	ELOC amplification	0.85	0.17	0.83	0.75	0.09	0.99	6	201	207
Metastasis	RECQL4 amplification	0.83	0.42	1.00	0.46	0.20	1.00	24	183	207
Metastasis	BRCA1 germline/somatic point mutation/deletion or fusion	0.83	0.44	1.00	0.42	0.23	1.00	21	122	143
Metastasis	AGO2 amplification	0.83	0.43	0.82	0.62	0.22	0.96	22	169	191
Metastasis	CDKN2B deletion/mutation	0.83	0.24	0.90	0.59	0.10	0.99	10	197	207
Metastasis	CDKN2A (p16INK4a) deletion	0.83	0.26	0.90	0.53	0.09	0.99	10	197	207
Metastasis	CDKN2A (p14ARF) deletion	0.82	0.26	0.90	0.54	0.09	0.99	10	197	207
Metastasis	PIK3R1 deletion/mutation	0.82	0.44	0.80	0.71	0.06	0.99	5	202	207
Metastasis	PIK3CA oncogenic mutation	0.81	0.34	0.94	0.41	0.12	0.99	16	191	207
Metastasis	CDKN2B deletion	0.81	0.23	0.90	0.47	0.08	0.99	10	197	207
Metastasis	TCF3 deletion	0.79	0.18	0.80	0.80	0.09	0.99	5	202	207
Metastasis	RTEL1 amplification	0.79	0.10	0.83	0.69	0.08	0.99	6	185	191
Metastasis	PIK3CA amplification/mutation	0.78	0.34	0.78	0.58	0.15	0.96	18	189	207
Metastasis	PPP2R1A oncogenic mutation	0.78	0.23	1.00	0.31	0.06	1.00	9	198	207
Metastasis	KDM5A amplification/mutation	0.76	0.43	1.00	0.33	0.04	1.00	5	202	207
Metastasis	BRCA1/2 germline/somatic point mutation/deletion or fusion	0.76	0.49	0.92	0.28	0.29	0.91	36	111	147
Metastasis	KDM5A amplification	0.76	0.44	0.80	0.31	0.03	0.98	5	202	207
Metastasis	BCL2L1 amplification	0.76	0.11	0.80	0.69	0.06	0.99	5	202	207
Metastasis	RAD21 amplification	0.75	0.21	0.93	0.49	0.12	0.99	14	193	207

Table S3.11. Biomarkers detected in Pancreatic Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	MEN1 oncogenic mutation	0.97	0.58	0.88	0.92	0.28	1.00	8	239	247
Primary	MEN1 deletion/mutation	0.97	0.55	0.88	0.92	0.27	1.00	8	239	247
Primary	GNAS oncogenic mutation	0.96	0.66	1.00	0.73	0.14	1.00	10	237	247
Primary	GNAS amplification/mutation	0.96	0.62	1.00	0.32	0.06	1.00	10	237	247
Primary	CDKN2B deletion/mutation	0.90	0.34	0.89	0.73	0.22	0.99	19	228	247
Primary	CDKN2A (p14ARF) deletion	0.88	0.30	0.85	0.81	0.28	0.98	20	227	247
Primary	KRAS oncogenic mutation	0.86	0.94	0.90	0.66	0.90	0.65	191	56	247
Primary	KRAS amplification/mutation	0.86	0.94	0.90	0.68	0.90	0.66	191	56	247
Primary	PIK3CA oncogenic mutation	0.84	0.12	0.83	0.53	0.04	0.99	6	241	247
Primary	TP53 oncogenic mutation	0.84	0.87	0.93	0.53	0.76	0.82	153	94	247
Primary	TP53 deletion/mutation	0.84	0.87	0.93	0.54	0.77	0.82	153	94	247
Primary	PIK3CA amplification/mutation	0.82	0.09	1.00	0.36	0.04	1.00	6	241	247
Primary	BRCA1/2 germline/somatic mutation/deletion or fusion	0.77	0.17	1.00	0.27	0.08	1.00	11	179	190
Primary	BRCA2 germline/somatic mutation/deletion or fusion	0.76	0.15	1.00	0.25	0.06	1.00	9	182	191
Metastasis	DAXX deletion/mutation	0.98	0.56	0.83	0.96	0.42	0.99	6	197	203
Metastasis	MEN1 oncogenic mutation	0.96	0.40	0.89	0.94	0.42	0.99	9	194	203
Metastasis	MEN1 deletion/mutation	0.96	0.38	0.89	0.93	0.38	0.99	9	194	203
Metastasis	KRAS oncogenic mutation	0.95	0.98	0.84	0.88	0.96	0.60	160	43	203
Metastasis	KRAS amplification/mutation	0.94	0.98	0.82	0.88	0.96	0.56	161	42	203
Metastasis	STK11 deletion/mutation	0.92	0.63	1.00	0.39	0.06	1.00	7	196	203
Metastasis	CDK6 amplification/mutation	0.92	0.19	1.00	0.56	0.06	1.00	6	197	203
Metastasis	CTNNB1 amplification/mutation	0.90	0.18	1.00	0.72	0.07	1.00	4	199	203
Metastasis	STK11 oncogenic mutation	0.89	0.43	1.00	0.48	0.06	1.00	6	197	203
Metastasis	CDK6 amplification	0.89	0.17	1.00	0.66	0.08	1.00	6	197	203
Metastasis	CTNNB1 oncogenic mutation	0.85	0.16	1.00	0.35	0.03	1.00	4	199	203
Metastasis	TP53 deletion/mutation	0.83	0.86	0.88	0.64	0.81	0.75	128	75	203
Metastasis	TP53 oncogenic mutation	0.83	0.86	0.87	0.67	0.82	0.75	128	75	203
Metastasis	TEK deletion	0.82	0.09	1.00	0.43	0.05	1.00	5	162	167
Metastasis	CDKN2A (p14ARF) deletion	0.78	0.39	0.84	0.54	0.33	0.92	43	160	203
Metastasis	CDKN2A (p16INK4a) deletion	0.77	0.41	0.87	0.54	0.35	0.93	45	158	203
Metastasis	GNAS amplification/mutation	0.77	0.25	0.80	0.56	0.04	0.99	5	198	203
Metastasis	CDKN2B deletion/mutation	0.77	0.37	0.90	0.50	0.31	0.95	40	163	203
Metastasis	KDM6A deletion/mutation	0.76	0.19	1.00	0.35	0.09	1.00	13	190	203
Metastasis	SMAD4 deletion	0.76	0.21	0.83	0.51	0.10	0.98	12	191	203
Metastasis	CDKN2B deletion	0.75	0.35	0.88	0.55	0.32	0.95	40	163	203

Table S3.12. Biomarkers detected in Prostate Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	ERF deletion/mutation	0.89	0.24	1.00	0.63	0.07	1.00	7	266	273
Primary	TMB-H	0.81	0.29	0.88	0.50	0.04	0.99	8	318	326
Primary	APC oncogenic mutation	0.81	0.16	0.93	0.50	0.08	0.99	14	318	332
Primary	APC deletion/mutation	0.79	0.15	0.87	0.38	0.06	0.98	15	317	332
Primary	PTEN deletion/mutation	0.78	0.23	0.91	0.52	0.17	0.98	33	299	332
Primary	SPOP oncogenic mutation	0.78	0.42	0.96	0.26	0.17	0.97	45	287	332
Primary	SPOP deletion/mutation	0.77	0.39	0.91	0.37	0.18	0.96	45	287	332
Primary	PTEN oncogenic mutation	0.76	0.09	0.92	0.47	0.07	0.99	13	319	332
Primary	TP53 deletion/mutation	0.76	0.55	0.83	0.47	0.34	0.89	83	249	332
Metastasis	DIS3 deletion/mutation	0.85	0.16	1.00	0.56	0.09	1.00	10	225	235
Metastasis	RBI deletion	0.85	0.20	0.85	0.79	0.19	0.99	13	222	235
Metastasis	CYSLTR2 deletion	0.84	0.13	0.86	0.67	0.10	0.99	7	169	176
Metastasis	RPS6KB2 amplification	0.82	0.11	1.00	0.62	0.05	1.00	5	230	235
Metastasis	CDKN2B deletion	0.82	0.25	1.00	0.20	0.04	1.00	7	228	235
Metastasis	ANKRD11 deletion/mutation	0.79	0.28	0.91	0.39	0.07	0.99	11	224	235
Metastasis	FYN deletion	0.79	0.14	0.78	0.73	0.10	0.99	9	226	235
Metastasis	BMPR1A deletion/mutation	0.79	0.10	1.00	0.22	0.04	1.00	7	228	235
Metastasis	PLCG2 deletion	0.78	0.21	0.92	0.36	0.08	0.99	13	222	235
Metastasis	DIS3 deletion	0.78	0.10	1.00	0.36	0.06	1.00	10	225	235
Metastasis	CDKN1B deletion	0.78	0.14	1.00	0.28	0.04	1.00	7	228	235
Metastasis	ETV6 deletion/mutation	0.77	0.12	1.00	0.33	0.03	1.00	5	230	235
Metastasis	ETV6 deletion	0.77	0.09	1.00	0.27	0.03	1.00	5	230	235
Metastasis	NKX3-1 deletion/mutation	0.76	0.20	0.85	0.58	0.11	0.98	13	222	235
Metastasis	CCND1 amplification	0.76	0.13	1.00	0.27	0.06	1.00	10	225	235
Metastasis	APC oncogenic mutation	0.76	0.17	0.86	0.44	0.09	0.98	14	221	235
Metastasis	PRDM14 amplification	0.75	0.19	1.00	0.24	0.11	1.00	15	161	176

Table S3.13. Biomarkers detected in Renal Cell Carcinoma

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	VEGFA amplification	0.98	0.62	1.00	0.93	0.25	1.00	3	135	138
Primary	VEGFA amplification/mutation	0.98	0.57	1.00	0.92	0.21	1.00	3	135	138
Primary	CCND3 amplification	0.97	0.42	1.00	0.94	0.27	1.00	3	135	138
Primary	TSC2 oncogenic mutation	0.94	0.42	1.00	0.78	0.12	1.00	4	134	138
Primary	NF2 deletion/mutation	0.94	0.30	1.00	0.60	0.07	1.00	4	134	138
Primary	NF2 oncogenic mutation	0.93	0.30	1.00	0.73	0.10	1.00	4	134	138
Primary	BAP1 deletion/mutation	0.90	0.45	0.83	0.83	0.31	0.98	12	126	138
Primary	PTEN deletion/mutation	0.89	0.16	1.00	0.77	0.11	1.00	4	134	138
Primary	PTEN oncogenic mutation	0.89	0.17	1.00	0.70	0.09	1.00	4	134	138
Primary	TP53 deletion/mutation	0.88	0.40	0.89	0.58	0.13	0.99	9	129	138
Primary	TP53 oncogenic mutation	0.88	0.41	1.00	0.55	0.13	1.00	9	129	138
Primary	TERT oncogenic mutation	0.87	0.33	1.00	0.53	0.10	1.00	7	131	138
Primary	VHL deletion/mutation	0.85	0.74	0.96	0.63	0.60	0.96	51	87	138
Primary	VHL oncogenic mutation	0.84	0.66	0.96	0.64	0.61	0.97	51	87	138
Primary	TERT amplification/mutation	0.82	0.29	1.00	0.54	0.12	1.00	8	130	138
Primary	PIK3CA amplification/mutation	0.80	0.33	1.00	0.30	0.04	1.00	4	134	138
Primary	PIK3CA oncogenic mutation	0.80	0.31	1.00	0.43	0.05	1.00	4	134	138
Primary	PBRM1 deletion/mutation	0.80	0.40	0.81	0.61	0.27	0.95	21	117	138
Primary	mTOR pathway	0.79	0.46	0.78	0.63	0.30	0.94	23	115	138
Metastasis	CDKN2B deletion/mutation	0.89	0.33	1.00	0.77	0.25	1.00	4	52	56
Metastasis	VHL oncogenic mutation	0.86	0.87	0.97	0.58	0.72	0.94	30	26	56
Metastasis	TERT amplification/mutation	0.86	0.26	1.00	0.40	0.11	1.00	4	52	56
Metastasis	TERT oncogenic mutation	0.86	0.47	1.00	0.42	0.09	1.00	3	53	56
Metastasis	VHL deletion/mutation	0.86	0.85	0.93	0.58	0.72	0.88	30	26	56

Table S3.14. Biomarkers detected in Thyroid Cancer

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	CDKN2A (p14ARF) deletion	0.93	0.46	1.00	0.64	0.17	1.00	4	56	60
Primary	TP53 deletion/mutation	0.91	0.78	0.86	0.94	0.67	0.98	7	53	60
Primary	CDKN2A (p16INK4a) deletion	0.90	0.54	1.00	0.73	0.21	1.00	4	56	60
Primary	BRAF amplification/mutation	0.89	0.93	0.83	0.80	0.81	0.83	30	30	60
Primary	BRAF oncogenic mutation	0.89	0.92	0.83	0.77	0.78	0.82	30	30	60
Primary	TP53 oncogenic mutation	0.89	0.76	0.86	0.94	0.67	0.98	7	53	60
Primary	NRAS amplification/mutation	0.87	0.29	1.00	0.72	0.16	1.00	3	57	60
Primary	NRAS oncogenic mutation	0.86	0.23	1.00	0.74	0.17	1.00	3	57	60
Primary	HRAS amplification/mutation	0.84	0.44	1.00	0.48	0.18	1.00	6	54	60
Primary	TERT oncogenic mutation	0.83	0.69	0.82	0.68	0.60	0.87	22	38	60
Primary	RTK pathway	0.83	0.48	0.88	0.25	0.15	0.93	8	52	60
Primary	TERT amplification/mutation	0.81	0.67	0.87	0.62	0.59	0.88	23	37	60
Metastasis	BRAF oncogenic mutation	0.96	0.95	0.94	0.78	0.76	0.94	31	40	71
Metastasis	BRAF amplification/mutation	0.96	0.95	0.94	0.85	0.83	0.94	31	40	71
Metastasis	RTK pathway	0.88	0.70	1.00	0.23	0.24	1.00	14	57	71
Metastasis	TERT amplification/mutation	0.87	0.78	0.93	0.67	0.65	0.94	28	43	71
Metastasis	HRAS amplification/mutation	0.85	0.31	1.00	0.43	0.10	1.00	4	67	71

Table S3.15. Biomarkers detected in Soft Tissue Sarcoma

Tumor Site	Biomarker	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
Primary	TERT oncogenic mutation	0.91	0.51	1.00	0.29	0.11	1.00	22	254	276
Primary	CARM1 amplification	0.91	0.20	0.80	0.86	0.11	0.99	5	221	226
Primary	CYSLTR2 deletion	0.88	0.17	0.89	0.79	0.15	0.99	9	217	226
Primary	ATRX deletion/mutation	0.88	0.21	0.92	0.64	0.10	0.99	12	264	276
Primary	JUN amplification/mutation	0.88	0.48	1.00	0.62	0.06	1.00	7	269	276
Primary	CDKN2A (p16INK4a) deletion	0.88	0.39	1.00	0.33	0.13	1.00	26	250	276
Primary	TP53 deletion	0.88	0.22	1.00	0.64	0.13	1.00	14	262	276
Primary	TMB-H	0.87	0.11	1.00	0.31	0.04	1.00	8	268	276
Primary	RB1 deletion/mutation	0.87	0.32	0.91	0.69	0.20	0.99	22	254	276
Primary	JUN amplification	0.87	0.44	1.00	0.58	0.06	1.00	7	269	276
Primary	RB1 deletion	0.86	0.17	0.83	0.78	0.14	0.99	12	264	276
Primary	CDKN2A (p14ARF) deletion	0.85	0.34	1.00	0.40	0.15	1.00	27	249	276
Primary	MTAP deletion/mutation	0.85	0.30	1.00	0.47	0.14	1.00	6	70	76
Primary	TP53 deletion/mutation	0.85	0.61	0.89	0.62	0.42	0.95	65	211	276
Primary	MDM2 amplification	0.84	0.70	0.84	0.59	0.31	0.94	50	226	276
Primary	MDM2 amplification/mutation	0.84	0.67	0.88	0.55	0.30	0.95	50	226	276
Primary	CDKN2B deletion	0.84	0.33	0.96	0.52	0.17	0.99	25	251	276
Primary	CDKN2B deletion/mutation	0.84	0.30	1.00	0.46	0.16	1.00	25	251	276
Primary	MTAP deletion	0.84	0.29	1.00	0.43	0.13	1.00	6	70	76
Primary	CTNNB1 oncogenic mutation	0.82	0.53	0.83	0.59	0.04	0.99	6	270	276
Primary	KDR amplification/mutation	0.81	0.27	0.83	0.41	0.03	0.99	6	270	276
Primary	TP53 oncogenic mutation	0.81	0.48	0.88	0.59	0.34	0.96	52	224	276
Primary	ATRX oncogenic mutation	0.81	0.09	0.88	0.61	0.06	0.99	8	268	276
Primary	KIT amplification/mutation	0.80	0.09	0.83	0.51	0.04	0.99	6	270	276
Primary	RB1 oncogenic mutation	0.80	0.13	0.90	0.60	0.08	0.99	10	266	276
Primary	ETV1 amplification/mutation	0.80	0.19	0.83	0.49	0.03	0.99	6	270	276
Primary	CTNNB1 amplification/mutation	0.80	0.53	0.83	0.56	0.04	0.99	6	270	276
Primary	CDK4 amplification	0.80	0.60	0.80	0.60	0.30	0.93	49	227	276
Primary	CDK4 amplification/mutation	0.80	0.60	0.78	0.60	0.30	0.93	49	227	276
Primary	CCND3 amplification	0.78	0.08	0.83	0.51	0.04	0.99	6	270	276
Primary	CCND3 amplification/mutation	0.78	0.06	0.83	0.51	0.04	0.99	6	270	276
Primary	FLCN amplification	0.77	0.15	0.83	0.71	0.06	0.99	6	270	276
Primary	DNA damage response	0.77	0.63	0.84	0.54	0.42	0.90	77	199	276
Primary	TERT amplification/mutation	0.76	0.39	0.90	0.23	0.12	0.95	29	247	276
Metastasis	WT1 fusion	0.99	0.82	0.83	0.98	0.71	0.99	6	83	89
Metastasis	ATRX deletion/mutation	0.91	0.57	0.90	0.67	0.26	0.98	10	79	89
Metastasis	TP53 deletion	0.91	0.55	1.00	0.62	0.21	1.00	8	81	89
Metastasis	ATRX deletion	0.90	0.41	1.00	0.85	0.24	1.00	4	85	89
Metastasis	RB1 oncogenic mutation	0.87	0.16	1.00	0.53	0.07	1.00	3	86	89
Metastasis	RB1 deletion/mutation	0.86	0.35	1.00	0.63	0.19	1.00	7	82	89
Metastasis	RB1 deletion	0.85	0.37	1.00	0.73	0.15	1.00	4	85	89
Metastasis	ATRX oncogenic mutation	0.84	0.36	1.00	0.64	0.17	1.00	6	83	89
Metastasis	TP53 deletion/mutation	0.83	0.68	0.86	0.64	0.52	0.91	28	61	89
Metastasis	CDKN2A (p16INK4a) deletion	0.81	0.24	1.00	0.31	0.10	1.00	6	83	89
Metastasis	HRD pathway	0.81	0.15	1.00	0.64	0.09	1.00	3	86	89
Metastasis	NF1 deletion/mutation	0.80	0.44	1.00	0.54	0.16	1.00	7	82	89
Metastasis	CDKN2B deletion	0.79	0.21	1.00	0.45	0.12	1.00	6	83	89
Metastasis	CDKN2B deletion/mutation	0.79	0.22	1.00	0.40	0.11	1.00	6	83	89
Metastasis	TP53 oncogenic mutation	0.77	0.45	0.90	0.57	0.40	0.95	21	68	89
Metastasis	MDM2 amplification	0.76	0.45	0.80	0.65	0.22	0.96	10	79	89
Metastasis	CDKN2A (p14ARF) deletion	0.75	0.28	1.00	0.35	0.10	1.00	6	83	89

Table S4: Performance Metrics for Biomarker Prediction on the TCGA validation cohort. Metrics include Area Under the Curve (AUC), Average Precision (AP), Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV). Sample counts are provided for each biomarker, indicating the number of positive, negative, and total samples. The performance of biomarkers in 18 TCGA projects that are associated with the 15 most common cancer types are summarized in the following sub-tables S4.1 - S4.18. Results for TCGA-CHOL, TCGA-KIRC, and TCGA-LUSC did not meet the baseline inclusion criteria of $AUC \geq 0.75$ and are therefore not included in this table.

Table S4.1. Biomarkers detected in TCGA-BLCA

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
FGFR3 amplification/mutation	FGFR3 oncogenic mutation	0.87	0.63	0.82	0.73	0.32	0.96	50	324	374
FGFR3 oncogenic mutation	FGFR3 oncogenic mutation	0.87	0.61	0.76	0.81	0.38	0.96	50	324	374
AKT2 amplification/mutation	AKT2 amplification	0.77	0.11	0.90	0.36	0.04	0.99	10	370	380
TBX3 amplification	TBX3 amplification	0.76	0.07	0.89	0.37	0.03	0.99	9	371	380

Table S4.2. Biomarkers detected in TCGA-BRCA

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
CDH1 oncogenic mutation	CDH1 oncogenic mutation	0.94	0.56	0.95	0.82	0.40	0.99	106	837	943
CDH1 deletion/mutation	CDH1 oncogenic mutation	0.94	0.56	0.96	0.83	0.41	0.99	106	837	943
CDH1 oncogenic mutation	CDH1 mutation	0.92	0.57	0.93	0.82	0.42	0.99	120	837	957
ETV6 amplification	ETV6 amplification	0.89	0.15	0.91	0.76	0.08	1.00	23	930	953
TP53 deletion/mutation	TP53 oncogenic mutation	0.88	0.82	0.92	0.55	0.52	0.93	333	622	955
TP53 oncogenic mutation	TP53 oncogenic mutation	0.88	0.82	0.92	0.56	0.53	0.93	333	622	955
TP53 oncogenic mutation	TP53 mutation	0.88	0.82	0.92	0.56	0.53	0.93	335	622	957
KDM5A amplification	KDM5A amplification	0.86	0.13	0.77	0.72	0.07	0.99	26	930	956
RAD52 amplification	RAD52 amplification	0.85	0.12	0.81	0.73	0.08	0.99	26	930	956
GATA3 amplification	GATA3 amplification	0.85	0.22	0.90	0.54	0.06	0.99	30	927	957
KDM5A amplification/mutation	KDM5A amplification	0.84	0.12	0.77	0.77	0.09	0.99	26	930	956
ERBB2 amplification	ERBB2 amplification	0.81	0.40	0.95	0.39	0.17	0.98	113	844	957
MAP3K1 oncogenic mutation	MAP3K1 oncogenic mutation	0.79	0.24	0.97	0.28	0.09	0.99	67	874	941
CDK12 amplification	CDK12 amplification	0.78	0.34	0.87	0.41	0.15	0.96	102	854	956
MAP3K1 deletion/mutation	MAP3K1 oncogenic mutation	0.78	0.21	0.99	0.27	0.09	1.00	67	874	941
ETV6 fusion	ETV6 alteration	0.77	0.10	0.92	0.30	0.05	0.99	36	921	957
ERBB2 amplification/mutation	ERBB2 amplification	0.77	0.36	0.86	0.45	0.17	0.96	113	844	957
FOXA1 deletion/mutation	FOXA1 oncogenic mutation	0.77	0.10	0.96	0.23	0.03	1.00	23	926	949
FOXA1 oncogenic mutation	FOXA1 oncogenic mutation	0.77	0.09	0.96	0.30	0.03	1.00	23	926	949

Table S4.3. Biomarkers detected in TCGA-COAD

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
SPEN oncogenic mutation	SPEN oncogenic mutation	0.96	0.52	0.78	0.91	0.20	0.99	9	321	330
SPEN deletion/mutation	SPEN oncogenic mutation	0.96	0.56	0.78	0.89	0.17	0.99	9	321	330
HNF1A oncogenic mutation	HNF1A oncogenic mutation	0.92	0.21	0.86	0.85	0.11	1.00	7	336	343
KMT2D oncogenic mutation	KMT2D oncogenic mutation	0.92	0.43	1.00	0.48	0.09	1.00	15	300	315
KMT2D deletion/mutation	KMT2D oncogenic mutation	0.91	0.39	1.00	0.57	0.10	1.00	15	300	315
RNF43 oncogenic mutation	RNF43 oncogenic mutation	0.90	0.37	1.00	0.40	0.09	1.00	18	313	331
RNF43 deletion/mutation	RNF43 oncogenic mutation	0.89	0.35	0.94	0.49	0.10	0.99	18	313	331
INPP4B oncogenic mutation	INPP4B mutation	0.88	0.13	0.78	0.84	0.11	0.99	9	339	348
BCOR deletion/mutation	BCOR oncogenic mutation	0.88	0.18	1.00	0.22	0.04	1.00	10	327	337
NOTCH2 oncogenic mutation	NOTCH2 mutation	0.88	0.25	1.00	0.36	0.06	1.00	14	334	348
CDH1 oncogenic mutation	CDH1 mutation	0.88	0.38	0.83	0.71	0.09	0.99	12	336	348
NF2 oncogenic mutation	NF2 mutation	0.88	0.23	1.00	0.44	0.04	1.00	8	340	348
SMARCB1 oncogenic mutation	SMARCB1 mutation	0.87	0.21	1.00	0.22	0.04	1.00	10	338	348
BCOR oncogenic mutation	BCOR oncogenic mutation	0.87	0.22	0.90	0.44	0.05	0.99	10	327	337
FAT1 oncogenic mutation	FAT1 mutation	0.86	0.49	0.93	0.52	0.20	0.98	40	308	348
MGA oncogenic mutation	MGA mutation	0.85	0.32	0.92	0.57	0.14	0.99	24	324	348
BRCA2 oncogenic mutation	BRCA2 oncogenic mutation	0.85	0.23	0.91	0.45	0.05	0.99	11	324	335
RNF43 oncogenic mutation	RNF43 mutation	0.84	0.45	0.94	0.40	0.15	0.98	35	313	348
BRAF oncogenic mutation	BRAF oncogenic mutation	0.84	0.51	0.93	0.43	0.19	0.98	43	301	344
PTCH1 oncogenic mutation	PTCH1 mutation	0.84	0.20	0.86	0.76	0.19	0.99	21	327	348
BRAF amplification/mutation	BRAF oncogenic mutation	0.84	0.49	0.98	0.23	0.15	0.99	43	301	344
FGFR3 oncogenic mutation	FGFR3 mutation	0.84	0.19	0.88	0.63	0.11	0.99	17	331	348
ANKRD11 oncogenic mutation	ANKRD11 mutation	0.83	0.39	0.76	0.79	0.27	0.97	33	315	348
ERF oncogenic mutation	ERF mutation	0.83	0.21	0.78	0.75	0.08	0.99	9	339	348
KMT2D oncogenic mutation	KMT2D mutation	0.83	0.51	0.90	0.48	0.22	0.97	48	300	348
MGA oncogenic mutation	MGA oncogenic mutation	0.83	0.15	0.92	0.57	0.07	0.99	12	324	336
ARID2 oncogenic mutation	ARID2 oncogenic mutation	0.83	0.14	0.91	0.62	0.08	1.00	11	322	333
BRCA2 deletion/mutation	BRCA2 oncogenic mutation	0.82	0.14	0.91	0.62	0.08	1.00	11	324	335
PARP1 oncogenic mutation	PARP1 mutation	0.82	0.25	0.91	0.25	0.04	0.99	11	337	348
TP53 oncogenic mutation	TP53 oncogenic mutation	0.82	0.83	0.89	0.56	0.71	0.81	191	156	347
NOTCH4 oncogenic mutation	NOTCH4 mutation	0.82	0.17	0.92	0.52	0.07	0.99	13	335	348
TP53 oncogenic mutation	TP53 mutation	0.82	0.83	0.89	0.56	0.72	0.81	192	156	348
ARID2 deletion/mutation	ARID2 oncogenic mutation	0.82	0.14	0.91	0.66	0.08	1.00	11	322	333
BCOR oncogenic mutation	BCOR mutation	0.82	0.23	0.95	0.44	0.10	0.99	21	327	348
TP53 deletion/mutation	TP53 oncogenic mutation	0.81	0.83	0.88	0.55	0.71	0.79	191	156	347
ARID1B oncogenic mutation	ARID1B mutation	0.81	0.23	0.92	0.32	0.10	0.98	26	322	348
NF1 oncogenic mutation	NF1 mutation	0.81	0.24	0.95	0.31	0.09	0.99	22	326	348
B2M deletion/mutation	B2M oncogenic mutation	0.81	0.16	0.93	0.44	0.07	0.99	15	331	346
BRAF oncogenic mutation	BRAF mutation	0.81	0.50	0.89	0.43	0.20	0.96	47	301	348
DAXX oncogenic mutation	DAXX mutation	0.80	0.07	0.86	0.55	0.04	0.99	7	341	348
MGA deletion/mutation	MGA oncogenic mutation	0.80	0.15	0.92	0.49	0.06	0.99	12	324	336
DNMT3B amplification	DNMT3B amplification	0.80	0.20	0.86	0.63	0.13	0.99	21	327	348
NCOA3 amplification	NCOA3 amplification	0.79	0.14	0.81	0.64	0.10	0.99	16	332	348
AXIN2 oncogenic mutation	AXIN2 mutation	0.79	0.25	0.91	0.35	0.09	0.98	23	325	348
CIC deletion/mutation	CIC oncogenic mutation	0.79	0.36	0.85	0.29	0.05	0.98	13	321	334
DICER1 oncogenic mutation	DICER1 mutation	0.78	0.14	0.94	0.31	0.07	0.99	17	331	348
ARID1B oncogenic mutation	ARID1B oncogenic mutation	0.78	0.08	1.00	0.32	0.04	1.00	10	322	332
RARA amplification	RARA amplification	0.78	0.18	1.00	0.37	0.04	1.00	8	340	348
SRC amplification	SRC amplification	0.78	0.12	0.87	0.65	0.10	0.99	15	332	347
RASA1 oncogenic mutation	RASA1 mutation	0.77	0.15	0.80	0.70	0.11	0.99	15	333	348
ASXL1 amplification	ASXL1 amplification	0.77	0.19	0.83	0.65	0.15	0.98	24	324	348
SETD2 oncogenic mutation	SETD2 oncogenic mutation	0.77	0.07	0.88	0.44	0.04	0.99	8	323	331
NSD3 amplification	NSD3 amplification	0.77	0.18	0.81	0.51	0.07	0.98	16	329	345
ARID2 oncogenic mutation	ARID2 mutation	0.76	0.25	0.81	0.62	0.15	0.98	26	322	348
PAX5 oncogenic mutation	PAX5 mutation	0.76	0.13	0.82	0.55	0.06	0.99	11	337	348
BLM oncogenic mutation	BLM mutation	0.76	0.23	0.88	0.28	0.06	0.98	16	332	348
ETV6 oncogenic mutation	ETV6 mutation	0.76	0.16	0.91	0.34	0.04	0.99	11	337	348
FLCN oncogenic mutation	FLCN mutation	0.75	0.05	0.86	0.61	0.04	1.00	7	341	348
GNA5 amplification	GNA5 amplification	0.75	0.15	0.95	0.50	0.11	0.99	21	327	348
RAD50 oncogenic mutation	RAD50 mutation	0.75	0.19	0.92	0.28	0.04	0.99	12	336	348
RTEL1 amplification	RTEL1 amplification	0.75	0.18	0.90	0.32	0.08	0.98	21	327	348
BRCA1 oncogenic mutation	BRCA1 mutation	0.75	0.08	0.89	0.40	0.04	0.99	9	339	348
TGFBR1 oncogenic mutation	TGFBR1 mutation	0.75	0.09	0.90	0.50	0.05	0.99	10	338	348
TNFAIP3 oncogenic mutation	TNFAIP3 mutation	0.75	0.10	1.00	0.25	0.04	1.00	10	338	348
CYLD oncogenic mutation	CYLD mutation	0.75	0.09	0.90	0.43	0.04	0.99	10	338	348

Table S4.4. Biomarkers detected in TCGA-ESCA

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
NOTCH1 oncogenic mutation	NOTCH1 oncogenic mutation	0.86	0.33	0.89	0.56	0.11	0.99	9	141	150
NOTCH1 deletion/mutation	NOTCH1 oncogenic mutation	0.79	0.21	0.89	0.52	0.11	0.99	9	141	150
NOTCH1 amplification/mutation	NOTCH1 oncogenic mutation	0.78	0.26	0.89	0.55	0.11	0.99	9	141	150
CDK12 amplification	CDK12 amplification	0.75	0.31	0.94	0.33	0.14	0.98	16	139	155
ERBB2 amplification/mutation	ERBB2 amplification	0.75	0.46	0.86	0.46	0.20	0.95	21	134	155

Table S4.5. Biomarkers detected in TCGA-GBM

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
ATRX deletion/mutation	ATRX oncogenic mutation	0.88	0.30	1.00	0.20	0.07	1.00	12	215	227
ATRX oncogenic mutation	ATRX oncogenic mutation	0.86	0.29	1.00	0.26	0.07	1.00	12	215	227
TP53 oncogenic mutation	TP53 oncogenic mutation	0.77	0.59	0.93	0.27	0.36	0.90	73	165	238
TP53 oncogenic mutation	TP53 mutation	0.77	0.59	0.93	0.27	0.36	0.90	73	165	238
MET amplification	MET amplification	0.75	0.08	0.90	0.51	0.05	0.99	10	360	370

Table S4.6. Biomarkers detected in TCGA-KIRP

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
SETD2 oncogenic mutation	SETD2 oncogenic mutation	0.84	0.16	1.00	0.35	0.07	1.00	12	245	257
SETD2 oncogenic mutation	SETD2 mutation	0.82	0.19	1.00	0.35	0.09	1.00	16	245	261
SETD2 deletion/mutation	SETD2 oncogenic mutation	0.79	0.13	0.83	0.53	0.08	0.98	12	245	257

Table S4.7. Biomarkers detected in TCGA-LGG

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
EGFR amplification/mutation	EGFR amplification	0.88	0.67	0.76	0.87	0.32	0.98	37	447	484
EGFR oncogenic mutation	EGFR oncogenic mutation	0.87	0.37	0.79	0.83	0.22	0.98	28	453	481
CIC deletion/mutation	CIC oncogenic mutation	0.86	0.57	0.89	0.65	0.36	0.96	87	384	471
CIC oncogenic mutation	CIC oncogenic mutation	0.85	0.57	0.80	0.73	0.40	0.94	87	384	471
FUBP1 oncogenic mutation	FUBP1 mutation	0.85	0.60	0.80	0.73	0.44	0.93	103	384	487
FUBP1 deletion/mutation	FUBP1 oncogenic mutation	0.83	0.31	0.87	0.63	0.19	0.98	45	441	486
FUBP1 oncogenic mutation	FUBP1 mutation	0.83	0.29	0.89	0.61	0.19	0.98	45	441	486
IDH1 amplification/mutation	IDH1 oncogenic mutation	0.81	0.91	0.91	0.54	0.86	0.64	373	114	487
IDH1 oncogenic mutation	IDH1 oncogenic mutation	0.80	0.92	0.87	0.55	0.86	0.57	373	114	487
IDH1 oncogenic mutation	IDH1 mutation	0.80	0.92	0.87	0.55	0.86	0.57	373	114	487
ATRX fusion	ATRX alteration	0.80	0.69	0.98	0.21	0.45	0.94	194	293	487
MTAP deletion/mutation	MTAP deletion	0.80	0.28	0.95	0.43	0.14	0.99	44	440	484
IDH2 oncogenic mutation	IDH2 oncogenic mutation	0.79	0.09	0.95	0.54	0.08	1.00	19	467	486
MDM4 amplification/mutation	MDM4 amplification	0.78	0.12	0.93	0.43	0.05	1.00	14	470	484
IDH2 oncogenic mutation	IDH2 mutation	0.76	0.09	0.90	0.54	0.08	0.99	20	467	487
CDK4 amplification	CDK4 amplification	0.76	0.13	0.83	0.61	0.08	0.99	18	465	483
CDK4 amplification/mutation	CDK4 amplification	0.75	0.15	0.78	0.56	0.06	0.98	18	465	483

Table S4.8. Biomarkers detected in TCGA-LIHC

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
BAP1 deletion/mutation	BAP1 oncogenic mutation	0.79	0.23	1.00	0.55	0.07	1.00	11	335	346
TP53 oncogenic mutation	TP53 oncogenic mutation	0.76	0.61	0.81	0.52	0.43	0.87	108	247	355
TP53 oncogenic mutation	TP53 mutation	0.76	0.61	0.81	0.52	0.43	0.87	108	247	355
TP53 deletion/mutation	TP53 oncogenic mutation	0.75	0.60	0.84	0.45	0.40	0.87	108	247	355

Table S4.9. Biomarkers detected in TCGA-LUAD

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
STK11 oncogenic mutation	STK11 oncogenic mutation	0.86	0.46	0.88	0.69	0.33	0.97	67	391	458
STK11 deletion/mutation	STK11 oncogenic mutation	0.86	0.47	0.87	0.68	0.32	0.97	67	391	458
STK11 oncogenic mutation	STK11 mutation	0.85	0.47	0.87	0.69	0.34	0.97	71	391	462
EGFR oncogenic mutation	EGFR oncogenic mutation	0.84	0.48	0.92	0.59	0.22	0.98	50	404	454
EGFR amplification/mutation	EGFR oncogenic mutation	0.83	0.47	0.92	0.50	0.18	0.98	50	404	454
MET amplification	MET amplification	0.82	0.11	1.00	0.25	0.03	1.00	11	445	456
FGFR1 amplification/mutation	FGFR1 amplification	0.81	0.13	1.00	0.34	0.04	1.00	12	437	449
FGFR1 amplification	FGFR1 amplification	0.81	0.11	1.00	0.33	0.04	1.00	12	437	449
EGFR oncogenic mutation	EGFR mutation	0.80	0.47	0.86	0.59	0.23	0.97	58	404	462
TP53 deletion/mutation	TP53 oncogenic mutation	0.79	0.79	0.96	0.26	0.58	0.85	240	222	462
TP53 oncogenic mutation	TP53 mutation	0.79	0.78	0.96	0.32	0.60	0.88	240	222	462
TP53 oncogenic mutation	TP53 oncogenic mutation	0.79	0.78	0.96	0.32	0.60	0.88	240	222	462
NSD3 amplification	NSD3 amplification	0.79	0.12	0.92	0.55	0.05	1.00	12	437	449
ATM deletion/mutation	ATM oncogenic mutation	0.77	0.20	0.95	0.34	0.06	0.99	20	424	444
RBI oncogenic mutation	RBI oncogenic mutation	0.76	0.10	0.94	0.28	0.05	0.99	17	436	453
KEAP1 oncogenic mutation	KEAP1 oncogenic mutation	0.75	0.23	0.87	0.52	0.18	0.97	45	377	422

Table S4.10. Biomarkers detected in TCGA-OV

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
TCF3 deletion	TCF3 deletion	0.95	0.41	1.00	0.55	0.07	1.00	3	94	97
STK11 deletion	STK11 deletion	0.93	0.44	1.00	0.51	0.06	1.00	3	94	97
TCF3 deletion/mutation	TCF3 deletion	0.93	0.43	1.00	0.53	0.06	1.00	3	94	97
DOT1L deletion	DOT1L deletion	0.92	0.38	1.00	0.59	0.07	1.00	3	94	97
PGR amplification	PGR amplification	0.91	0.41	1.00	0.54	0.09	1.00	4	93	97
GNA11 deletion	GNA11 deletion	0.91	0.32	1.00	0.62	0.08	1.00	3	94	97
MED12 oncogenic mutation	MED12 mutation	0.89	0.18	1.00	0.72	0.10	1.00	3	95	98
DOT1L fusion	DOT1L alteration	0.87	0.19	1.00	0.64	0.08	1.00	3	100	103
CARM1 amplification	CARM1 amplification	0.86	0.52	0.88	0.60	0.16	0.98	8	89	97
SMARCA4 amplification	SMARCA4 amplification	0.86	0.27	1.00	0.60	0.18	1.00	8	89	97
FGF3 amplification/mutation	FGF3 amplification	0.84	0.13	1.00	0.35	0.05	1.00	3	94	97
CMTR2 deletion/mutation	CMTR2 deletion	0.82	0.22	1.00	0.47	0.09	1.00	5	92	97
FGF4 amplification	FGF4 amplification	0.82	0.10	1.00	0.59	0.07	1.00	3	94	97
CCND1 amplification	CCND1 amplification	0.81	0.11	1.00	0.29	0.04	1.00	3	94	97
FGF19 amplification	FGF19 amplification	0.81	0.10	1.00	0.55	0.07	1.00	3	94	97
FGF3 amplification	FGF3 amplification	0.80	0.09	1.00	0.48	0.06	1.00	3	94	97
CMTR2 deletion	CMTR2 deletion	0.79	0.21	0.80	0.50	0.08	0.98	5	92	97
KEAP1 amplification	KEAP1 amplification	0.79	0.23	0.86	0.54	0.13	0.98	7	90	97
ZFHX3 deletion	ZFHX3 deletion	0.77	0.21	0.80	0.46	0.07	0.98	5	92	97
CARD11 deletion	CARD11 deletion	0.76	0.11	1.00	0.28	0.04	1.00	3	94	97

Table S4.11. Biomarkers detected in TCGA-PRAD

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
TP53 oncogenic mutation	TP53 mutation	0.77	0.30	0.89	0.50	0.18	0.97	45	353	398
TP53 oncogenic mutation	TP53 oncogenic mutation	0.76	0.28	0.89	0.50	0.18	0.97	44	353	397
TP53 deletion/mutation	TP53 oncogenic mutation	0.76	0.27	0.84	0.52	0.18	0.96	44	353	397
SPOP deletion/mutation	SPOP oncogenic mutation	0.75	0.31	0.90	0.34	0.13	0.97	39	357	396

Table S4.12. Biomarkers detected in TCGA-READ

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
BCOR oncogenic mutation	BCOR mutation	0.96	0.47	1.00	0.55	0.06	1.00	3	109	112
RAD52 amplification	RAD52 amplification	0.95	0.34	1.00	0.28	0.04	1.00	3	108	111
MSH2/MSH6 deficient	MSH2 alteration	0.95	0.47	1.00	0.84	0.19	1.00	4	108	112
MGA oncogenic mutation	MGA oncogenic mutation	0.93	0.19	1.00	0.65	0.07	1.00	3	107	110
NOTCH2 amplification	NOTCH2 amplification	0.93	0.21	1.00	0.39	0.04	1.00	3	109	112
KDM5A amplification/mutation	KDM5A amplification	0.92	0.28	1.00	0.31	0.04	1.00	3	108	111
KDM5A amplification	KDM5A amplification	0.90	0.16	1.00	0.29	0.04	1.00	3	108	111
BRCA1 germline/somatic point mutation/deletion or fusion	BRCA1 alteration	0.89	0.16	1.00	0.68	0.08	1.00	3	109	112
MSH2 oncogenic mutation	MSH2 mutation	0.89	0.28	1.00	0.72	0.12	1.00	4	108	112
DICER1 oncogenic mutation	DICER1 mutation	0.89	0.25	1.00	0.35	0.04	1.00	3	109	112
FOXO1 amplification	FOXO1 amplification	0.88	0.24	1.00	0.66	0.07	1.00	3	109	112
BLM oncogenic mutation	BLM mutation	0.87	0.27	1.00	0.39	0.06	1.00	4	108	112
ARID1A oncogenic mutation	ARID1A oncogenic mutation	0.87	0.62	1.00	0.32	0.06	1.00	5	106	111
MGA deletion/mutation	MGA oncogenic mutation	0.87	0.12	1.00	0.65	0.07	1.00	3	107	110
RBML0 oncogenic mutation	RBML0 oncogenic mutation	0.85	0.34	1.00	0.50	0.07	1.00	4	107	111
EGFR fusion	EGFR alteration	0.83	0.23	1.00	0.28	0.04	1.00	3	109	112
RAD51 amplification	RAD51 alteration	0.83	0.39	1.00	0.34	0.04	1.00	3	109	112
RAD51 oncogenic mutation	RAD51 mutation	0.83	0.11	1.00	0.39	0.04	1.00	3	109	112
PTEN oncogenic mutation	PTEN mutation	0.83	0.24	1.00	0.42	0.09	1.00	6	106	112
PTEN oncogenic mutation	PTEN oncogenic mutation	0.83	0.24	1.00	0.42	0.09	1.00	6	106	112
ARID1A deletion/mutation	ARID1A oncogenic mutation	0.82	0.53	0.80	0.50	0.07	0.98	5	106	111
RBML0 oncogenic mutation	RBML0 mutation	0.82	0.30	1.00	0.50	0.09	1.00	5	107	112
BRCA1 alteration	BRCA1 alteration	0.80	0.10	1.00	0.22	0.03	1.00	3	109	112
TP53 deletion/mutation	TP53 oncogenic mutation	0.79	0.90	0.85	0.41	0.78	0.52	80	32	112
TP53 oncogenic mutation	TP53 mutation	0.79	0.90	0.85	0.41	0.78	0.52	80	32	112
TP53 oncogenic mutation	TP53 oncogenic mutation	0.79	0.90	0.85	0.41	0.78	0.52	80	32	112
FLT1 amplification/mutation	FLT1 amplification	0.78	0.25	0.83	0.21	0.06	0.96	6	106	112
ARID1A oncogenic mutation	ARID1A mutation	0.77	0.53	0.83	0.32	0.06	0.97	6	106	112
RB1 oncogenic mutation	RB1 mutation	0.77	0.09	1.00	0.34	0.04	1.00	3	109	112
FLT1 amplification	FLT1 amplification	0.77	0.22	0.83	0.25	0.06	0.96	6	106	112
SMAD4 deletion	SMAD4 deletion	0.76	0.17	1.00	0.28	0.04	1.00	3	109	112

Table S4.13. Biomarkers detected in TCGA-SARC

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
ROS1 amplification/mutation	ROS1 amplification	0.90	0.16	1.00	0.20	0.04	1.00	7	239	246
NF1 oncogenic mutation	NF1 oncogenic mutation	0.89	0.29	1.00	0.56	0.06	1.00	7	238	245
NF1 deletion	NF1 deletion	0.89	0.25	1.00	0.49	0.08	1.00	10	235	245
NF1 oncogenic mutation	NF1 mutation	0.87	0.28	1.00	0.56	0.09	1.00	10	238	248
NF1 deletion/mutation	NF1 oncogenic mutation	0.86	0.32	1.00	0.40	0.05	1.00	7	238	245
YAP1 amplification	YAP1 amplification	0.84	0.11	1.00	0.26	0.04	1.00	8	237	245
MDM2 amplification	MDM2 amplification	0.84	0.56	0.98	0.42	0.28	0.99	46	200	246
CDK4 amplification/mutation	CDK4 amplification	0.84	0.56	0.86	0.53	0.28	0.95	43	203	246
MDM2 amplification/mutation	MDM2 amplification	0.83	0.53	0.98	0.41	0.28	0.99	46	200	246
CDK4 amplification	CDK4 amplification	0.83	0.55	0.91	0.51	0.28	0.96	43	203	246
EGFR amplification	EGFR amplification	0.83	0.09	0.80	0.65	0.05	0.99	5	241	246
TSC2 oncogenic mutation	TSC2 mutation	0.83	0.11	1.00	0.24	0.03	1.00	6	242	248
BIRC3 amplification	BIRC3 amplification	0.81	0.09	1.00	0.35	0.05	1.00	8	237	245
FYN amplification	FYN amplification	0.80	0.19	0.88	0.67	0.08	0.99	8	238	246
JUN amplification	JUN amplification	0.76	0.17	1.00	0.31	0.10	1.00	18	228	246
JUN amplification/mutation	JUN amplification	0.76	0.15	1.00	0.36	0.11	1.00	18	228	246
CHEK1 deletion/mutation	CHEK1 deletion	0.76	0.11	0.80	0.67	0.05	0.99	5	240	245
CHEK1 deletion	CHEK1 deletion	0.76	0.25	0.80	0.51	0.03	0.99	5	240	245
NF1 deletion/mutation	NF1 deletion	0.75	0.09	1.00	0.40	0.07	1.00	10	235	245

Table S4.14. Biomarkers detected in TCGA-SKCM

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
FAT1 oncogenic mutation	FAT1 mutation	0.86	0.27	1.00	0.47	0.10	1.00	4	68	72
SMARCA4 fusion	SMARCA4 alteration	0.86	0.43	1.00	0.39	0.07	1.00	3	69	72

Table S4.15. Biomarkers detected in TCGA-STAD

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
ZFHX3 deletion/mutation	ZFHX3 oncogenic mutation	0.93	0.25	1.00	0.46	0.06	1.00	11	332	343
KMT2B deletion/mutation	KMT2B oncogenic mutation	0.92	0.46	1.00	0.68	0.18	1.00	23	326	349
KMT2B oncogenic mutation	KMT2B oncogenic mutation	0.92	0.46	1.00	0.65	0.17	1.00	23	326	349
KMT2D deletion/mutation	KMT2D oncogenic mutation	0.92	0.55	0.91	0.75	0.21	0.99	22	303	325
KMT2D oncogenic mutation	KMT2D oncogenic mutation	0.91	0.56	0.95	0.71	0.19	1.00	22	303	325
CDH1 oncogenic mutation	CDH1 oncogenic mutation	0.91	0.23	0.85	0.81	0.15	0.99	13	324	337
ZFHX3 oncogenic mutation	ZFHX3 oncogenic mutation	0.90	0.29	1.00	0.50	0.06	1.00	11	332	343
CDH1 deletion/mutation	CDH1 oncogenic mutation	0.90	0.21	0.85	0.76	0.12	0.99	13	324	337
MGA deletion/mutation	MGA oncogenic mutation	0.89	0.17	1.00	0.58	0.08	1.00	12	334	346
MGA oncogenic mutation	MGA oncogenic mutation	0.89	0.16	1.00	0.44	0.06	1.00	12	334	346
KMT2A deletion/mutation	KMT2A oncogenic mutation	0.88	0.32	0.88	0.77	0.08	1.00	8	325	333
JAK1 oncogenic mutation	JAK1 mutation	0.88	0.34	0.83	0.82	0.14	0.99	12	346	358
CIC oncogenic mutation	CIC mutation	0.88	0.24	0.95	0.66	0.15	1.00	21	337	358
KMT2B oncogenic mutation	KMT2B mutation	0.87	0.46	0.94	0.65	0.21	0.99	32	326	358
MGA oncogenic mutation	MGA mutation	0.87	0.26	0.96	0.44	0.11	0.99	24	334	358
FANCA oncogenic mutation	FANCA mutation	0.86	0.20	0.82	0.83	0.13	0.99	11	347	358
KMT2A oncogenic mutation	KMT2A oncogenic mutation	0.86	0.31	0.88	0.61	0.05	0.99	8	325	333
ATR oncogenic mutation	ATR oncogenic mutation	0.86	0.22	0.90	0.54	0.05	0.99	10	336	346
KMT2D oncogenic mutation	KMT2D mutation	0.85	0.66	0.85	0.71	0.35	0.96	55	303	358
CIC deletion/mutation	CIC oncogenic mutation	0.84	0.14	0.83	0.61	0.07	0.99	12	337	349
CIC oncogenic mutation	CIC oncogenic mutation	0.84	0.12	0.92	0.66	0.09	1.00	12	337	349
PTPRD deletion/mutation	PTPRD oncogenic mutation	0.83	0.16	1.00	0.22	0.03	1.00	7	330	337
TP53BP1 deletion/mutation	TP53BP1 oncogenic mutation	0.83	0.14	0.86	0.50	0.03	0.99	7	340	347
ATR deletion/mutation	ATR oncogenic mutation	0.82	0.21	0.80	0.66	0.07	0.99	10	336	346
ASXL2 oncogenic mutation	ASXL2 mutation	0.82	0.13	0.89	0.51	0.04	0.99	9	349	358
ERBB3 amplification	ERBB3 amplification	0.81	0.07	1.00	0.62	0.06	1.00	8	351	359
BRCA2 oncogenic mutation	BRCA2 mutation	0.80	0.28	0.76	0.78	0.20	0.98	25	333	358
PBRM1 oncogenic mutation	PBRM1 oncogenic mutation	0.80	0.12	0.80	0.86	0.14	0.99	10	342	352
ATR oncogenic mutation	ATR mutation	0.79	0.23	0.77	0.54	0.10	0.97	22	336	358
SMARCE1 amplification	SMARCE1 amplification	0.78	0.27	1.00	0.32	0.10	1.00	24	335	359
ZFHX3 oncogenic mutation	ZFHX3 mutation	0.77	0.28	0.85	0.50	0.12	0.98	26	332	358
ARID1B oncogenic mutation	ARID1B mutation	0.77	0.15	0.80	0.54	0.07	0.98	15	343	358
PTEN deletion/mutation	PTEN oncogenic mutation	0.77	0.17	0.91	0.63	0.14	0.99	22	335	357
CTNNB1 oncogenic mutation	CTNNB1 oncogenic mutation	0.77	0.12	0.82	0.41	0.04	0.99	11	338	349
CTNNB1 amplification/mutation	CTNNB1 oncogenic mutation	0.76	0.10	0.82	0.54	0.05	0.99	11	338	349
KMT2A oncogenic mutation	KMT2A mutation	0.76	0.29	0.76	0.61	0.17	0.96	33	325	358
CDH1 fusion	CDH1 alteration	0.76	0.29	0.97	0.32	0.15	0.99	40	321	361
SMARCA4 oncogenic mutation	SMARCA4 mutation	0.75	0.20	0.88	0.31	0.06	0.98	17	341	358
PTEN oncogenic mutation	PTEN mutation	0.75	0.19	0.96	0.37	0.09	0.99	23	335	358

Table S4.16. Biomarkers detected in TCGA-THCA

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
BRAF oncogenic mutation	BRAF oncogenic mutation	0.90	0.91	0.96	0.61	0.78	0.93	285	203	488
BRAF oncogenic mutation	BRAF mutation	0.90	0.91	0.96	0.61	0.78	0.93	285	203	488
BRAF amplification/mutation	BRAF oncogenic mutation	0.90	0.91	0.96	0.60	0.77	0.92	285	203	488
HRAS oncogenic mutation	HRAS mutation	0.89	0.15	0.88	0.79	0.12	0.99	16	472	488
HRAS oncogenic mutation	HRAS oncogenic mutation	0.89	0.15	0.88	0.79	0.12	0.99	16	472	488
HRAS amplification/mutation	HRAS oncogenic mutation	0.89	0.14	1.00	0.51	0.07	1.00	16	472	488
NRAS amplification/mutation	NRAS oncogenic mutation	0.86	0.38	0.77	0.81	0.26	0.98	39	449	488

Table S4.17. Biomarkers detected in TCGA-UCEC

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
PTEN deletion/mutation	PTEN oncogenic mutation	0.87	0.91	0.81	0.70	0.88	0.69	294	159	453
PTEN oncogenic mutation	PTEN oncogenic mutation	0.87	0.90	0.79	0.77	0.87	0.67	294	159	453
PTEN oncogenic mutation	PTEN mutation	0.86	0.91	0.79	0.77	0.87	0.66	304	159	463
TP53 deletion/mutation	TP53 oncogenic mutation	0.85	0.77	0.93	0.49	0.51	0.92	167	295	462
TP53 oncogenic mutation	TP53 mutation	0.85	0.75	0.93	0.48	0.51	0.93	168	295	463
TP53 oncogenic mutation	TP53 oncogenic mutation	0.85	0.75	0.93	0.48	0.50	0.93	167	295	462
BRCA1 deletion/mutation	BRCA1 oncogenic mutation	0.85	0.10	0.91	0.76	0.09	1.00	11	422	433
RUNX1 deletion/mutation	RUNX1 oncogenic mutation	0.84	0.12	1.00	0.58	0.05	1.00	10	439	449
NOTCH2 oncogenic mutation	NOTCH2 oncogenic mutation	0.82	0.13	1.00	0.42	0.05	1.00	12	411	423
TRIP13 amplification	TRIP13 amplification	0.82	0.14	0.92	0.54	0.05	1.00	12	458	470
SDHA amplification	SDHA amplification	0.81	0.17	0.92	0.42	0.04	0.99	12	458	470
CCNE1 amplification/mutation	CCNE1 amplification	0.80	0.27	0.83	0.68	0.18	0.98	36	434	470
CCNE1 amplification	CCNE1 amplification	0.80	0.23	0.78	0.76	0.21	0.98	36	434	470
CIC oncogenic mutation	CIC oncogenic mutation	0.79	0.14	0.87	0.48	0.06	0.99	15	415	430
CALR amplification	CALR amplification	0.79	0.11	0.83	0.65	0.09	0.99	18	452	470
NOTCH3 amplification	NOTCH3 amplification	0.79	0.13	0.91	0.50	0.08	0.99	22	448	470
NOTCH2 deletion/mutation	NOTCH2 oncogenic mutation	0.78	0.14	1.00	0.35	0.04	1.00	12	411	423
NOTCH2 oncogenic mutation	NOTCH2 mutation	0.78	0.28	0.96	0.42	0.17	0.99	52	411	463
RUNX1 oncogenic mutation	RUNX1 oncogenic mutation	0.78	0.09	0.90	0.55	0.04	1.00	10	439	449
NOTCH2 amplification/mutation	NOTCH2 oncogenic mutation	0.78	0.20	1.00	0.35	0.04	1.00	12	411	423
B2M oncogenic mutation	B2M mutation	0.78	0.10	0.92	0.33	0.04	0.99	13	450	463
CTNNB1 amplification/mutation	CTNNB1 oncogenic mutation	0.77	0.54	0.78	0.54	0.32	0.90	96	342	438
BRD4 amplification/mutation	BRD4 amplification	0.77	0.11	0.95	0.43	0.08	0.99	22	448	470
CTNNB1 oncogenic mutation	CTNNB1 oncogenic mutation	0.77	0.54	0.84	0.51	0.32	0.92	96	342	438
KEAP1 amplification	KEAP1 amplification	0.77	0.09	0.83	0.53	0.07	0.99	18	451	469
CIC deletion/mutation	CIC oncogenic mutation	0.76	0.13	1.00	0.22	0.04	1.00	15	415	430
BRD4 amplification	BRD4 amplification	0.75	0.13	0.95	0.42	0.07	0.99	22	448	470

Table S4.18. Biomarkers detected in TCGA-UCS

Biomarker (MSK) - Predicted label	Biomarker (TCGA) - Ground truth label	AUC	AP	Sensitivity	Specificity	PPV	NPV	Positive	Negative	Total
SMARCA4 amplification	SMARCA4 amplification	0.81	0.22	1.00	0.35	0.11	1.00	4	48	52
ERBB2 amplification/mutation	ERBB2 amplification	0.80	0.31	1.00	0.40	0.15	1.00	5	47	52
ERBB2 amplification	ERBB2 amplification	0.80	0.30	1.00	0.38	0.15	1.00	5	47	52
PIK3CA oncogenic mutation	PIK3CA oncogenic mutation	0.80	0.68	0.94	0.44	0.47	0.94	18	34	52
PIK3CA amplification/mutation	PIK3CA oncogenic mutation	0.79	0.60	0.94	0.41	0.46	0.93	18	34	52
PIK3CA oncogenic mutation	PIK3CA mutation	0.78	0.67	0.95	0.44	0.49	0.94	19	34	53
CARM1 amplification	CARM1 amplification	0.78	0.19	1.00	0.56	0.16	1.00	4	48	52
ARHGAP35 deletion/mutation	ARHGAP35 oncogenic mutation	0.77	0.39	1.00	0.30	0.13	1.00	5	47	52
CDK12 amplification	CDK12 amplification	0.76	0.18	1.00	0.54	0.15	1.00	4	48	52

Table S5: MSK Development cohort counts

Cancer Type	Tumor Site	train			tune			test		
		patients	samples	slides	patients	samples	slides	patients	samples	slides
Non-Small Cell Lung Cancer	Primary	2410	2691	3096	332	364	414	689	781	899
	Metastasis	1064	1158	1358	143	156	174	311	338	397
Pancreatic Cancer	Primary	850	859	917	128	130	141	246	247	258
	Metastasis	638	656	718	95	102	121	195	203	223
Prostate Cancer	Primary	1207	1232	1330	190	196	212	327	332	356
	Metastasis	704	818	897	105	121	134	201	235	253
Esophagogastric Cancer	Primary	770	803	885	114	116	126	212	225	251
	Metastasis	246	254	283	32	33	34	62	62	65
Soft Tissue Sarcoma	Primary	980	1028	1135	136	139	165	266	276	301
	Metastasis	334	358	400	55	59	65	82	89	101
Colorectal Cancer	Primary	2205	2241	2524	299	305	346	630	644	734
	Metastasis	852	886	978	131	138	151	230	241	272
Thyroid Cancer	Primary	237	240	269	29	29	29	60	60	63
	Metastasis	257	280	313	40	42	44	67	71	78
Germ Cell Tumor	Primary	175	178	188	28	28	29	54	54	60
	Metastasis	140	149	160	15	15	18	38	41	43
Cancer of Unknown Primary	Primary	256	261	291	39	40	40	75	77	83
	Metastasis	573	579	630	68	71	78	183	186	205
Renal Cell Carcinoma	Primary	489	514	547	74	77	81	128	138	145
	Metastasis	225	236	258	29	29	29	55	56	60
Breast Cancer	Primary	1893	2052	2186	236	251	278	546	592	629
	Metastasis	1541	1748	1976	238	264	294	458	535	604
Ovarian Cancer	Primary	577	581	618	87	87	94	164	164	180
	Metastasis	671	684	722	100	104	114	203	207	221
Skin Cancer, Non-Melanoma	Primary	174	180	188	22	26	27	56	57	62
	Metastasis	80	85	93	10	11	12	23	24	26
Endometrial Cancer	Primary	1123	1131	1219	165	167	182	343	344	366
	Metastasis	312	321	350	39	43	44	82	85	95

Table S5: MSK Development cohort counts (continued)

Cancer Type	Tumor Site	train			tune			test		
		patients	samples	slides	patients	samples	slides	patients	samples	slides
Hepatobiliary Cancer	Primary	536	559	600	70	73	82	154	158	175
	Metastasis	180	191	210	26	28	30	55	56	57
Melanoma	Primary	370	379	417	52	53	58	117	121	141
	Metastasis	576	613	670	92	97	110	188	204	227
Glioma	Primary	1053	1164	1288	161	179	196	310	341	391
	Metastasis	2	2	2	1	2	2	1	1	1
Bladder Cancer	Primary	1077	1217	1300	129	143	148	282	321	333
	Metastasis	264	291	314	37	37	39	77	81	84
Gastrointestinal Stromal Tumor	Primary	244	257	270	32	33	35	75	83	85
	Metastasis	97	123	131	14	17	18	29	37	38
Small Cell Lung Cancer	Primary	109	110	118	12	12	14	30	30	33
	Metastasis	94	99	106	11	11	11	26	26	26
Gastrointestinal Neuroendocrine Tumor	Primary	55	55	58	9	11	13	17	17	18
	Metastasis	55	56	59	5	5	5	17	17	19
Uterine Sarcoma	Primary	105	105	121	19	19	22	29	29	32
	Metastasis	97	99	112	11	12	13	26	27	28
Peripheral Nervous System	Primary	146	163	177	23	24	24	39	43	45
	Metastasis	82	98	104	17	20	20	19	20	21
Vaginal Cancer	Primary	31	32	35	2	2	2	6	6	8
	Metastasis	8	8	8	-	-	-	4	4	4
Salivary Gland Cancer	Primary	81	83	89	19	20	22	17	17	18
	Metastasis	116	137	154	17	18	19	26	28	29
Bone Cancer	Primary	217	228	247	29	34	36	45	48	54
	Metastasis	95	106	118	12	16	18	23	32	34
Mesothelioma	Primary	183	196	213	24	24	24	50	54	61
	Metastasis	31	31	33	4	4	4	6	6	6
Histiocytosis	Primary	21	21	26	4	4	5	6	7	7
	Metastasis	3	3	3	-	-	-	1	1	1
Miscellaneous Brain Tumor	Primary	20	20	20	8	8	9	9	9	9
	Metastasis	1	1	1	-	-	-	3	3	3

Table S5: MSK Development cohort counts (continued)

Cancer Type	Tumor Site	train			tune			test		
		patients	samples	slides	patients	samples	slides	patients	samples	slides
Thymic Tumor	Primary	28	28	28	6	6	6	14	15	15
	Metastasis	18	19	22	3	3	3	3	5	5
Appendiceal Cancer	Primary	103	103	111	6	6	7	30	30	30
	Metastasis	104	109	119	17	19	19	37	38	40
Sex Cord Stromal Tu- mor	Primary	27	27	29	3	3	3	6	6	6
	Metastasis	23	24	26	6	7	9	7	8	9
Mature B-Cell Neo- plasms	Primary	98	99	107	16	18	19	22	22	24
	Metastasis	7	7	8	1	1	1	1	1	1
Anal Cancer	Primary	48	48	49	12	13	13	18	18	20
	Metastasis	25	26	32	3	3	3	7	7	7
CNS Cancer	Primary	98	105	110	12	12	13	25	26	30
	Metastasis	2	2	3	1	1	1	2	2	2
Tubular Adenoma of the Colon	Primary	16	16	17	3	3	3	3	4	4
Cervical Cancer	Primary	116	118	131	20	20	21	40	40	42
	Metastasis	53	55	61	9	12	14	17	21	22
Nerve Sheath Tumor	Primary	62	69	72	9	10	11	17	19	20
	Metastasis	13	16	17	-	-	-	3	4	5
Embryonal Tumor	Primary	27	32	39	5	7	7	4	4	4
	Metastasis	2	2	3	-	-	-	-	-	-
Head and Neck Cancer	Primary	186	193	207	27	27	30	65	65	70
	Metastasis	177	187	198	27	28	33	49	50	51
Mature T and NK Neo- plasms	Primary	22	22	24	-	-	-	8	8	9
Sellar Tumor	Primary	59	62	69	11	11	11	16	16	19
	Metastasis	3	3	4	-	-	-	-	-	-
Small Bowel Cancer	Primary	63	65	69	12	14	14	25	27	31
	Metastasis	24	25	27	3	3	3	9	10	12
Adrenocortical Carci- noma	Primary	34	34	38	1	1	1	8	8	8
	Metastasis	14	14	15	2	3	3	10	10	10

Table S5: MSK Development cohort counts (continued)

Cancer Type	Tumor Site	train			tune			test		
		patients	samples	slides	patients	samples	slides	patients	samples	slides
Ampullary Cancer	Primary	62	62	68	7	7	8	13	13	14
	Metastasis	20	21	22	-	-	-	8	8	9
Pheochromocytoma	Primary	8	8	8	3	3	4	-	-	-
	Metastasis	3	3	3	-	-	-	-	-	-
B-Lymphoblastic	Primary	9	9	10	3	3	4	4	4	5
Leukemia/Lymphoma	Metastasis	-	-	-	-	-	-	1	1	1
	Primary	5	5	6	1	1	1	2	2	2
Penile Cancer	Metastasis	1	1	2	-	-	-	2	2	2
	Primary	24	32	40	6	6	7	9	11	11
Wilms Tumor	Metastasis	9	12	15	1	1	1	4	4	5
	Primary	26	26	27	5	5	5	8	8	9
Breast Sarcoma	Metastasis	7	7	9	-	-	-	3	3	3
	Primary	26	26	27	5	5	5	8	8	9
Peritoneal Cancer, NOS	Primary	7	7	7	-	-	-	1	1	1
	Metastasis	6	6	7	2	2	2	3	3	3
Retinoblastoma	Primary	45	46	53	8	8	9	14	14	20
	Metastasis	-	-	-	1	1	1	-	-	-
Gestational Trophoblastic Disease	Primary	4	4	4	1	1	1	3	3	3
	Metastasis	4	4	5	1	1	1	-	-	-
Miscellaneous Neuroepithelial Tumor	Primary	14	14	14	1	1	1	4	4	4
	Metastasis	3	3	3	-	-	-	1	1	1
Gastrointestinal Neuroendocrine Tumors of Esophagus	Primary	3	3	5	1	1	1	-	-	-
	Metastasis	-	-	-	-	-	-	1	1	1
Adrenocortical Adenoma	Primary	1	1	1	-	-	-	-	-	-
	Metastasis	-	-	-	-	-	-	1	1	1
Mastocytosis	Primary	1	1	1	-	-	-	-	-	-
Rhabdoid Cancer	Primary	4	4	5	-	-	-	-	-	-
Parathyroid Cancer	Primary	3	3	4	-	-	-	-	-	-
Hodgkin Lymphoma	Primary	6	6	6	2	2	2	1	1	1

Table S5: MSK Development cohort counts (continued)

Cancer Type	Tumor Site	train			tune			test		
		patients	samples	slides	patients	samples	slides	patients	samples	slides
Soft Tissue Cancer	Primary	6	6	7	-	-	-	1	1	2
Angiomatoid Fibrous	Primary	3	3	4	-	-	-	-	-	-
Histiocytoma										
Pineal Tumor	Primary	6	7	9	1	1	1	2	2	2
Blood Cancer, NOS	Primary	1	1	1	-	-	-	1	1	2
Adenocarcinoma	In Situ	1	1	1	1	1	1	-	-	-
Primary	CNS	Primary	2	2	2	-	-	-	-	-
Melanocytic Tumors										
Lacrimal Gland Tumor	Primary	1	1	1	-	-	-	-	-	-
Choroid Plexus Tumor	Primary	4	5	5	1	1	1	2	3	3
Melanocytoma	Primary	1	1	3	-	-	-	1	1	1
T-Lymphoblastic	Primary	-	-	-	-	-	-	1	1	1
Leukemia/Lymphoma										
Renal Neuroendocrine	Tumor	Metastasis	1	1	1	-	-	-	-	-
Unknown		Metastasis	-	-	-	-	-	1	1	2

Table S6: TCGA Validation dataset counts

project	samples	slides
TCGA-BRCA	1054	1122
TCGA-KIRC	513	518
TCGA-THCA	506	519
TCGA-LGG	490	833
TCGA-UCEC	489	542
TCGA-LUSC	478	512
TCGA-LUAD	468	530
TCGA-HNSC	450	472
TCGA-COAD	412	414
TCGA-PRAD	403	449
TCGA-GBM	386	857
TCGA-BLCA	386	457
TCGA-LIHC	364	372
TCGA-STAD	363	386
TCGA-SKCM	292	305
TCGA-KIRP	275	299
TCGA-CESC	269	279
TCGA-SARC	254	600
TCGA-PAAD	183	203
TCGA-PCPG	176	194
TCGA-ESCA	156	158
TCGA-TGCT	149	245
TCGA-READ	141	141
TCGA-THYM	121	180
TCGA-KICH	109	109
TCGA-OV	106	107
TCGA-UVM	80	80
TCGA-MESO	75	87
TCGA-ACC	56	227
TCGA-UCS	53	87
TCGA-DLBC	44	44
TCGA-CHOL	39	39

8.3 Supplementary Figures

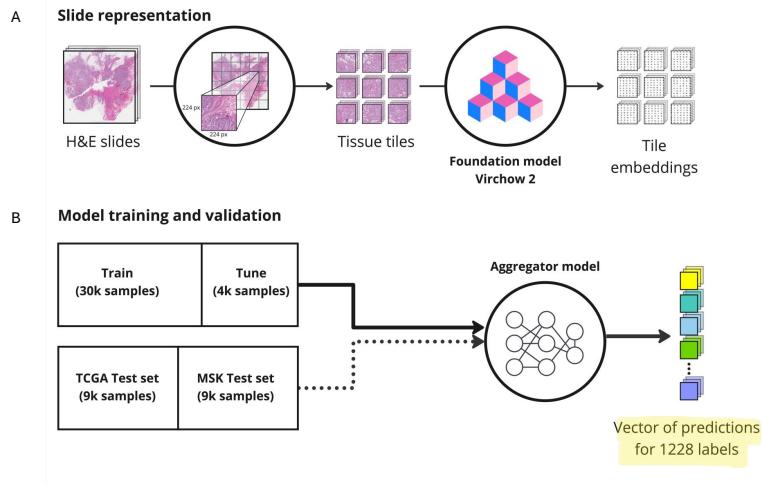


Figure S1: **A:** Schematic diagram of generating slide representations. The image is split into tiles of size 224×224 pixels. The tiles are filtered and only tiles representing tissue are kept. Next, the tissue tiles are fed to Virchow2, which generates an embedding of length 2,560 for each tile. Each slide is thus represented as a collection of its tissue tile embeddings. **B:** Schematic flow chart of model training and validation. The aggregator model receives the tile embeddings of the WSIs in the train set (processed as described in **A**) and runs validation during training on the tune set. Once trained, the model is evaluated on two unseen test sets: the TCGA and MSK cohorts.

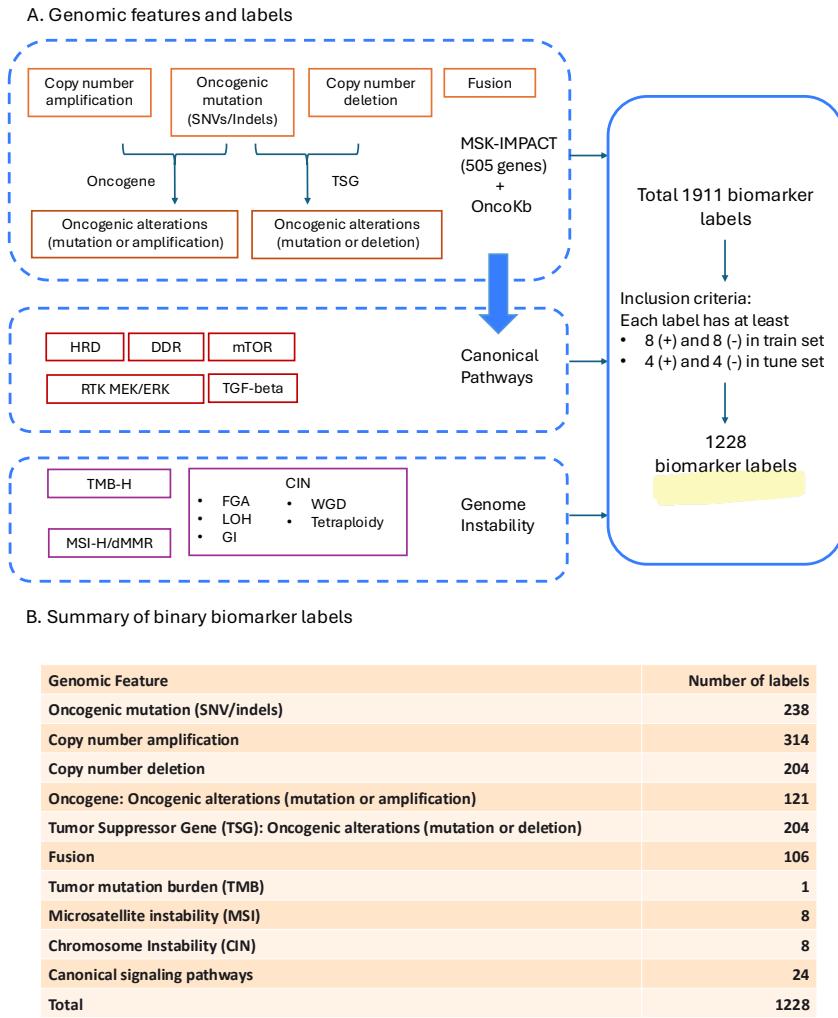


Figure S2: Genomic feature and binary biomarker label derivation. **A:** Overview of genomic features, including 1) gene level alteration labels of 505 genes from MSK-IMPACT panel, 2) alterations in a group of genes participating in 5 canonical signaling pathways, 3) genome instability: tumor mutation burden (TMB), microsatellite instability high (MSI-H) or defects in mismatch repair genes (dMMR), and chromosomal instability (CIN) measured by fraction of genome altered (FGA), loss-of-heterozygosity (LOH), genome instability index (GI), whole genome doubling (WGD), and tetraploidy. The alterations include mutations (SNVs/Indels), copy number aberration (amplification and deletion), and fusion events. The oncogenic status was determined based on OncoKB annotation. **B:** Summary table showing the number of labels in different label categories.

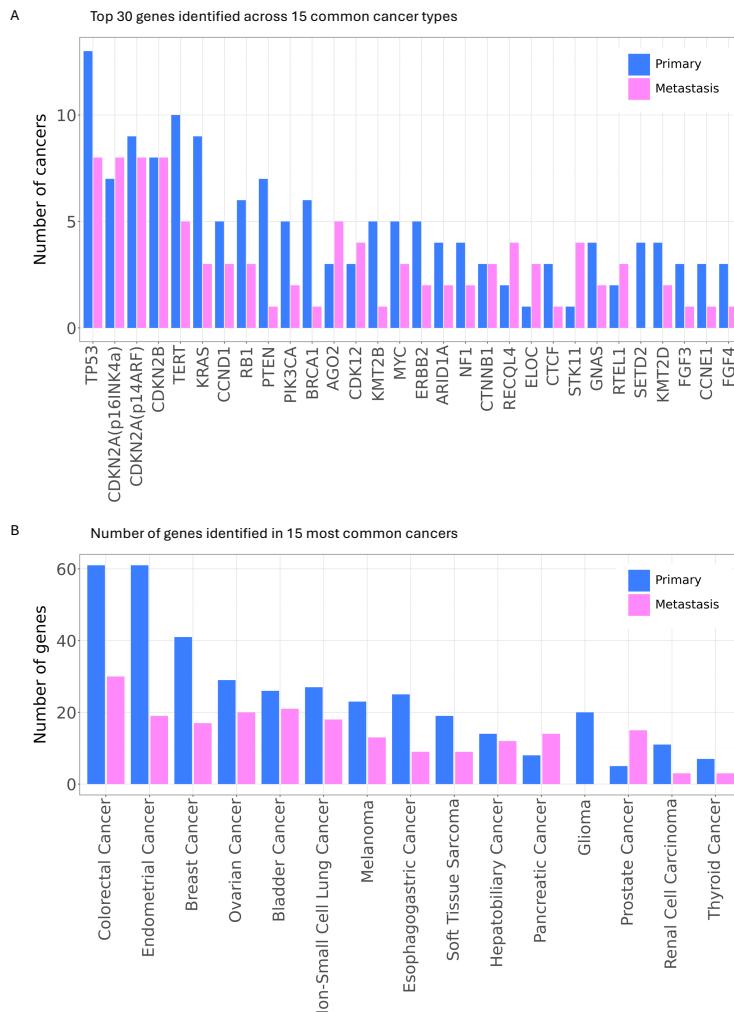
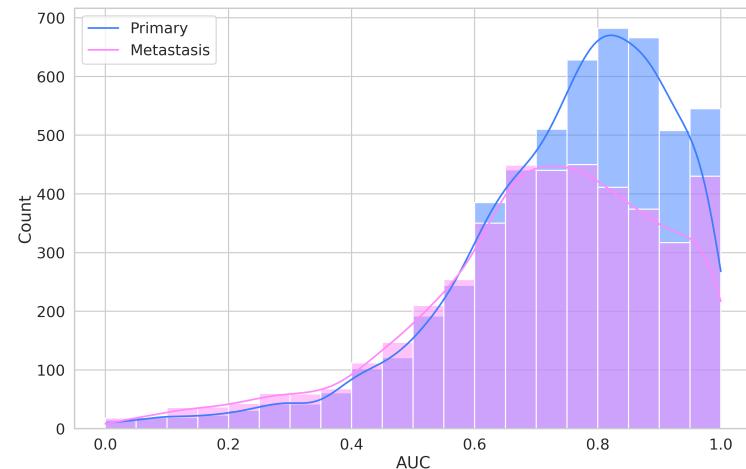


Figure S3: Gene biomarkers identified in the 15 most common cancers. **A:** Top 30 gene biomarkers identified across 15 most common cancers, with $AUC > 0.75$. **B:** Distribution of number of genes biomarkers identified across 15 most common cancers.

A. Distribution of the AUC metric



B. Distribution of the AUC metric across 15 most common cancer types

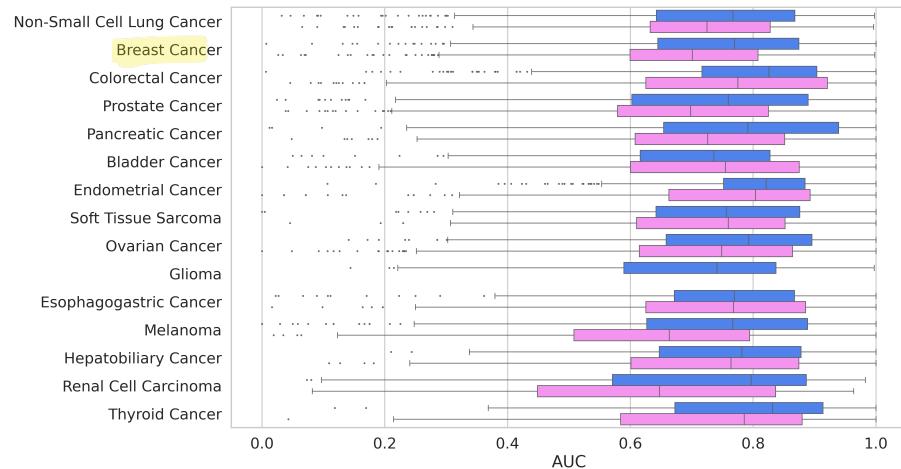


Figure S4: Distribution of the AUC for all biomarker labels in the MSK test set results. **A:** Histogram showing the distribution of the AUC score across all cancer types, comparing primary and metastatic samples. **B:** Boxplots showing the distribution of the AUC score in each of the 15 common cancer types, comparing primary and metastatic samples.

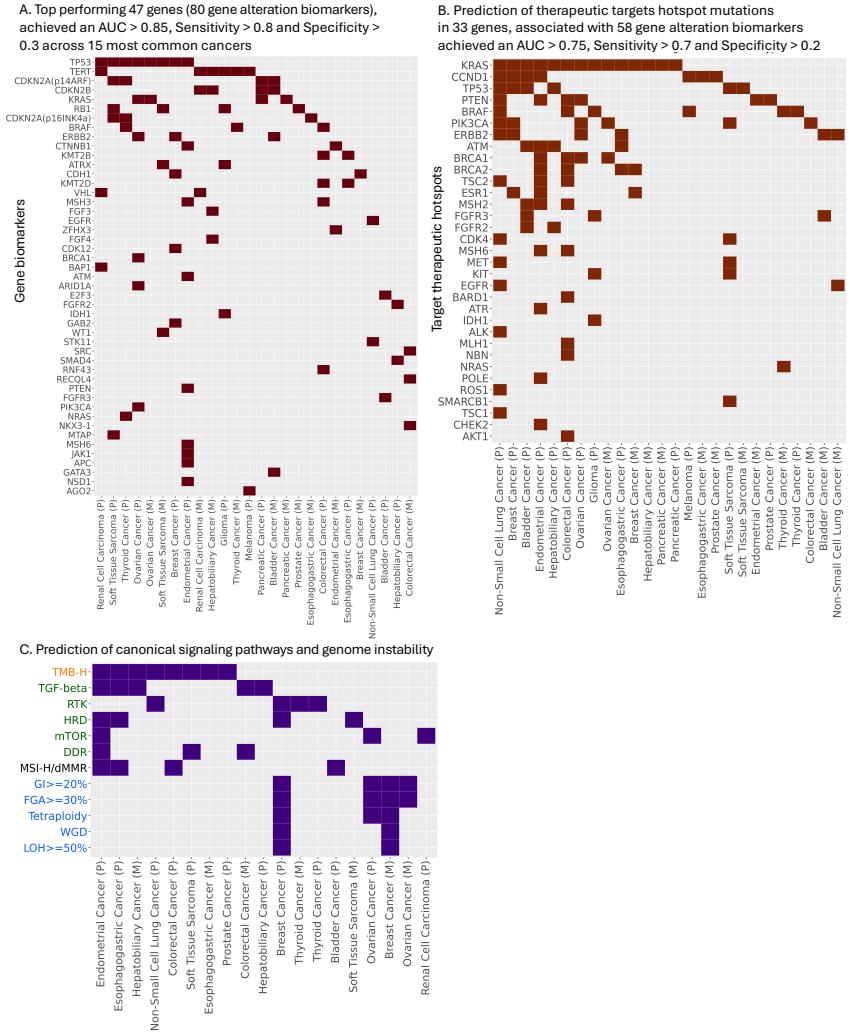


Figure S5: Highlighted biomarker predictions passing filtering criteria. **A:** Top performing biomarkers (N=80) with AUC> 0.85, sensitivity> 0.8, specificity> 0.3, representing 47 genes, across 15 most common cancers. Biomarker labels with more than 50 samples, positive sample ratio > 2% were included. **B:** Prediction of specific therapeutic hotspot mutations, highlighting the targeted gene harboring the hotspot mutations detected by our AI model, which achieved an AUC> 0.75, sensitivity> 0.7, specificity> 0.2, with at least 5 positive samples in the corresponding 58 genetic alteration biomarkers of 33 genes. **C:** Prediction of 5 canonical signaling pathways (TGF- β , RTK, HRD, mTOR, DDR), tumor mutation burden (TMB), microsatellite instability (MSI), and chromosomal instability measures: genome instability index (GI), fraction of genome altered (FGA), tetraploidy, whole genome doubling (WGD) and loss-of-heterozygosity (LOH), highlighted the biomarker predictions, achieving an AUC> 0.75, sensitivity> 0.75, specificity> 0.2, with positive sample ratio> 2%.

References

- [1] Paulina Krzyszczyk, Alison Acevedo, Erika J Davidoff, Lauren M Timmins, Ileana Marrero-Berrios, Misaal Patel, Corina White, Christopher Lowe, Joseph J Sherba, Clara Hartmanshenn, et al. The growing role of precision and personalized medicine for cancer treatment. *Technology*, 6(03n04):79–100, 2018.
- [2] Uttpal Anand, Abhijit Dey, Arvind K Singh Chandel, Rupa Sanyal, Amarnath Mishra, Devendra Kumar Pandey, Valentina De Falco, Arun Upadhyay, Ramesh Kandimalla, Anupama Chaudhary, et al. Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. *Genes & Diseases*, 10(4):1367–1401, 2023.
- [3] Katherine S Virgo, R Bryan Rumble, Ronald de Wit, David S Mendelson, Thomas J Smith, Mary-Ellen Taplin, James L Wade III, Charles L Bennett, Howard I Scher, Paul L Nguyen, et al. Initial management of noncastrate advanced, recurrent, or metastatic prostate cancer: ASCO guideline update. *Journal of Clinical Oncology*, 39(11):1274–1305, 2021.
- [4] Katharina Beyer, Lisa Moris, Michael Lardas, Anna Haire, Francesco Bartletta, Simone Scuderi, Megan Molnar, Ronald Herrera, Abdul Rauf, Riccardo Campi, et al. Diagnostic and prognostic factors in patients with prostate cancer: a systematic review. *BMJ Open*, 12(4):e058267, 2022.
- [5] Ryon P Graf, Virginia Fisher, Joaquin Mateo, Ole V Gjoerup, Russell W Madison, Kira Raskina, Hanna Tukachinsky, James Creeden, Rachel Cunningham, Richard SP Huang, et al. Predictive Genomic Biomarkers of Hormonal Therapy Versus Chemotherapy Benefit in Metastatic Castration-resistant Prostate Cancer. *European Urology*, 81(1):37–47, 2022.
- [6] Andrew NJ Tutt, Judy E Garber, Bella Kaufman, Giuseppe Viale, Debora Fumagalli, Priya Rastogi, Richard D Gelber, Evandro de Azambuja, Anitra Fielding, Judith Balmaña, et al. Adjuvant Olaparib for Patients with BRCA1- or BRCA2-Mutated Breast Cancer. *New England Journal of Medicine*, 384(25):2394–2405, 2021.
- [7] Spyridon P Basourakos, Michael Tzeng, Patrick J Lewicki, Krishnan Patel, Bashir Al Hussein Al Awamlih, Siv Venkat, Jonathan E Shoag, Michael A Gorin, Christopher E Barbieri, and Jim C Hu. Tissue-Based Biomarkers for the Risk Stratification of Men With Clinically Localized Prostate Cancer. *Frontiers in Oncology*, 11:676716, 2021.
- [8] Scott E Eggner, R Bryan Rumble, Andrew J Armstrong, Todd M Morgan, Tony Crispino, Philip Cornford, Theodorus van der Kwast, David J Grignon, Alex J Rai, Neeraj Agarwal, et al. Molecular Biomarkers in Localized Prostate Cancer: ASCO Guideline. *Journal of Clinical Oncology*, 38(13):1474–1494, 2020.

- [9] Zeina Habli, Walid AlChamaa, Raya Saab, Humam Kadara, and Massoud L Khraiche. Circulating Tumor Cell Detection Technologies and Clinical Utility: Challenges and Opportunities. *Cancers*, 12(7):1930, 2020.
- [10] Gabriella Cirmena, Martina Dameri, Francesco Ravera, Piero Fregatti, Alberto Ballestrero, and Gabriele Zoppoli. Assessment of Circulating Nucleic Acids in Cancer: From Current Status to Future Perspectives and Potential Clinical Applications. *Cancers*, 13(14):3460, 2021.
- [11] Albert Juan Ramon, Chaitanya Parmar, Oscar M Carrasco-Zevallos, Carlos Csiszer, Stephen SF Yip, Patricia Raciti, Nicole L Stone, Spyros Triantos, Michelle M Quiroz, Patrick Crowley, et al. Development and deployment of a histopathology-based deep learning algorithm for patient prescreening in a clinical trial. *Nature Communications*, 15(1):4690, 2024.
- [12] Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, Anant Madabhushi, Parantu Shah, Michaela Spitzer, and Shanrong Zhao. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6):463–477, 2019.
- [13] Jakob Nikolas Kather, Lara R Heij, Heike I Grabsch, Chiara Loeffler, Amelie Echle, Hannah Sophie Muti, Jeremias Krause, Jan M Niehues, Kai AJ Sommer, Peter Bankhead, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature Cancer*, 1(8):789–799, 2020.
- [14] Heather D Couture. Deep Learning-Based Prediction of Molecular Tumor Biomarkers from H&E: A Practical Review. *Journal of Personalized Medicine*, 12(12):2022, 2022.
- [15] Hui Qu, Mu Zhou, Zhennan Yan, He Wang, Vinod K Rustgi, Shaoting Zhang, Olivier Gevaert, and Dimitris N Metaxas. Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning. *npj Precision Oncology*, 5(1):87, 2021.
- [16] Sarah Volinsky-Fremond, Nanda Horeweg, Sonali Andani, Jurriaan Barkey Wolf, Maxime W Lafarge, Cor D de Kroon, Gitte Ørtoft, Estrid Høgdall, Jouke Dijkstra, Jan J Jobsen, et al. Prediction of recurrence risk in endometrial cancer with multimodal deep learning. *Nature Medicine*, pages 1–12, 2024.
- [17] Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, Nicolo Fusi, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature Medicine*, pages 1–12, 2024.

- [18] Eric Zimmermann, Eugene Vorontsov, Julian Viret, Adam Casson, Michal Zelechowski, George Shaikovski, Neil Tenenholz, James Hall, Thomas Fuchs, Nicolo Fusi, Siqi Liu, and Kristen Severson. Virchow 2: Scaling Self-Supervised Mixed Magnification Models in Pathology, 2024.
- [19] Donavan T Cheng, Talia N Mitchell, Ahmet Zehir, Ronak H Shah, Ryma Benayed, Aijazuddin Syed, Raghu Chandramohan, Zhen Yu Liu, Helen H Won, Sasinya N Scott, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *The Journal of molecular diagnostics*, 17(3):251–264, 2015.
- [20] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.
- [21] Kazuaki Takahashi, Masataka Takenaka, Aikou Okamoto, David DL Bowtell, and Takashi Kohno. Treatment Strategies for ARID1A-Deficient Ovarian Clear Cell Carcinoma. *Cancers*, 13(8):1769, 2021.
- [22] Yuanwang Pan, Han Han, Hai Hu, Hua Wang, Yueqiang Song, Yuan Hao, Xinyuan Tong, Ayushi S Patel, Selim Misirliglu, Sittinon Tang, et al. KMT2D deficiency drives lung squamous cell carcinoma and hypersensitivity to RTK-RAS inhibition. *Cancer Cell*, 41(1):88–105, 2023.
- [23] Alessandro Gambella, Luca Bertero, Milena Rondón-Lagos, Ludovica Verdun Di Cantogno, Nelson Rangel, Chiara Pitino, Alessia Andrea Ricci, Luca Mangherini, Isabella Castellano, and Paola Cassoni. FISH Diagnostic Assessment of MDM2 Amplification in Liposarcoma: Potential Pitfalls and Troubleshooting Recommendations. *International Journal of Molecular Sciences*, 24(2):1342, 2023.
- [24] Chia-Chin Wu, Hannah C Beird, Salah-Eddine Lamhamdi-Cherradi, Melinda Soeung, Davis Ingram, Danh D Truong, Robert W Porter, Sandhya Krishnan, Latasha Little, Curtis Gumbus, et al. Multi-site desmoplastic small round cell tumors are genetically related and immune-cold. *npj Precision Oncology*, 6(1):21, 2022.
- [25] Joo Kyung Park, Woo Hyun Paik, Kyoungbun Lee, Ji Kon Ryu, Sang Hyub Lee, and Yong-Tae Kim. DAXX/ATRX and MEN1 genes are strong prognostic markers in pancreatic neuroendocrine tumors. *Oncotarget*, 8(30):49796, 2017.
- [26] Junhui Hu, Ping Tan, Moe Ishihara, Nicholas A Bayley, Shiruyeh Schokrpur, Jeremy G Reynoso, Yangjun Zhang, Raymond J Lim, Camelia Dumitras, Lu Yang, et al. Tumor heterogeneity in VHL drives metastasis in

clear cell renal cell carcinoma. *Signal Transduction and Targeted Therapy*, 8(1):155, 2023.

- [27] Abigail K Suwala, Marius Felix, Dennis Friedel, Damian Stichel, Daniel Schrimpf, Felix Hinz, Ekkehard Hewer, Leonille Schweizer, Hildegard Dohmen, Ute Pohl, et al. Oligosarcomas, IDH-mutant are distinct and aggressive. *Acta Neuropathologica*, 143(2):263–281, 2022.
- [28] Charles Swanton. My cancer genome: a unified genomics and clinical trial portal. *The Lancet Oncology*, 13(7):668–669, 2012.
- [29] Marilyn E Holt, Kathleen F Mittendorf, Michele LeNoue-Newton, Neha M Jain, Ingrid Anderson, Christine M Lovly, Travis Osterman, Christine Micheel, and Mia Levy. My cancer genome: coevolution of precision oncology and a molecular oncology knowledgebase. *JCO Clinical Cancer Informatics*, 5:995–1004, 2021.
- [30] Debyani Chakravarty, Jianjiong Gao, Shaun M Phillips, Ritika Kundra, Hui Zhang, Jing Wang, Joshua E Rudolph, Rona Yaeger, Tenley Soumerai, Moriah H Nissan, Michael T Chang, Steven Kundel, Garth Fell, Zachary Heins, Radhika Chandramohan, Ritu Somwar, Agnes Viale, David M Hyman, Vasilis Syrgkanis, Anthony Funnell, Shaun Badges, Armin Danesh, Wei Song, Nils Weinhold, Samuel F Bakhoum, Marlies Verstoep, Lillian Nichols, Laura Borsu, Jessie Foo, Jill P Mesirov, Michael Peto, Gustavo Rossi, Michael C Wendt, Emek Demir, Dora Dias-Santagata, Craig H Mermel, Marc Ladanyi, Alan Ashworth, Boris Reva, David B Solit, and Nikolaus Schultz. OncoKB: A Precision Oncology Knowledge Base. *JCO Precision Oncology*, 1:1–16, 2017.
- [31] Jonathan R Suehnholz, Amanda Lee, Samuel K Jackson, Hannah Y Kim, Michael J Thompson, Daniela Rivera, Sophia L Martinez, Vikram Patel, Emily J Moore, Katherine L Wong, and Jason Liu. Advancements in Cancer Treatment: A Decade of Progress. *Cancer Discovery*, 13(7):789–805, 2023.
- [32] Vincenzo Di Nunno, Lidia Gatto, Alicia Tosoni, Stefania Bartolini, and Enrico Franceschi. Implications of braf v600e mutation in gliomas: Molecular considerations, prognostic value and treatment evolution. *Frontiers in Oncology*, 12:1067252, 2023.
- [33] Sandra M Swain, Mythili Shastry, and Erika Hamilton. Targeting HER2-positive breast cancer: advances and future directions. *Nature reviews Drug discovery*, 22(2):101–126, 2023.
- [34] Stephanie PL Saw, Xiuning Le, Lizza EL Hendriks, and Jordi Remon. New Treatment Options for Patients with Oncogene-Addicted Non-Small Cell Lung Cancer Focusing on EGFR-mutant tumors. *American Society of Clinical Oncology Educational Book*, 44(3):e432516, 2024.

- [35] Mark A Socinski, Nathan A Pennell, and Kurtis D Davies. Met exon 14 skipping mutations in non–small-cell lung cancer: an overview of biology, clinical outcomes, and testing considerations. *JCO precision oncology*, 5, 2021.
- [36] Remi Veillon, Hiroshi Sakai, Xiuning Le, Enriqueta Felip, Alexis B Corrot, Egbert F Smit, Keunchil Park, Frank Griesinger, Christian Britschgi, Yi-Long Wu, et al. Safety of tepotinib in patients with met exon 14 skipping nsclc and recommendations for management. *Clinical lung cancer*, 23(4):320–332, 2022.
- [37] Mirat Shah, Hima Lingam, Xin Gao, Haley Gittleman, Mallorie H Fiero, Danielle Krol, Nikolett Biel, Tiffany K Ricks, Wentao Fu, Salaheldin Hamed, et al. US Food and Drug Administration Approval Summary: Elacestrant for Estrogen Receptor–Positive, Human Epidermal Growth Factor Receptor 2–Negative, ESR1–Mutated Advanced or Metastatic Breast Cancer. *Journal of Clinical Oncology*, 42(10):1193–1201, 2024.
- [38] Kamaneh Montazeri and Joaquim Bellmunt. Erdafitinib for the treatment of metastatic bladder cancer. *Expert review of clinical pharmacology*, 13(1):1–6, 2020.
- [39] Mrinal M Gounder, Narasimhan P Agaram, Sally E Trabucco, Victoria Robinson, Richard A Ferraro, Sherri Z Millis, Anita Krishnan, Jessica Lee, Steven Attia, Wassim Abida, et al. Clinical genomic profiling in the management of patients with soft tissue and bone sarcoma. *Nature Communications*, 13(1):3406, 2022.
- [40] Lisa Elze, Rachel S van der Post, Janet R Vos, Arjen R Mensenkamp, Mirjam SC de Hullu, Iris D Nagtegaal, Nicoline Hoogerbrugge, Richarda M de Voer, and Marjolijn JL Ligtenberg. Microsatellite instability in noncolorectal and nonendometrial malignancies in patients with Lynch syndrome. *JNCI: Journal of the National Cancer Institute*, 115(7):853–860, 2023.
- [41] Francisco Martínez-Jiménez, Ali Movasati, Sascha Remy Brunner, Luan Nguyen, Peter Priestley, Edwin Cuppen, and Arne Van Hoeck. Pan-cancer whole-genome comparison of primary and metastatic solid tumours. *Nature*, 618(7964):333–341, 2023.
- [42] Ashton C Berger, Anil Korkut, Rupa S Kanchi, Apurva M Hegde, Walter Lenoir, Wenbin Liu, Yuexin Liu, Huihui Fan, Hui Shen, Visweswaran Ravikumar, et al. A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. *Cancer Cell*, 33(4):690–705, 2018.
- [43] Yang Liu, Nilay S Sethi, Toshinori Hinoue, Barbara G Schneider, Andrew D Cherniack, Francisco Sanchez-Vega, Jose A Seoane, Farshad Farshidfar, Reanne Bowlby, Mirazul Islam, et al. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. *Cancer Cell*, 33(4):721–735, 2018.

- [44] Gary H Lyman and Harold L Moses. Biomarker tests for molecularly targeted therapies—the key to unlocking precision medicine. *New England Journal of Medicine*, 375(1):4–6, 2016.
- [45] Yue Zhou, Lei Tao, Jiahao Qiu, Jing Xu, Xinyu Yang, Yu Zhang, Xinyu Tian, Xinqi Guan, Xiaobo Cen, and Yinglan Zhao. Tumor biomarkers for diagnosis, prognosis and targeted therapy. *Signal Transduction and Targeted Therapy*, 9(1):132, 2024.
- [46] Higinio Dopeso, Andrea M Gazzo, Fatemeh Derakhshan, David N Brown, Pier Selenica, Sahar Jalali, Arnaud Da Cruz Paula, Antonio Marra, Edaise M da Silva, Thais Basili, et al. Genomic and epigenomic basis of breast invasive lobular carcinomas lacking CDH1 genetic alterations. *npj Precision Oncology*, 8(1):33, 2024.
- [47] Fresia Pareja, Higinio Dopeso, Yi Kan Wang, Andrea M Gazzo, David N Brown, Monami Banerjee, Pier Selenica, Jan H Bernhard, Fatemeh Derakhshan, Edaise M da Silva, et al. A genomics-driven artificial intelligence-based model classifies breast invasive lobular carcinoma and discovers cdh1 inactivating mechanisms. *Cancer Research*, 2024.
- [48] Changxia Shao, Gerald Li, Lingkang Huang, Scott Pruitt, Emily Castellanos, Garrett Frampton, Kenneth R Carson, Tamara Snow, Gaurav Singal, David Fabrizio, et al. Prevalence of High Tumor Mutational Burden and Association With Survival in Patients With Less Common Solid Tumors. *JAMA Network Open*, 3(10):e2025109–e2025109, 2020.
- [49] Raf Sciot. MDM2 Amplified Sarcomas: A Literature Review. *Diagnostics*, 11(3):496, 2021.
- [50] Shuo Zhang, Tinghe Fang, Yexuan He, Weichen Feng, Zhuoyang Yu, Yaoyao Zheng, Chi Zhang, Shuai Hu, Zhuojun Liu, Jia Liu, et al. VHL mutation drives human clear cell renal cell carcinoma progression through PI3K/AKT-dependent cholesteryl ester accumulation. *Ebiomedicine*, 103, 2024.
- [51] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, 25(8):1301–1309, 2019.
- [52] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2019.