Supplementary Information

1. Choosing input features for developing OncoNPC

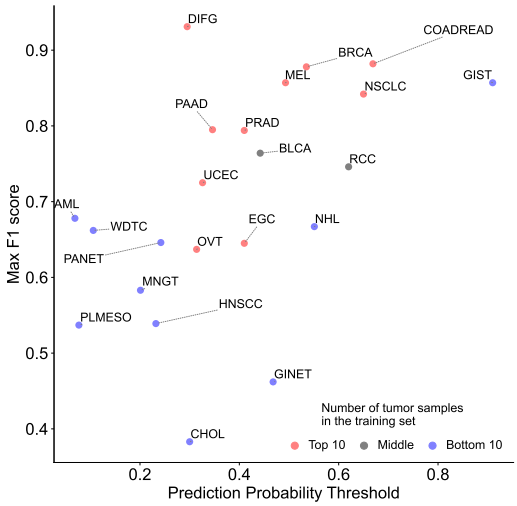
When choosing features for developing OncoNPC, it is important to acknowledge that CUP tumors often exhibit fundamentally different characteristics from known primaries, including stage, location/biopsy site, and pathological features. Including such characteristics specific to known primaries in the model may introduce a bias and potentially overconfident wrong predictions. Thus, we have chosen to prioritize the use of targeted panel sequencing data, which provides relatively unbiased and reliable biological signals about the tumors. Furthermore, targeted panel sequencing data is increasingly becoming a routine part of cancer care in many centers, which enhances the model's translational potential.

Pathology findings can be important information for determining primary sites. However, inconsistent guidelines across cancer centers make it challenging to integrate them into the model [1]. Furthermore, for patients with CUP, only 25% of tumors have immunohistochemistry (IHC) results that indicate a single primary diagnosis [2]. This lack of consistency in guidelines, coupled with the subpar and uncertain performance of pathology in identifying primary sites for challenging tumors, can result in unreliable and inconsistent primary site predictions for CUP tumors.

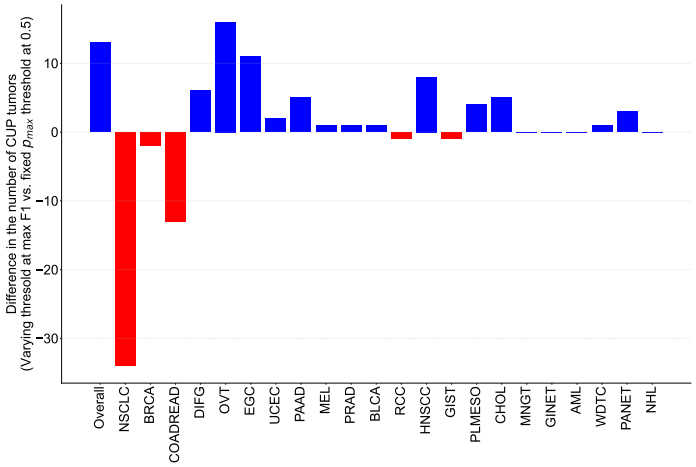
With regards to patient demographics such as ethnicity and environmental variables, it has been demonstrated that these factors can improve disease predictions [3, 4]. However, it is crucial to exercise caution when including such factors in machine learning models, as they may lead to unwanted disparities in model outcomes across different patient subgroups [5]. While acknowledging the potential benefits of incorporating sensitive patient demographics and environmental variables in predicting cancer types and relevant clinical outcomes, we recognize that this is an area that requires further investigation. Thus, we leave it as a future work to study the effects of such variables.

2. Optimizing F1 scores across cancer types using varying prediction probability thresholds

Throughout our clinical analyses, we have utilized a fixed prediction probability threshold to identify tumor samples with a moderately high degree of confidence (e.g., $p_{\max} \geq 0.5$) in their primary cancer type predictions. However, an alternative approach is to use varying prediction probability thresholds for different cancer types, which achieves the best F1 score for each individual cancer type. In order to determine the optimal F1 score and corresponding prediction probability threshold for each cancer type, we conducted an analysis, which is illustrated in Supplementary Fig. S1a.



(a)



(b)

Figure S1: **Utilization of varying thresholds for prediction probabilities that maximizes the F1 score for each cancer**. **(a)** The scatter plot shows the maximum F1 score and the corresponding prediction probability threshold for each cancer type. The red dots correspond to the top 10 cancer types with the highest number of tumor samples in the training set, while the blue dots correspond to the bottom 10 cancer types. The gray dots correspond to all the other cancer types. **(b)** Impact of varying the threshold for achieving maximum F1 for each cancer on cancer type predictions for CUP tumors at the DFCI. We compared the number of CUP tumors across predicted cancer types, obtained by using varying prediction probability thresholds that optimize the F1 score for each cancer, with those obtained by using a fixed $p_{\max}$ of 0.5 for all cancer types. The blue bars correspond to the increase in the number of tumors when using the varying threshold strategy, while the red bars correspond to the decrease in the number of tumors.

Our analysis indicates that the bottom 10 minority cancer types exhibit considerably lower maximum F1 scores compared to the top 10 majority cancer types (F1 score 0.601 vs. 0.799), which is consistent with the findings presented in Results: OncoNPC accurately classifies 22 known cancer types. Additionally, the minority cancer types reached their maximum F1 scores at lower prediction

probability thresholds in comparison to the majority cancer types (prediction probability threshold 0.316 vs. 0.445), suggesting that the predictions for the minority cancer types are generally less confident.

Furthermore, we investigated the impact of varying the threshold for achieving maximum F1 for each cancer on cancer type predictions for CUP tumors at the DFCI. To do that, we compared the number of CUP tumors across predicted cancer types, obtained by using varying prediction probability thresholds that optimize the F1 score for each cancer, with those obtained by using a fixed $p_{\max}$ of 0.5 for all cancer types (as it is used as the exclusion criterion in our CUP analyses; see Extended Data Fig. 6). In Supplementary Fig. S1b, we visualized the resulting differences in the number of CUP tumors across predicted cancer types. The blue bars correspond to the increase in the number of tumors when using the varying threshold strategy, while the red bars correspond to the decrease in the number of tumors.

Overall, using the varying threshold strategy resulted in an increase in the number of included CUP tumors from 798 to 811 (13 additional tumors included). Although there was a reduction in the number of included tumors for NSCLC, BRCA, and COADREAD predicted CUP tumors (with decreases of 34, 2, and 13 tumors, respectively), the overall increase was due to many other cancer types. Notably, OVT, EGC, and HNSCC were the top three contributors to the overall increase, with increases of 16, 11, and 8 tumors, respectively.

3. OncoNPC performance under real-world dataset shifts and difficult-to-predict cancers

Real-world dataset shifts

OncoNPC achieved robust performance against potential dataset shifts due to the factors including cancer center, biopsy site type, sequence panel version, and patient ethnicity (Main Fig. 2e). OncoNPC showed comparable performance, measured in area under the precision-recall curve (AUC-PR), for tumor samples from DFCI (AUC-PR = 0.893, n = 3,690) and those from MSK (AUC-PR = 0.850, n = 3,331). OncoNPC performance for those from VICC was slightly lower (AUC-PR = 0.760, n = 268). Refer to Extended Data Fig. 2a for more detailed center-specific OncoNPC performance. OncoNPC showed comparable performance for primary tumor samples (AUC-PR = 0.872, n = 4,525) and metastatic tumor samples (AUC-PR = 0.869, n = 2,605), demonstrating its capability to predict the primary cancer types of metastatic tumors without compromising its performance. Furthermore, as shown in Main Fig. 2e, we investigated its performance across sequence panel versions utilized at DFCI. The OncoNPC performance on tumor samples from earlier versions of DFCI sequence panels (OncoPanel v1: AUC-PR = 0.821, n = 414 and OncoPanel v2: AUC-PR = 0.887, n = 1,050) was slightly lower than the performance on the tumor samples from the most recent panel (OncoPanel v3: AUC-PR = 0.907, n = 2,226) which also contained the largest number of genes. As all tumor samples have been collected from OncoPanel v3 since October 2016, we expect our model to make high-quality predictions in a prospective setting. Finally, OncoNPC demonstrated reliable performance for all patient ethnicities, achieving an AUC-PR of over 0.8 for each ethnicity. However, the model performed better for white patients compared to other ethnicities (White: AUC-PR =

0.877, n = 6,035, Asian: AUC-PR = 0.812, n = 342, Other: AUC-PR = 0.834, n = 347, Black: AUC-PR = 0.844, n = 312, and Spanish/Hispanic: AUC-PR = 0.825, n = 253). This discrepancy in performance may be attributed to the fact that the training cohort predominantly consisted of white patients (83.2% White, 4.46% Asian, 4.78% Other, 4.27% Black, and 3.25% Spanish/Hispanic).

OncoPanel version variability on primary site predictions

To investigate the impact of OncoPanel version variability during the model training on primary site predictions, we added 7 additional one-hot-encoded input features that represented different panel versions (3 versions at DFCI, 3 versions at MSK, and 2 versions at VICC). We then retrained the model and analyzed changes in performance and the proportion of predicted cancer types in held-out CKP and CUP tumor samples. Supplementary Fig. S2a shows the F1 score differences across 22 cancer types between the original OncoNPC model and the model with the panel features. Negative values in red indicate cases where the model with the panel features outperformed the original OncoNPC model, while positive values in blue indicate the opposite. Supplementary Fig. S2b shows the absolute differences in the proportions of 22 predicted cancer types in percentage between the original OncoNPC model and the new model in held-out CKP tumors (blue) and CUP tumors (green), respectively.

Regarding the changes in model performance, we observed only a negligible change in the overall F1 score when compared to the original OncoNPC model (weighted overall F1 difference of -3.68×10^{-4}). However, we did observe larger performance changes for some minority cancer types, such as PLMESO (Pleural Mesothelioma), AML (Acute Myeloid Leukemia), and NHL (Non-Hodgkin Lymphoma) with a mean absolute difference in F1 score of 0.0491 among them. With regard to the proportion of predicted cancer types among CKP tumors and CUP tumors, we observed mean absolute differences in percentages of 0.0549% and 0.253%, respectively. Our findings suggest that incorporating the panel variability into the model training process has only minor effects on the model performance and predicted cancer types. Additionally, any resulting small changes are likely influenced by stochastic elements of the training process, such as weight initializations.

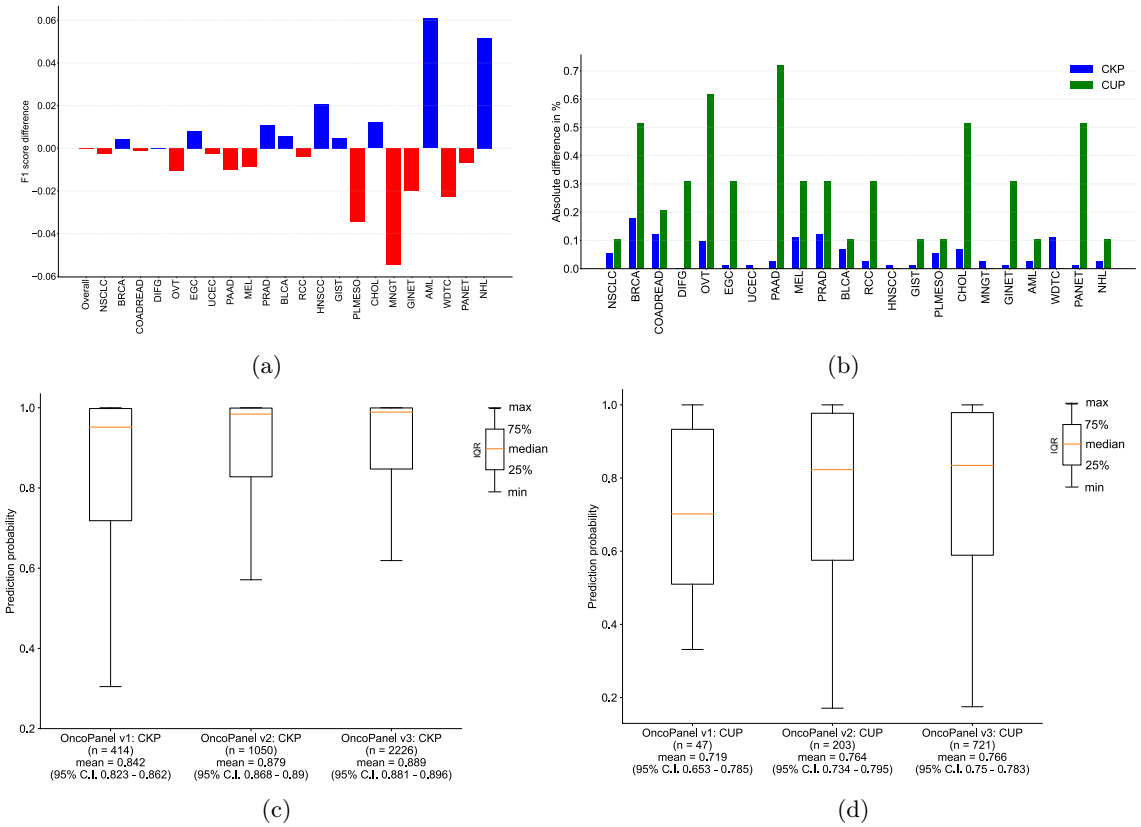


Figure S2: **OncoNPC performance with respect to sequencing panel variability**. We added 7 panel version features, retrained the model, and analyzed changes in performance and predicted cancer types in the held-out CKP and CUP tumor samples. **(a)** F1 score differences across 22 cancer types between the original OncoNPC model and the model with the panel features. Negative values in red indicate cases where the model with the panel features outperformed the original OncoNPC model, while positive values in blue indicate the opposite. We observed a negligible change in overall F1 score (-3.68×10^{-4}) compared to the original OncoNPC model. **(b)** Absolute differences in the proportions of 22 predicted cancer types in percentage between the original OncoNPC model and the new model in held-out CKP tumors (blue) and CUP tumors (green), respectively. For both **(a)** and **(b)**, the cancer types on the x-axis are sorted in a decreasing order of the number of tumor samples. We observed mean absolute differences in percentages of 0.0549% and 0.253% among CKP tumors and CUP tumors, respectively. The panel variability, when incorporated into the model training process, has only minor effects on the model performance and predicted cancer types. **(c)**, **(d)** Box plots comparing prediction confidences ($p_{\max}$) across OncoPanel versions for **(c)** CKP tumors and **(d)** CUP tumors at DFCI. The figures show medians, lower and upper quartiles, as well as the mean and 95% confidence intervals, along with the number of tumor samples.

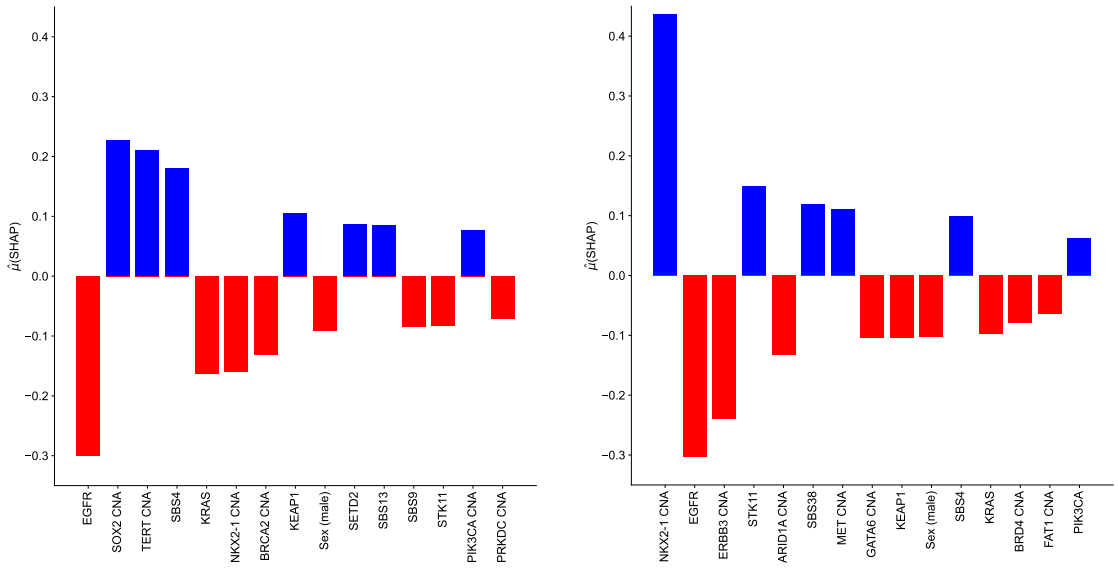
Regarding the impacts of OncoPanel versions on the prediction confidences, we have provided box plots of the prediction confidences ($p_{\max}$) of the final OncoNPC model for both CKP (Cancer with Known Primary) tumors and CUP tumors across three different OncoPanel versions at DFCI, as shown in Supplementary Fig. S2d and Supplementary Fig. S2c, respectively. For CKP tumors, the prediction confidences of OncoPanel v1 sequenced tumors were significantly lower than those of OncoPanel v2 and v3 sequenced tumors (with $p_{\max}$ mean of 0.842 and 95% C.I. of 0.823 - 0.862

for OncoPanel v1, as compared to 0.879 and 95% C.I. of 0.868 - 0.890, and 0.889 and 95% C.I. of 0.881 - 0.896 for OncoPanel v2 and v3, respectively). For CUP tumors, the prediction confidences of OncoPanel v1 sequenced tumors were not significantly different from those of OncoPanel v2 and v3 sequenced tumors (with $p_{\max}$ mean of 0.719 and 95% C.I. of 0.653 - 0.785 for OncoPanel v1, as compared to 0.764 and 95% C.I. of 0.734 - 0.795, and 0.766 and 95% C.I. of 0.750 - 0.783 for OncoPanel v2 and v3, respectively). Finally, we want to highlight that OncoPanel v1, which targets the fewest genes compared to OncoPanel v2 and v3 (304 genes vs. 326 and 447 genes, respectively), only accounts for 10.2% of CKP tumors and 4.8% of CUP tumors. The vast majority of tumors at DFCI were sequenced using the latest version of the panel, OncoPanel v3 (representing 61.6% of CKP tumors and 74.3% of CUP tumors). We anticipate that this percentage will continue to rise as new tumor samples at DFCI are routinely sequenced with OncoPanel v3.

Difficult-to-predict cancers

Five cancer types with the lowest F1 were CHOL (0.368), GINET (0.436), PLMESO (0.479), HNSCC (0.500), and MNGT (0.545). Nevertheless, when samples were predicted with high confidence ($p_{\max} \geq 0.9$), the performance improved as shown in Extended Data Fig. 2a: CHOL (precision: 0.783, recall: 0.500, F1: 0.610), GINET (precision: 1.00, recall: 0.478, F1: 0.647), PLMESO (precision: 0.765, recall: 0.684, F1: 0.722), HNSCC (precision: 0.818, recall: 0.692, F1: 0.750), and MNGT (precision: 1.00, recall: 0.778, F1: 0.875). This demonstrates that the OncoNPC was still able to make high-quality predictions for a subset of tumor samples in rare cancer types, for which training data was limited.

We further investigated the factors underlying the suboptimal performance of relatively more common cancer types such as CHOL and HNSCC. As shown in the confusion matrix in Extended Data Fig. 1a, a sizable proportion of HNSCC and CHOL tumors were inaccurately classified as NSCLC (28 out of 134 HNSCC tumors and 26 out of 126 CHOL tumors). In order to gain insight into the OncoNPC's reasoning behind the misclassification of HSNCC and CHOL tumor samples as NSCLC, we selected a subset of samples (n = 14 for HSNCC and n = 7 for CHOL) that were mis-classified as NSCLC with a moderately high prediction probability (i.e., $p_{\max} \geq 0.5$). We then investigated the SHAP values for NSCLC in these samples, as illustrated in Supplementary Fig. S3, which shows the 15 most important features for each group based on the mean SHAP values. The features are plotted on the x-axis, and the corresponding mean SHAP values on the y-axis, with positive values (in blue) corresponding to features that favor the prediction towards NSCLC, and negative values (in red) corresponding to features that discourage the prediction towards NSCLC.

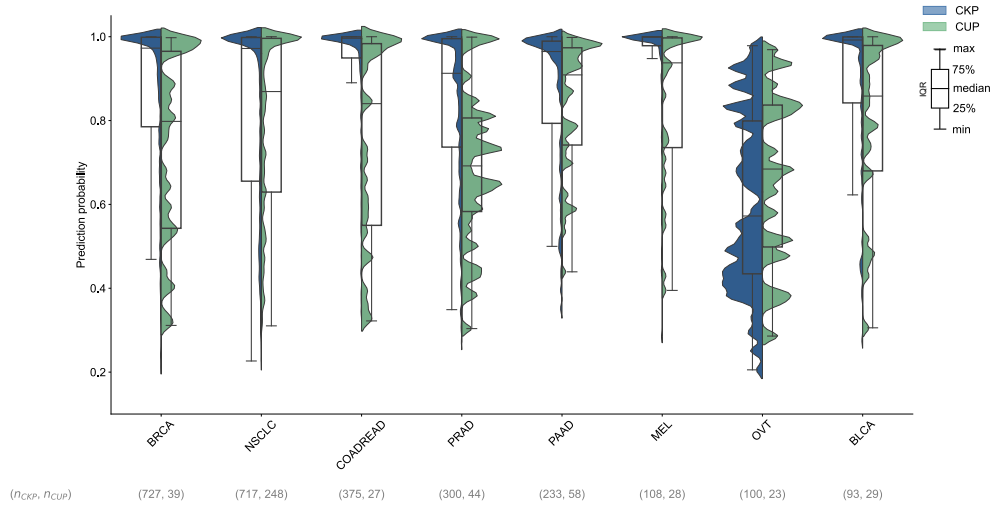


(a) HNSCC tumors (n = 14) classified as NSCLC (b) CHOL tumors (n = 7) classified as NSCLC

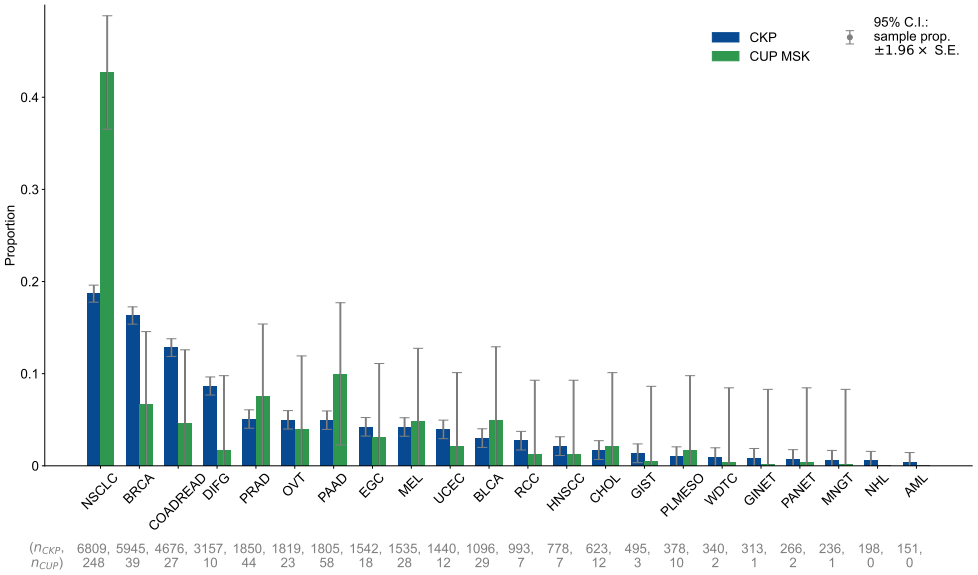
Figure S3: **Top 15 most important features for (a) HNSCC tumors and (b) CHOL tumors classified as NSCLC**. OncoNPC predicted these tumor samples as NSCLC with prediction probability ≥ 0.5. On the y-axis, the mean SHAP values of each feature are plotted, with positive values (in blue) corresponding to features that favor the prediction towards NSCLC and negative values (in red) corresponding to features that discourage the prediction towards NSCLC. For HNSCC tumors, the key features for favoring the NSCLC prediction are copy number alterations (CNA) in the *SOX2* and *TERT* genes, as well as the tobacco smoking-related mutation signature SBS4. For CHOL tumors, key features include CNA in the *NKX2-1* gene, mutation in the *STK11* gene, and the mutation signature SBS38 (whose association is currently unknown). Notably, none of the tumor samples showed a mutation in *EGFR*.

Among the HNSCC tumors, the model relied heavily on copy number alterations (CNA) in the *SOX2* and *TERT* genes, as well as the tobacco smoking-related mutation signature SBS4 to predict NSCLC. Interestingly, the SBS4 mutation signature was significantly more enriched in the NSCLC-classified HNSCC tumors compared to the rest of the HNSCC tumors (0.123 vs. 0.091). Upon reviewing the patient charts, we discovered that the average pack-years of those 14 patients were 31, and only two of the tumors came from HPV-positive patients, indicating that smoking-related signatures played a significant role in incorrect predictions, while HPV status played a minor role.

For CHOL tumors, key features included CNA in the *NKX2-1* gene, mutation in the *STK11* gene, mutation signature SBS38 (currently unknown), and SBS4 (tobacco). The absence of EGFR/KRAS mutations in these tumors negatively impacted the model's prediction (indicated by red bars), but were outweighed by the other factors in these samples. For both cancer types, it thus appears that a smoking signature together with some genomic hallmarks of NSCLC (*KEAP1*, STK11, and *PIK3CA*) contributed to an NSCLC mis-classification in these samples (Main Fig. 3d and Supplementary Fig. S10j). Notably, OncoNPC's interpretability allows us to delve deeper into the model's errors from a clinical perspective for specific cases.

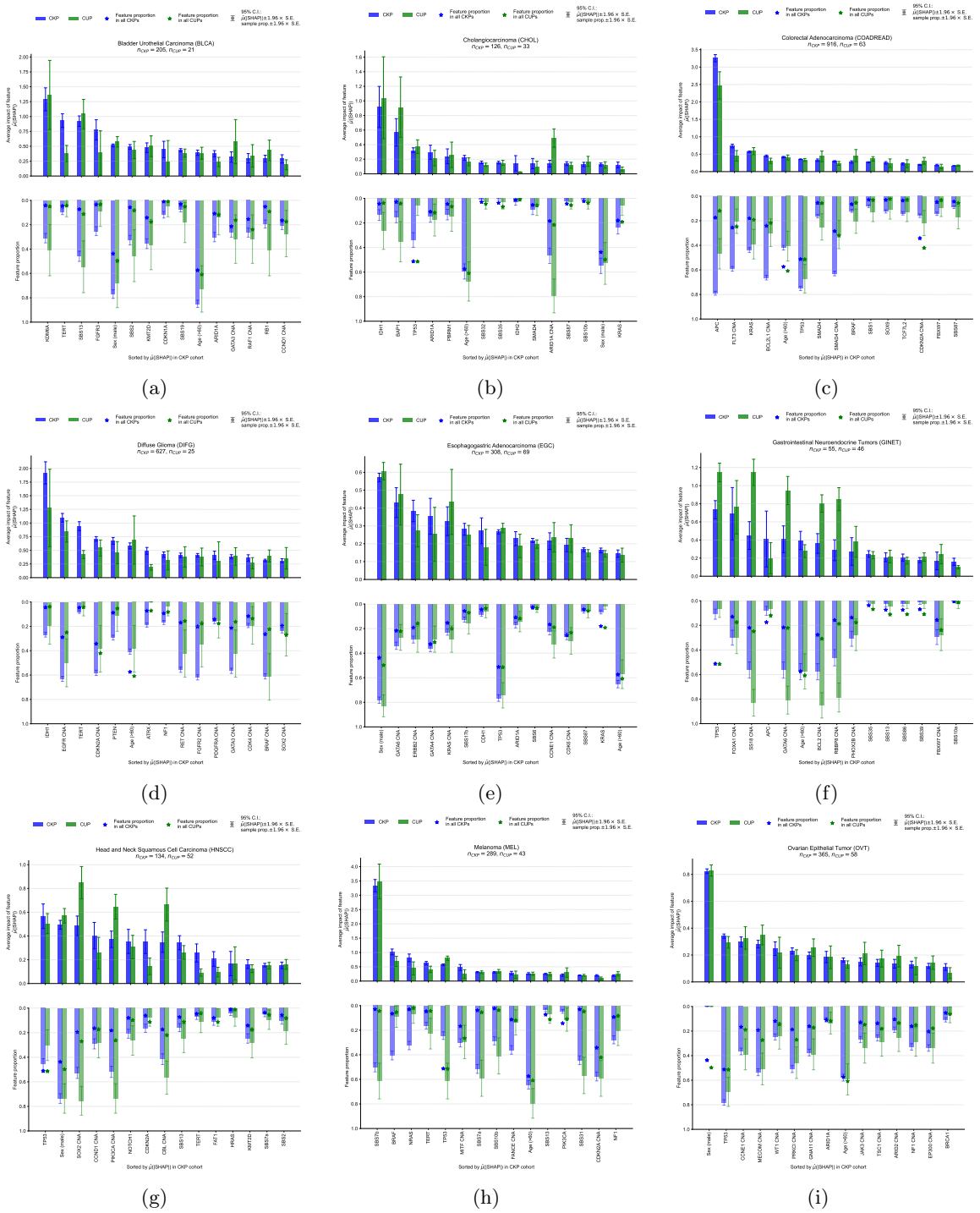


(a)



(b)

Figure S4: **Applying OncoNPC to MSK CUP tumor samples. (a)** Empirical distributions of prediction probabilities for correctly predicted held-out CKP tumor samples (n = 3,429) and MSK CUP tumor samples (n = 496), broken down by CKP cancer types (blue) and their corresponding OncoNPC predicted cancer types for CUP tumors (green). Only OncoNPC classifications with at least 20 CUP tumor samples are shown. **(b)** Proportion of each CKP cancer type and the corresponding OncoNPC predicted CUP cancer type. All training CKP tumor samples (n = 36,445) and all MSK CUP tumor samples (n = 581) are shown. For both **(a)** and **(b)**, the cancer types (x-axis) are ordered by the number of CKP tumor samples in each cancer type

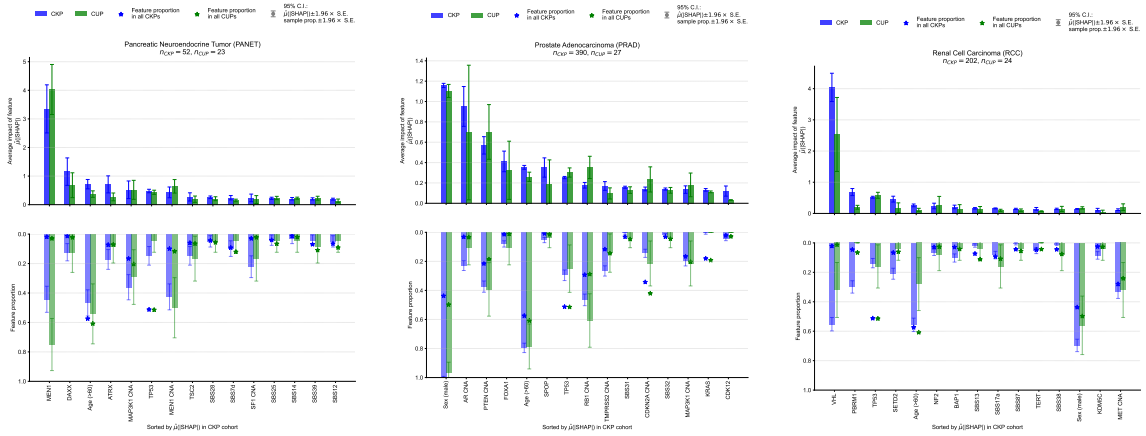


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure S5



(j) (k) (l)

Figure S5: **Interpreting OncoNPC predictions**. Top 15 most important features, based on mean absolute SHAP values (i.e., $\hat{\mu}(|\text{SHAP}|)$ [6]), for cancer types that had a minimum of 20 CUP tumors classified.

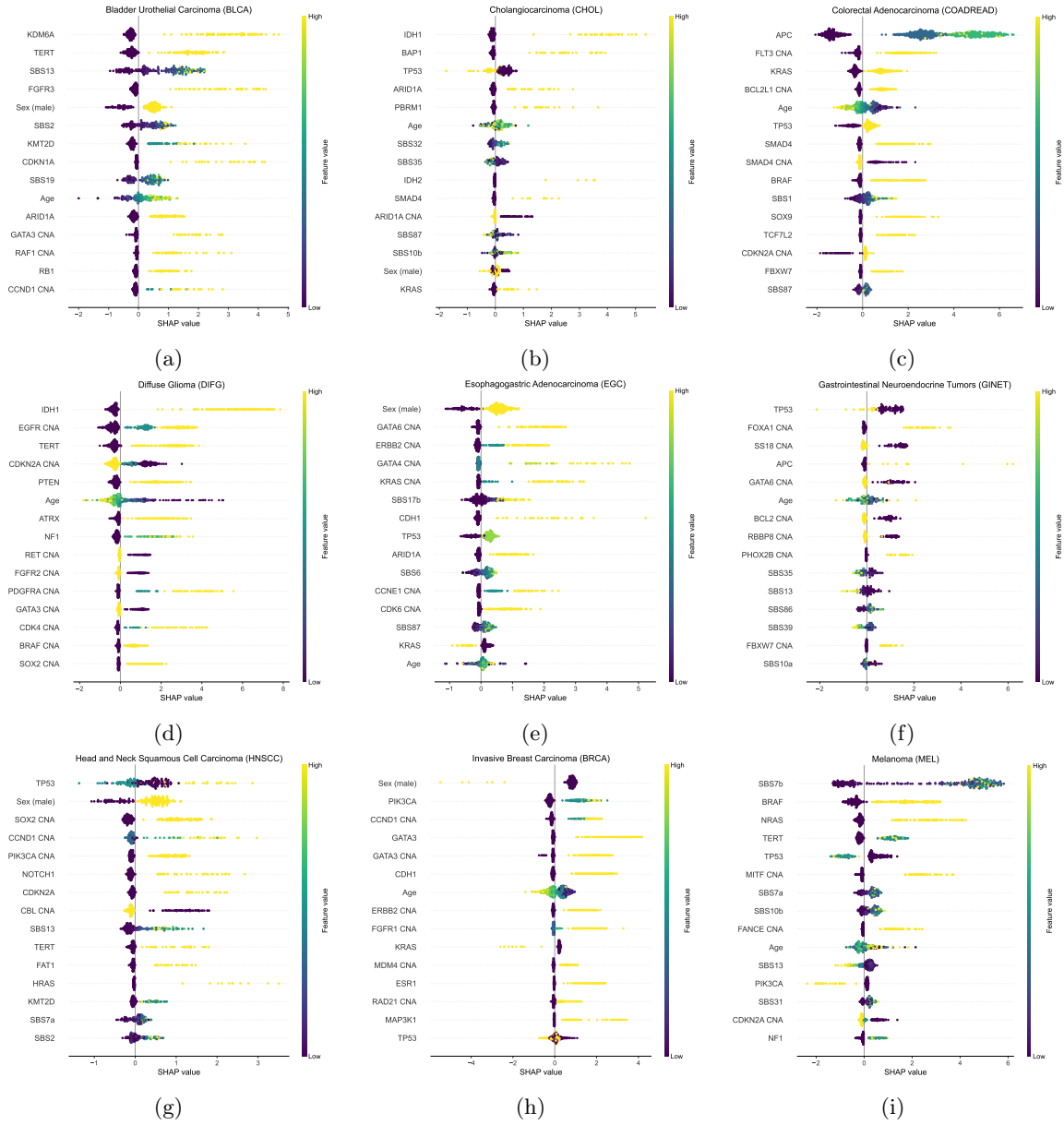

Bladder Urothelial Carcinoma (BLCA)
KDM6A
TERT
SBS13
FGFR3
Sex (male)
SBS2
KMT2D
CDKN1A
SBS19
Age
ARID1A
GATA3 CNA
RAF1 CNA
RB1
CCND1 CNA
SHAP value
High
Low
Feature value
(a)
Cholangiocarcinoma (CHOL)
IDH1
BAP1
TP53
ARID1A
PBRM1
Age
SBS32
SBS35
IDH2
SMAD4
ARID1A CNA
SBS87
SBS10b
Sex (male)
KRAS
SHAP value
High
Low
Feature value
(b)
Colorectal Adenocarcinoma (COADREAD)
APC
FLT3 CNA
KRAS
BCL2L1 CNA
Age
TP53
SMAD4
SMAD4 CNA
BRAF
SBS1
SOX9
TCF7L2
CDKN2A CNA
FBXW7
SBS87
SHAP value
High
Low
Feature value
(c)
Diffuse Glioma (DIFG)
IDH1
EGFR CNA
TERT
CDKN2A CNA
PTEN
Age
ATRX
NF1
RET CNA
FGFR2 CNA
PDGFRA CNA
GATA3 CNA
CDK4 CNA
BRAF CNA
SOX2 CNA
SHAP value
High
Low
Feature value
(d)
Esophagogastric Adenocarcinoma (EGC)
Sex (male)
GATA6 CNA
ERBB2 CNA
GATA4 CNA
KRAS CNA
SBS17b
CDH1
TP53
ARID1A
SBS6
CCNE1 CNA
CDK6 CNA
SBS87
KRAS
Age
SHAP value
High
Low
Feature value
(e)
Gastrointestinal Neuroendocrine Tumors (GINET)
TP53
FOXA1 CNA
SS18 CNA
APC
GATA6 CNA
Age
BCL2 CNA
RBBP8 CNA
PHOX2B CNA
SBS35
SBS13
SBS86
SBS39
FBXW7 CNA
SBS10a
SHAP value
High
Low
Feature value
(f)
Head and Neck Squamous Cell Carcinoma (HNSCC)
TP53
Sex (male)
SOX2 CNA
CCND1 CNA
PIK3CA CNA
NOTCH1
CDKN2A
CBL CNA
SBS13
TERT
FAT1
HRAS
KMT2D
SBS7a
SBS2
SHAP value
High
Low
Feature value
(g)
Invasive Breast Carcinoma (BRCA)
Sex (male)
PIK3CA
CCND1 CNA
GATA3
GATA3 CNA
CDH1
Age
ERBB2 CNA
FGFR1 CNA
KRAS
MDM4 CNA
ESR1
RAD21 CNA
MAP3K1
TP53
SHAP value
High
Low
Feature value
(h)
Melanoma (MEL)
SBS7b
BRAF
NRAS
TERT
TP53
MITF CNA
SBS7a
SBS10b
FANCE CNA
Age
SBS13
PIK3CA
SBS31
CDKN2A CNA
NF1
SHAP value
High
Low
Feature value
(i)

Figure S6

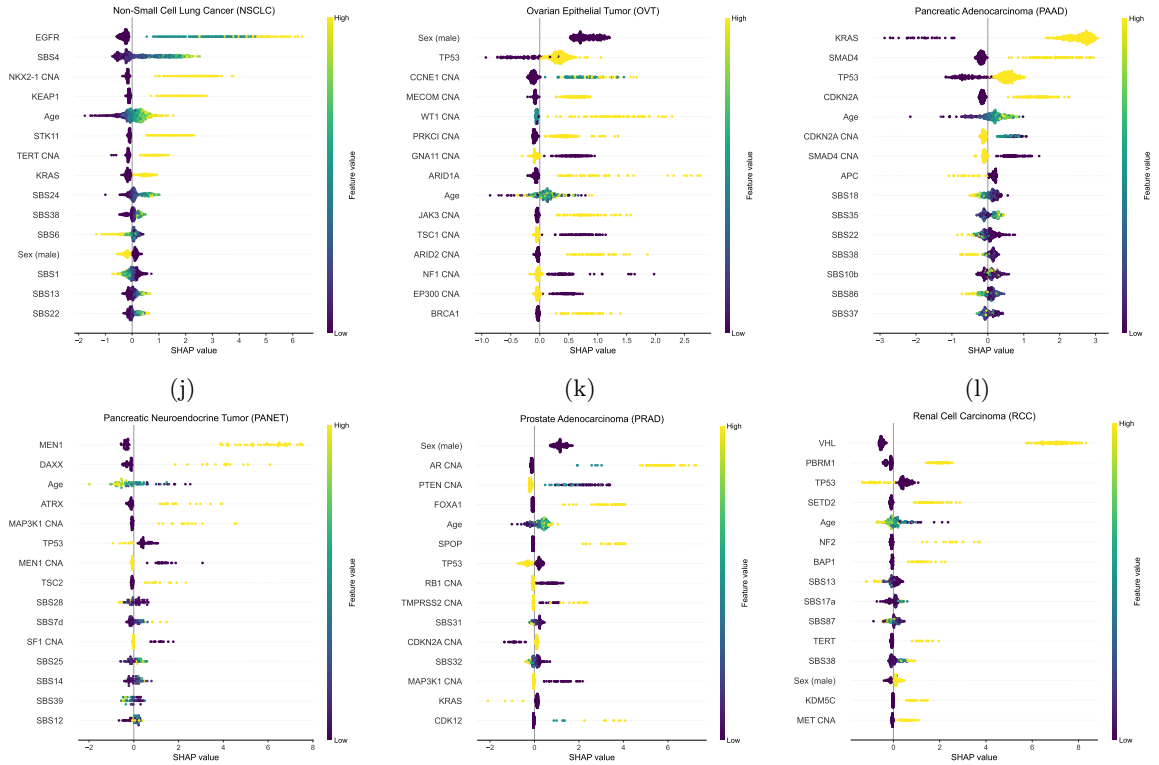


Figure S6: **SHAP summary plot** [6] for cancer types that had a minimum of 20 CUP tumor samples classified. SAHP values (i.e., impact on OncoNPC predictions) are shown on the x-axis, while feature values are shown as a color map (from purple to yellow). In each plot, CUP and CKP tumor samples were combined into a single cohort for the corresponding cancer.

4. Robustness of OncoNPC performance with respect to key genomics input features

In order to identify the key panel features, we took the average of absolute SHAP values for each input feature across 22 cancer types. We then utilized these aggregated feature importance values to rank the input features, which were primarily related to age, sex, mutation, and CNA events. The top 20 most crucial features with their aggregated SHAP values are shown in Supplementary Table S1, while Supplementary Data 3 provides a full list of the ranked input features along with their corresponding aggregated SHAP values.

Table S1: Top 20 most globally important features with their aggregated SHAP values. To obtain the aggregated SHAP values, we took the average of absolute SHAP values for each input feature across 22 cancer types. We utilized these aggregated SHAP values to rank the input features, which were primarily related to age, sex, mutation, and CNA events. See Supplementary Data 3 for a full list of the ranked input features along with their corresponding aggregated SHAP values

OncoNPC features	Aggregated SHAP
Sex	0.659
TP53	0.523
Age	0.376
KRAS	0.326
CDKN2A CNA	0.180
PIK3CA	0.158
APC	0.120
BRAF	0.117
MYC CNA	0.103
EGFR CNA	0.095
TERT	0.088
PTEN	0.078
KIT	0.069
ARID1A	0.067
KMT2D	0.065
CCND1 CNA	0.063
CDKN2B CNA	0.053
NF2	0.053
SMAD4 CNA	0.053
ERBB2 CNA	0.050

Furthermore, in order to investigate the impact of input genomics features on OncoNPC's robustness, we performed a feature ablation study, where we chose the most important genes based on their aggregated SHAP values from the previous analysis and gradually reduced them from all 846 features associated with those genes, as well as age and sex, to only the top 10% (i.e., top 29 fea-

tures). In each feature configuration, we re-trained the model with the same set of hyperparameters and evaluated its performance on the held-out CKP (Cancer with Known Primary) tumor samples (n = 7,289), which were utilized throughout this work. The results of this analysis are presented in Extended Data Fig. 3, which shows the breakdown of OncoNPC performance in F1 score by 22 cancer types across increasing prediction confidence. The cancer types on the y-axis are sorted in a decreasing order of the number of tumor samples. Finally, Supplementary Data 4 provides a list of input features that correspond to the selected genes in each configuration.

OncoNPC demonstrates good overall performance, even when using just the top 50% of features compared to using all the chosen features, with an overall weighted F1 score of 0.757 vs. 0.777 at a minimum $p_{\max}$ threshold of 0, and an F1 score of 0.950 vs. 0.960 at the minimum $p_{\max}$ threshold of 0.9. However, when using only the top 10% of features, the performance drops significantly across all cancer types compared to using all the chosen features, with an overall weighted F1 score of 0.512 at a minimum $p_{\max}$ threshold of 0.0 and 0.892 at a minimum $p_{\max}$ threshold of 0.9. Interestingly, minority cancer types, such as MEL (Melanoma), RCC (Renal Cell Carcinoma), GIST (Gastrointestinal Stromal Tumor), and GINET (Gastrointestinal Neuroendocrine Tumors), showed improved performance when we selected only the top 50% of features compared to the full feature setting. Specifically, MEL had an F1 score of 0.771 vs. 0.761, RCC had an F1 score of 0.769 vs. 0.759, GIST had an F1 score of 0.833 vs. 0.824, and GINET had an F1 score of 0.488 vs. 0.475 at a minimum $p_{\max}$ threshold of 0. We suspect that dropping features regularizes the feature space and improves generalizability over some of the minority cancer types.

In summary, our feature ablation study demonstrates OncoNPC achieves good performance even with a reduced set of genes. While prospective users should still exercise caution and conduct a thorough evaluation on their own patient population before integrating the model into their clinical workflow, our results suggest that the OncoNPC model may be applicable to a broader range of cancer centers, especially those with resource constraints that limit the number of genes that can be targeted in their sequencing panels.

5. Difference in OncoNPC prediction confidence across CUP and CKP cohorts

We investigated how OncoNPC prediction confidence varied across different cohorts, including CUP tumors at DFCI and MSK, as well as rare CKP tumors at DFCI. Specifically, we referred to the cohort of the rare CKP tumors whose cancer types were not considered during the development of OncoNPC as "DFCI excluded CKP tumors". We note that all cohorts in this analysis were not seen by OncoNPC during the model training. The results are presented in Extended Data Fig. 2b. As expected, the held-out test CKP tumors from cancer types that were considered in OncoNPC had the highest prediction confidence (mean 0.881, 95% C.I. 0.875 - 0.887), whereas held-out CKP tumors from cancer types that were not considered in OncoNPC (i.e., DFCI excluded CKP tumors) had the lowest prediction confidence (mean 0.674, 95% C.I. 0.667 - 0.681), suggesting that OncoNPC is able to properly down-weight the confidence for target cancer types that were not included in the OncoNPC model training. Interestingly, the overall prediction confidence of DFCI CUP tumors

(mean 0.764, 95% C.I. 0.75 - 0.778) and MSK CUP tumors (mean 0.761, 95% C.I. 0.743 - 0.779) were comparable to that of all DFCI CKP tumors including the rare tumors (mean 0.769, 95% C.I. 0.764 - 0.774). Though further study is merited, the similarity in prediction confidence between all CUP and CKP tumors suggests that CUP tumors may harbor rare primary cancers, resulting in a comparable mixture of known cancer types as seen in CKP tumors. Finally, we observed that among the held-out CKP tumor samples, prediction confidence was highest for tumors at DFCI, followed by tumors at MSK and VICC (DFCI: mean 0.88, 95% C.I. 0.875 - 0.887, MSK: mean 0.856, 95% C.I. 0.85 - 0.863, and VICC: mean 0.804, 95% C.I. 0.778 - 0.830; see Extended Data Fig. 2c).

6. Accuracy of germline imputation for panel-sequenced tumors

To impute common SNPs in the tumor data, ultra low coverage off-target reads and the 1000 Genomes Phase 3 reference panel with the STITCH algorithm were used [7]. This method has been extensively evaluated in previous research and was implemented using the default parameters that were previously published [8]. In the germline data, patients were genotyped on the Illumina Multi-Ethnic Global (MEG) SNP array and imputed to the 1000 Genomes Phase 3 reference panel using the Michigan Imputation Server. To ensure high-quality results, each cohort underwent variant-level quality control and was then limited to SNPs in HapMap3, which are typically well-imputed in GWAS and capture the majority of genetic variation [9].

To estimate the polygenic risk scores (PRS), common germline variants were first imputed directly from tumor-sequencing using the previously described approach that combines off-target reads and a haplotype reference panel [8]. Although this method has a mean SNP imputation accuracy of 0.86 across the genome, it may have biases towards on-target genes (where proximal coverage is high) and against deleted regions (where coverage is low). To evaluate the accuracy of the tumor off-target PRS, a subset of 1509 tumor samples from the Mass General Brigham (MGB) Biobank, including 57 CUP tumors, were used to obtain direct germline SNP array genotyping. The mean tumor-germline correlation was 0.92 (s.e. 0.006) for all tumors and 0.91 (s.e. 0.01) for CUP tumors across the cancer types for which PRS were calculated. No significant differences were observed for any individual cancer PRS (see Supplementary Fig. S7).

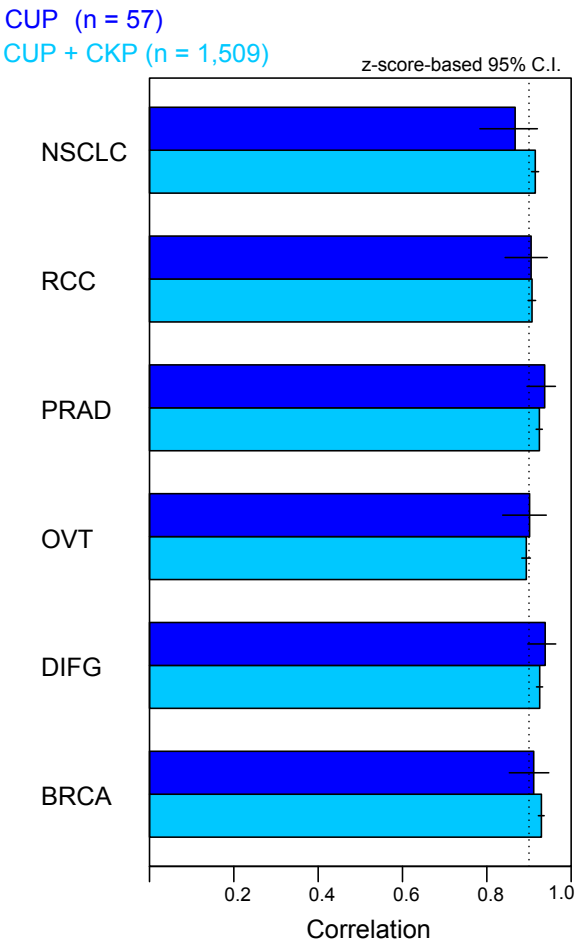


Figure S7: **Correlation between the germline PRS and imputed PRS across cancer types for which the PRS were calculated**. The light blue bars represent all CUP and CKP samples (n = 1,509), while the blue bars represent CUP samples (n = 57) with matched germline genotyping. The vertical dotted line at 0.9 serves as a reference point for correlation.

7. CUP-CKP metastatic survival comparison

We estimated median survival times of patients across CUP-metastatic CKP pairs using the Kaplan-Meier estimator [10] to account for patients lost to follow-up. For the CUP cohort, we excluded patients with CUP that were lost to follow up at the time of tumor sequencing and those whose primary cancer types were predicted with low probability (see the exclusion criteria in Extended Data Fig. 6). The resulting CUP cohort (n = 685), was then restricted to OncoNPC cancer types with more than 35 CUP patients. For the CKP metastatic cohort, we excluded patients lost to follow up at the tumor sequencing time in the same manner and chose patients with one of the known cancers, where either the biopsy was metastatic or the patient had an ICD-10 code indicative of secondary malignant neoplasms within a year prior to sequencing dates. A total of 521 and 5,937 patients were thus retained from the CUP cohort and metastatic CKP cohort, respectively: NSCLC ($n_{CUP} = 200$, $n_{met\text{-}CKP} = 1{,}559$), PAAD ($n_{CUP} = 80$, $n_{met\text{-}CKP} = 357$), BRCA ($n_{CUP} = 67$, $n_{met\text{-}CKP} = 1{,}656$), COADREAD ($n_{CUP} = 54$, $n_{met\text{-}CKP} = 1{,}198$), HNSCC ($n_{CUP} = 44$, $n_{met\text{-}CKP} = 216$), EGC ($n_{CUP} = 40$, $n_{met\text{-}CKP} = 336$), and OVT ($n_{CUP} = 36$, $n_{met\text{-}CKP} = 615$). Note that patients with CUP, whose predicted cancer type is GINET ($n_{CUP} = 39$, $n_{CKP} = 118$), were excluded due to the fact that the estimated survival function for the CUP cohort never reached 50 percent.

We then investigated if cancer-specific prognosis is shared between CUP predicted cancers and their corresponding CKP metastatic cancers. Utilizing overall survival data linked to the National Death Index and in-house follow-up data (see Methods), we found that median survival times of CUP-metastatic CKP pairs were significantly correlated across the cancer types (Spearman's ρ: 0.964, p-value: 4.54×10^{-4}; Main Fig. 4b). This significant relationship provides evidence that genetics-based OncoNPC predictions capture prognostic signals specific to each predicted cancer type. While correlated, median survival times were significantly lower for patients with CUP compared to those with metastatic CKP: CUP median survival 14.0 months (95% C.I. 11.9 - 15.8, n = 685) vs. metastatic CKP median survival 23.1 months (95% C.I. 21.8 - 24.2, n = 7,797). This is expected as CUPs are an advanced metastatic cancer with limited treatment options [11]. The absolute difference in median survival was significant across all predicted CUP - metastatic CKP pairs with the exception of Pancreatic Adenocarcinoma (CUP PAAD median survival 8.61 months 95% C.I. 5.09 - 10.8 vs. metastatic CKP PAAD median survival 6.73 months 95% C.I. 5.98 - 8.02), known to be a particularly deadly cancer type.

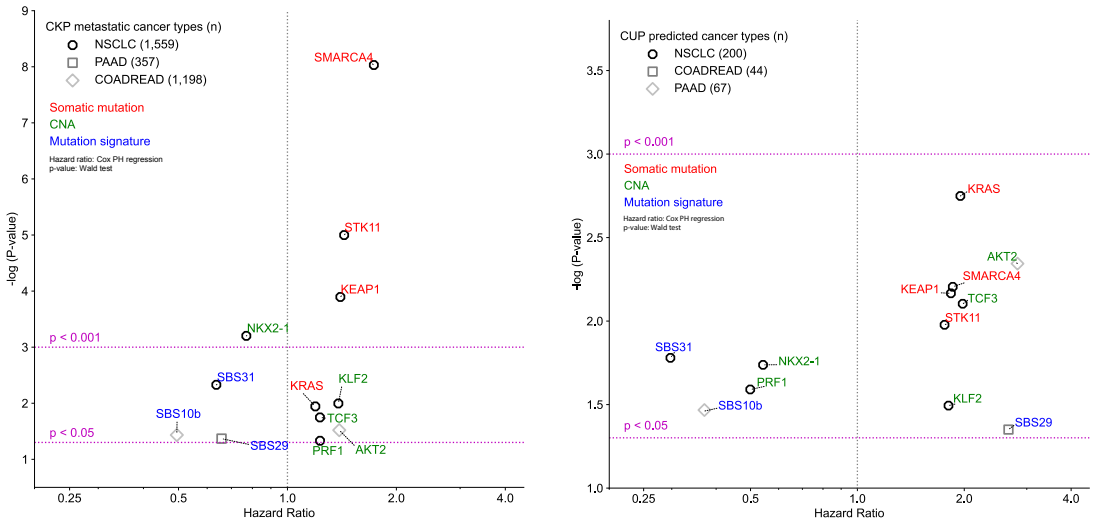
8. Prognostic somatic variants shared in CUP-metastatic CKP pairs

To identify prognostic somatic variants shared between CUP-metastatic CKP pairs, we restricted to the 7 common OncoNPC subgroups with at least 35 CUP tumors: NSCLC, PAAD, BRCA, COADREAD, HNSCC, EGC, GINET, and OVT. We utilized the processed input somatic features encoded in the OncoNPC model (see Methods). To ensure sufficient statistical power, we considered somatic features (i.e., mutated genes and CNA genes) present in at least 15 tumors in both the CUP and metastatic CKP cohorts, with each pair having the matching cancer type. Finally, we also considered all 96 mutational signature features.

After selecting the cancer types to consider in the CUP-metastatic CKP pairs and candidate

somatic variants for each pair, we iteratively tested each feature for association with survival in both CUP and corresponding metastatic CKP cohorts, where each pair had the matching cancer type. A multivariable Cox Proportional Hazard regression [12] model was used with time-to-death from sequencing as the outcome. To adjust for baseline effects, we included age at the time of sequencing, sex, tumor sequencing panel version, mutational burden (i.e., sum of total somatic mutations in each tumor sample), and CNA burden (i.e., sum of total CNA events in each tumor sample) as covariates. Finally, to identify shared prognostic somatic variants for each CUP-metastatic CKP pair, we retained somatic variants which passed Schoenfield residuals-based proportional hazard tests with p-value threshold of 0.05 (Python `lifelines` [13]) and were nominally significant (association p-value < 0.05) for both CUP and CKP cancer types in each pair.

Three out of 14 tested CUP-metastatic CKP pairs, specifically NSCLC, PAAD, and COAD-READ, showed shared prognostic somatic variants significantly associated with overall survival with nominal p-value cut-off at 0.05 (Supplementary Fig. S8a and S8b). In patients with known or classified NSCLC, three somatic mutations were associated with poor survival in both groups: *SMARCA4* (CUP: H.R. 1.86, 95% C.I. 1.19 - 2.89, p-value 6.23×10^{-3}, CKP mets: H.R. 1.73, 95% C.I. 1.44 - 2.09, p-value 9.30×10^{-9}), *STK11* (CUP: H.R. 1.76, 95% C.I. 1.14 - 2.71, p-value 1.05×10^{-2}, CKP mets: H.R. 1.43, 95% C.I. 1.22 - 1.68, p-value 1.00×10^{-5}), and *KEAP1* (CUP: H.R. 1.83, 95% C.I. 1.18 - 2.85, p-value 6.82×10^{-3}, CKP mets: H.R. 1.40, 95% C.I. 1.18 - 1.66, p-value 1.27×10^{-4}). These associations of somatic mutations in *SMARCA4*, *STK11*, and *KEAP1* genes with overall survival are well established for NSCLC [14–16]. Interestingly, a CNA event in *NKX2-1* was associated with improved survival in the patients from the NSCLC pair (CUP: H.R. 0.542, 95% C.I. 0.326 - 0.901, p-value 1.83×10^{-2}, CKP mets: H.R. 0.770, 95% C.I. 0.662 - 0.894, p-value 6.28×10^{-4}), consistent with prior meta-analyses [17]. In patients with known or classified COADREAD tumors, SBS10b mutation signature, linked to polymerase epsilon exonuclease domain mutations [18], was associated with longer overall survival (CUP: H.R. 0.371, 95% C.I. 0.148 - 0.928, p-value 3.41×10^{-2}, CKP mets: H.R. 0.495, 95% C.I. 0.255 - 0.958, p-value 3.68×10^{-2}). Finally, in patients with known or classified PAAD tumors, the SBS29 mutation signature (commonly found in tumor samples from individuals with a tobacco chewing habit [18]) was associated with poor survival in CUPs but nominally protective in metastatic CKPs (CUP: H.R. 2.66, 95% C.I. 1.02 - 6.93, p-value 4.46×10^{-2}, CKP mets: H.R. 0.657, 95% C.I. 0.438 - 0.986, p-value 4.28×10^{-2}). Although these somatic associations remain to be validated in independent cohorts, by categorizing patients with CUP based on their OncoNPC predictions, we were able to identify prognostic somatic variants, consistent with recent research findings.



(a) (b)

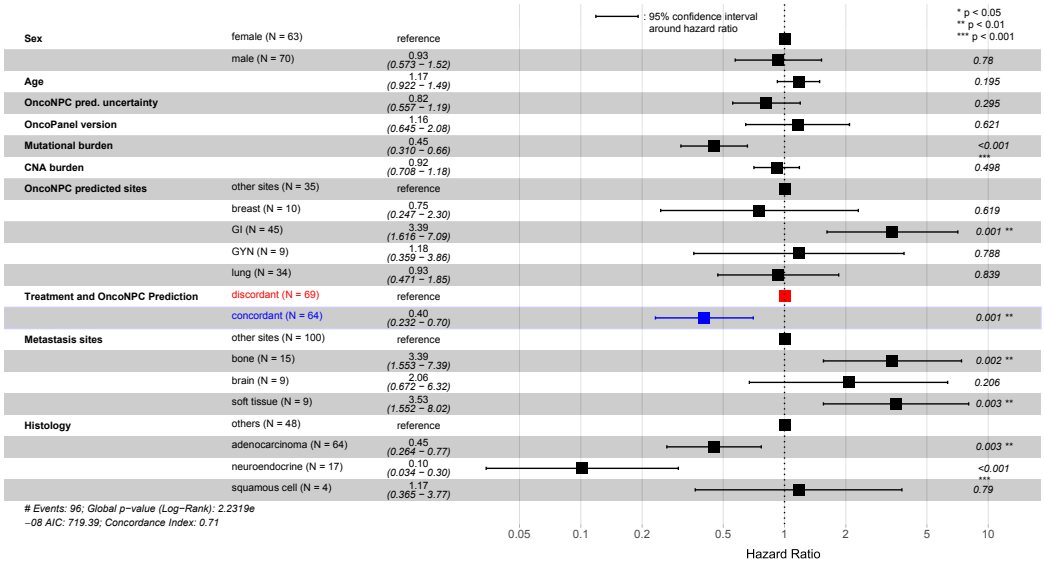
Figure S8: **Prognostic somatic biomarkers shared between CUP and metastic CKP tumors across OncoNPC predicted cancers. (a), (b)** Prognostic somatic variants significantly associated with overall survival, shared between three different CUP **(a)**-metastatic CKP **(b)** pairs (NSCLC: ○, PAAD: □, and COADREAD: ◇). Variant types are indicated by colors: red for somatic mutations, green for CNAs, and blue for mutation signatures. To estimate hazard ratio for each somatic biomarker, a multivariable Cox Proportional Hazard regression [12] model was used with time-to-death from sequencing as the outcome, while adjusting for baseline effects including age, sex, tumor sequencing panel version, mutational burden, and CNA burden.

9. Determining treatment-OncoNPC concordance

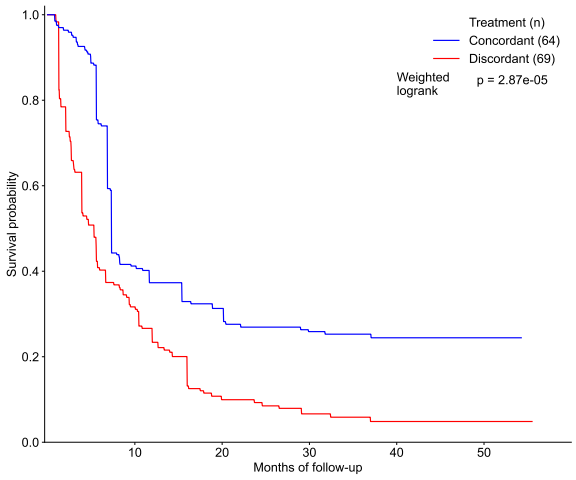
Concordance of OncoNPC predicted cancer type with the first palliative treatment assignment at DFCI was classified in one of four categories: 1) "TRUE": the OncoNPC cancer type matched the clinically proven/suspected tumor type and the treatment received, which was dictated by NCCN guidelines and/or standard of care, within the clinical context provided by the medical record; 2) "FALSE": the OncoNPC cancer type did not match the clinically proven/suspected cancer type and the treatment was not appropriate per NCCN guidelines or standard of care, in most reasonable situations, and within the context of the medical record; 3) "SOFT FALSE": the OncoNPC cancer type did not match the clinically proven/suspected cancer type, but the treatment received was not chosen based on NCCN guidelines or standard of care, owing to the unique clinical context provided by the medical record, 4) "EMPIRIC": treatment received was empiric treatment for cancer of unknown primary (e.g., carboplatin/taxol or gemcitabine/cisplatin) with the corresponding clinical rationale. In cases where patients received these regimens but not with the clinical intent of empiric CUP treatment (i.e., as regimens intended for treating other tumor types), the treatment was not labeled as "EMPIRIC" and the case was instead evaluated in context of the proven/suspected tumor type. In our analysis, we considered the TRUE group as the concordant group, and FALSE and SOFT FALSE groups as the discordant group. We did not include the EMPIRIC group, which is typically a more challenging patient population with systematically worse outcomes [19].

10. Additional discoveries in the treatment concordance analysis

Upon reassessing the charts of 158 patients in the cohort, there were some tumors in both groups where patients had a prior history of a known primary cancer, and the tumor being evaluated as a CUP was clinically felt to be a recurrence of that primary cancer: 13 patients (16.9%) in the concordant group and 12 patients (14.8%) in the discordant group. Notably, even after removing these samples from both groups, the analysis still showed a favorable outcome in the concordant group (see Supplementary Fig. S9; multivariable Cox regression: H.R. 0.400, 95% C.I. $0.232 - 0.700$, p-value 1.31×10^{-3} and IPTW Kaplan-Meier estimator: weighted log-rank test p-value 2.87×10^{-5}). This indicates that OncoNPC predictions continue to provide clinical benefits for the remaining CUP cases without a prior history of a known primary cancer.



(a)



(b)

Figure S9: **Estimating impact of the treatment-OncoNPC concordance on survival for patients with CUP who do not have a prior history of known primary cancers**. Among the original CUP treatment concordance analysis cohort (n = 158), we removed 13 patients (16.9%) in the concordant group and 12 patients (14.8%) in the discordant group who had history of a known primary cancer. Notably, after removing these samples from both groups, our results showed a favorable outcome in the concordant group: **(a)** Multivariable Cox regression: H.R. 0.400, 95% C.I. 0.232 – 0.700, p-value 1.31×10^{-3}, and **(b)** IPTW Kaplan-Meier estimator: weighted log-rank test p-value 2.87×10^{-5}.

The multivariable Cox regression (Main Fig. 5a) additionally identified significant hazardous effects of age, gastrointestinal (GI) cancer types predicted by OncoNPC, and bone metastasis (H.R. 1.27, 95% C.I. 1.02 – 1.58, p-value 3.10×10^{-2}, H.R. 4.20, 95% C.I. 2.06 – 8.55, p-value 7.78×10^{-5}, and H.R. 3.73, 95% C.I. 1.84 – 7.59, p-value 2.71×10^{-4}, respectively), and significantly protective effects

of tumor mutational burden (TMB), as well as adenocarcinoma and neuroendocrine tumor group determined by the histopathology results (H.R. 0.537, 95% C.I. 0.388 - 0.742, p-value 1.64×10^{-4}, H.R. 0.439, 95% C.I. 0.272 - 0.710, p-value 7.85×10^{-4} and H.R. 0.0854, 95% C.I. 0.0298 - 0.245, p-value 4.79×10^{-6}, respectively). In the IPTW Kaplan-Meier analysis, we found that treatment concordance with the OncoNPC prediction was associated with Gastrointestinal (GI) cancer types (coefficient 1.916, 95% C.I. 0.627 - 3.205, p-value 3.57×10^{-3}), whereas male sex and OncoNPC prediction uncertainty (i.e., entropy of predictive probability distribution over the considered cancer types) were inversely associated with receiving concordant treatment (coefficient -1.259, 95% C.I. -2.283 - -0.234, p-value 1.61×10^{-2}, and coefficient -1.693, 95% C.I. -2.458 - -0.927, p-value 1.46×10^{-5}; see Supplementary Fig. S10). These associations with treatment concordance are consistent with likely GI CUPs being more clinically identifiable and low OncoNPC confidence CUPs being less clinically and pathologically identifiable (see Supplementary Note 11.4 for a pathology-based evaluation of OncoNPC predictions for CUP tumors). We note, however, that the IPTW approach specifically adjusts for these systematic differences when estimating the effect of the treatment concordance on survival.

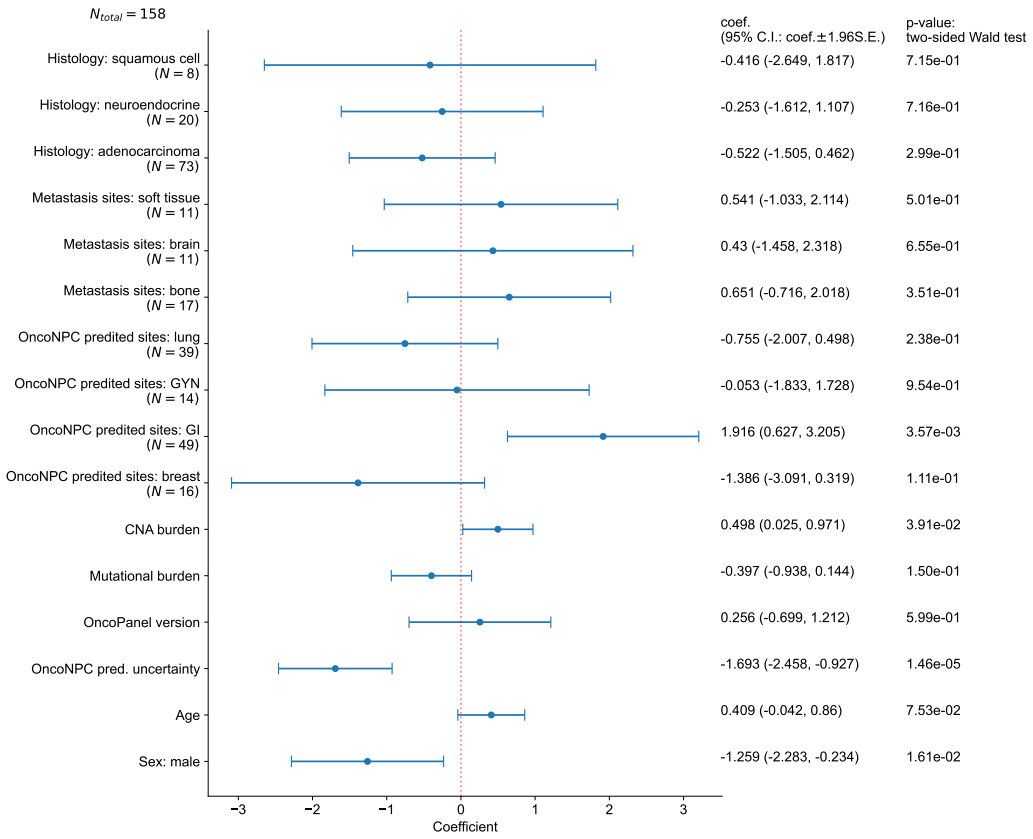


Figure S10: **Summary of coefficients for estimating treatment-OncoNPC concordance.** Formally, we estimated out-of-sample $P(A|X)$, where A corresponds to the treatment-OncoNPC concordance, using a logistic regression model in a 10-fold cross-fitting. The coefficients were obtained from the first fold. See Methods.

11. OncoNPC-guided actionable variants in patients with CUP

We investigated if OncoNPC classifications could inform genomically-guided, site-specific treatment options that are typically available for cancers with known primaries. We utilized OncoKB [20] as a knowledge base and considered three different categories of potentially actionable somatic variants: oncogenic mutation, amplification, and fusion (see Methods). OncoNPC cancer type predictions enabled identification of potentially actionable somatic variants across CUP tumor samples (total 22.8% of the eligible CUP tumor samples; see Extended Data Fig. 9). The majority of potentially actionable somatic variants for patients with CUP were oncogenic mutations (183 counts; 87.1%), followed by amplifications (22 counts; 9.52%) and fusions (7 counts; 3.33%) as shown in Extended Data Fig. 9a. The four most frequent oncogenic mutations were in *PIK3CA*, *KRAS*, *ALK*, and *ERBB2* genes, occurring in CUP tumor samples classified as BRCA (*PIK3CA* and *ERBB2* genes) and NSCLC (*KRAS*, *ALK*, and *ERBB2* genes). Overall, among the eligible CUPs whose prediction confidences are greater than 0.5 (N = 794; see Extended Data Fig. 6 for the exclusion criteria), OncoNPC predictions identified potentially actionable somatic variants for 11.5% of the CUP tumor samples for Level 1 therapeutic level (FDA-approved drugs), 3.63% for Level 2 (Standard care), 6.64% for Level 3 (Clinical evidence), and 1.00% for Level 4 (Biological evidence), summing up to the total 22.8% of the eligible CUP tumor samples (Extended Data Fig. 9b).

12. Performance of OncoNPC vs. gene expression and whole genome sequencing-based classifiers

OncoNPC vs. gene expression profiling (GEP)-based classifier

Gene expression profiling is a method that enables a comprehensive and holistic understanding of cellular behavior, providing insight into unique expression patterns specific to various cancer types. In contrast, targeted panel sequencing focuses on the detection of known cancer-associated genes and mutations within them. This approach typically involves a limited number of genes, ranging from tens to hundreds. For example, OncoPanel, a targeted panel sequencing platform at the DFCI, covers 447 key cancer genes.

We assessed the confidence of classifiers using targeted panel sequencing data versus gene expression profiling by comparing the proportions of confident samples in our cohort of unresolved CUP tumor samples (n = 158; see the exclusion criteria in Extended Data Fig. 6) to those of confident samples in a previous study that used a gene expression profiling-based classifier for clinicopathology-unresolved CUPs (n = 146) [21]. The gene expression profiling data in the previous study were obtained from a NanoString panel that targeted 225 genes differentially expressed across 18 tumor classes and viral transcripts encoding capsid proteins for HPV16 L1, HPV18 L1, and Merkel cell polyomavirus (VP2). The previous study found that high-medium confidence (>0.8) classifications were made for 56.2% (82/146) of cases, while in our cohort of 158 CUP cases, 96 patients had a prediction confidence for primary sites greater than 0.8 (60.8%; 95/158), which is comparable to the findings based on gene expression profiling. It is not possible to compare clinical and biological characteristics of patients with high confidence predictions across two classifiers. However, as dis-

cussed in [21], a DNA sequencing panel may be more diagnostically valuable than gene expression profiling for CUP tumors, as they often have an atypical transcriptional profile but retain diagnostic mutation features.

OncoNPC vs. whole genome sequencing (WGS)-based classifier

Whole genome sequencing (WGS) allows for the analysis of the entire genome, including all somatic and germline mutations, providing a comprehensive understanding of the mutational landscape of cancer cells. Unlike targeted panel sequencing, which captures mutations in a specific set of genes, WGS detects mutations in any region of the genome, providing information on regional mutational density features that cannot be accessed otherwise. A WGS-based classifier developed by [22] found that the distribution of somatic passenger mutations across the genome is the most predictive feature for cancer type prediction, reflecting chromatin accessibility to DNA repair complexes and relating to the epigenetic state of the cancer cell of origin.

This is further supported by the performance comparison of classifiers. Although the OncoNPC classifier based on targeted panel sequencing achieved an accuracy of 79.0% for held-out primary tumors (n = 4,525) and 79.8% for metastatic tumors (n = 2,605) across 22 different cancer types, the WGS-based classifier [22] outperformed OncoNPC for a larger number of cancer types. Specifically, the WGS-based classifier achieved an accuracy of 88% for held-out primary tumors (n = 1,436) across 24 cancer types and 83% for metastatic tumors (n = 2,028) across 16 cancer types.

Finally, we would like to note that comparing our targeted panel sequencing-based classifier, OncoNPC, to classifiers based on other sequencing technologies in an apple-to-apple manner is challenging due to differences in patient populations, modeled cancer types, and other technical factors such as the architecture of the machine learning models used. Nevertheless, we believe a key advantage of an NGS targeted panel sequencing-based algorithm is the widespread use of this assay in routine clinical practice as well as widely available commercial platforms. In particular, we see a major opportunity for the use of OncoNPC for CUP tumors in community clinics where access to detailed diagnostic evaluations and expert pathology reviews is limited.

13. Insights from chart reviews of 158 patients with CUP

Discordance between treatments and OncoNPC predictions in patients with CUP

In order to identify potential reasons behind the discordance, we have looked at pathological findings as well as other factors such as ECOG performance status and issues pertaining to insurance or access to medical care. A trained oncologist reviewed the medical chart records of all 158 patients in the cohort at DFCI, and we found that there was no significant difference in ECOG performance status between the treatment-OncoNPC prediction concordant and discordant groups. The median ECOG was 1 for both groups, and the mean ECOG was 0.816 and 0.778 for the concordant and discordant groups, respectively. Only one patient in the discordant group received emergency and empiric treatment due to superior vena cava syndrome, but otherwise, all patients in the discordant group were treated without any extenuating circumstances, insurance issues, or access barriers.

Diagnostic utility of OncoPanel identified mutation signatures

After reviewing the genomic specimen reports for the patients with CUP in the cohort, we have identified Tobacco, APOBEC, and UVA mutations signatures in 4 and 14 patients belonging to the concordant (Tobacco = 2, APOBEC = 0, UVA = 2) and discordant (Tobacco = 4, APOBEC = 4, UVA = 6) groups, respectively. We have provided details on the OncoPanel identified mutation signatures and actionable variants for each patient in Supplementary Data 5. Based on the patient chart reviews, we found that out of the 18 cases with detectable mutation signatures (UVA, tobacco, and APOBEC), only 5 cases (27.7%) were diagnostically useful. Notably, only the UVA and tobacco signatures were helpful in identifying cutaneous and thoracic malignancies, respectively, while the APOBEC signature was not useful in diagnosis. In two cases (11.1%), the presence of a UVA signature directly led to the diagnosis of melanoma and subsequent use of immunotherapy. In another case, a patient who had never smoked and presented with a pulmonary mass and visceral metastases was initially thought to have lung cancer; however, OncoPanel testing showed a strong UV damage signal, favoring a diagnosis of melanoma over non-small cell lung cancer. The patient was subsequently treated with pembrolizumab, which led to significant improvement in all disease sites after three cycles. In this patient's tumor sample, we would like to highlight that OncoNPC accurately predicted melanoma.

Retrospective pathology-based evaluation of OncoNPC predictions

To assess the concordance of OncoNPC prediction with pathology-based suspected primaries, we retrospectively reviewed available pathological data containing histology and immunohistochemistry information. A certified oncologist at the DFCI reviewed the pathology notes, containing histologic and morphologic features, immunohistochemical stains, and integration of these features by a subspecialized (disease-specific) pathologist, for the 158 patients with CUP in the treatment-concordance analysis (see Extended Data Fig. 6 for the exclusion criteria). It is worth noting that these patients had received comprehensive diagnostic work-up and were treated at DFCI. We were able to identify top 3 pathology-based suspected primaries for 129 of these patients, and among them, the top 5 most frequently suspected primaries were Gastrointestinal (26.8%), Pancreatic (18.3%), Lung (11.3%), Cholangiocarcinoma (11.3%), and Gynecologic (mullerian, ovary, and endometrial; 6.57%).

Subsequently, we examined the level of agreement between the OncoNPC primary site prediction and the top 3 suspected primary sites based on pathology. For this analysis, we considered the prediction to be concordant if it matched any of the top 3 suspected primary sites, and discordant if it did not. Overall, we found that there was pathology-OncoNPC concordance in 67 out of 129 cases (51.9%), significantly higher than the top-3 concordance of randomly selected OncoNPC predictions (19.9%, 95% C.I. 19.7% - 20.1%). Within the treatment-OncoNPC concordant group (n = 64), we found the pathology-OncoNPC concordance in 53 out of 64 cases (82.8%, 95% C.I. 73.6% - 92.1%), significantly higher than the concordance observed in the treatment-OncoNPC discordant group (n = 65), which was only 14 out of 65 cases (21.5%, 95% C.I. 11.5% - 31.5%). It is worth noting that the higher pathology-OncoNPC concordance in the treatment concordant group is expected due to the definition of the group. Intriguingly, the 14 putatively concordant cases in the treatment

discordant group are likely instances where a primary was suspected by pathology but did not factor into the ultimate treatment decision. Overall, the relatively low overall concordance rate highlights the difficulties in diagnosing highly metastatic or poorly differentiated tumors through pathology, as previously mentioned in [2, 23].

Next, we compared the level of uncertainty in OncoNPC predictions, measured by entropy, in two groups of patients. The first group comprised patients with suspected primaries based on pathology findings (n = 129), while the second group consisted of patients with no suspected primary site (n = 29), indicating inconclusive pathology results and an additional degree of pathological complexity. Interestingly, we found no significant difference in prediction uncertainty (0.770, 95% C.I. 0.664 - 0.876 and 0.796, 95% C.I. 0.522 - 1.07 for those with and without pathology-based suspected primaries, respectively; note higher means the prediction is less certain), suggesting that tumors without IHC-based suspected primaries were not more challenging to classify genetically. Additionally, we compared the prediction uncertainty between the pathology-OncoNPC discordant and concordant groups (n = 62 and n = 67, respectively), where CUP tumors were defined as pathology-OncoNPC concordant if the OncoNPC prediction matched any of the top 3 suspected primary sites (and discordant otherwise). The discordant group had significantly higher prediction uncertainty (0.958, 95% C.I. 0.809 - 1.11 for the discordant group vs. 0.597, 95% C.I. 0.458 - 0.735 for the concordant group), indicating that discordant tumors were more challenging to classify genetically.

Finally, we conducted a retrospective investigation into the diagnostic utility of OncoNPC predictions for CUP tumors when multiple primaries are suspected based on pathology findings. We present two case studies here:

In one case, a patient had urothelial cancer years prior to presenting with an ampullary mass. Based on initial pathological findings, it was initially thought to be an intestinal type ampullary carcinoma and treated with first-line chemotherapy (FOLFOX regimen) per NCCN guidelines. However, further examination showed that the mass was more likely consistent with urothelial cancer, and this was confirmed by the positive response to the NCCN-recommended first-line therapy for locally advanced or metastatic bladder cancer (i.e., gemcitabine and cisplatin). OncoNPC predicted Bladder Urothelial Carcinoma for this patient's tumor with a high prediction confidence of 0.999.

In another case, a patient with a history of bilateral breast cancers presented with a lung lesion of unknown primary. OncoNPC predicted Invasive Breast Carcinoma for this patient's tumor with a confidence of 0.996, while pathology findings suggested possible recurrence of breast cancer, primary lung adenocarcinoma, or mesothelioma. The case was ultimately treated as lung cancer with immunotherapy, but the patient had a poor response and passed away two months later, raising the possibility of a primary breast cancer.

Table S2: Demographic details of patients with CUP in the concordant and discordant treatment groups, who *received their initial treatments after the results of the OncoPanel sequencing were available to clinicians*. The OncoNPC predicted cancer groups, except for the GI group, match the cancer groups defined in Main Table 1. The GI group in this analysis comprises the upper GI group, consisting of Cholangiocarcinoma (CHOL), Esophagogastric Adenocarcinoma (EGC), and Pancreatic Adenocarcinoma (PAAD), as well as Colorectal Adenocarcinoma (COADREAD).

	Concordant treatment group (n = 21)	**Discordant treatment group (n = 12)**
Sex; male-female ratio	0.429-0.571	0.417-0.583
Age at seqeuncing (95% C.I.)	65 (60.4 - 69.6)	62 (57.3 - 66.7)
OncoNPC prediction uncertainty (in entropy; 95% C.I.)	0.517 (0.253 - 0.782)	0.862 (0.484 - 1.24)
OncoPanel version (proportion in %)		
v1	1 (4.762%)	
v2	2 (9.524%)	1 (8.333%)
v3	18 (85.714%)	10 (83.333%)
Mutational burden (95% C.I.)	0.029 (0.016 - 0.043)	0.031 (0.02 - 0.042)
CNA burden (95% C.I.)	0.231 (0.175 - 0.286)	0.256 (0.2 - 0.313)
OncoNPC predicted cancer groups (proportion in %)		
Lung	5 (23.81%)	3 (25.0%)
Breast	1 (4.762%)	4 (33.333%)
GI	3 (14.286%)	
Gyn	4 (19.048%)	1 (8.333%)
Others	8 (38.095%)	4 (33.333%)
Metastatic sites (proportion in %)		
Brain	2 (9.524%)	2 (16.667%)
Bone	3 (14.286%)	
Soft tissue		1 (8.333%)
Others	16 (76.19%)	9 (75.0%)
Histology (proportion in %)		
Adenocarcinoma	9 (42.857%)	4 (33.333%)
Neuroendocrine	4 (19.048%)	
Squamous cell	1 (4.762%)	1 (8.333%)
Others	7 (33.333%)	7 (58.333%)
Treatment start date (95% C.I.)	2018-10-26 (2018-3-5 - 2019-6-18)	2018-7-2 (2017-9-9 - 2019-4-24)

Table S3: Center-specific number of held-out CKP tumor samples, broken down by cancer types and prediction confidences (i.e., $p_{\max}$) thresholds.

		Minimum $p_{\max}$ threshold			
		0.0	0.5	0.7	0.9
Overall	DFCI	3690	3438	3047	2502
	MSK	3331	3012	2608	2112
	VICC	268	230	192	136
Non-Small Cell Lung Cancer (NSCLC)	DFCI	811	735	644	533
	MSK	717	618	520	430
	VICC	36	27	23	19
Invasive Breast Carcinoma (BRCA)	DFCI	600	572	514	433
	MSK	727	675	598	474
	VICC	68	62	48	35
Colorectal Adenocarcinoma (COADREAD)	DFCI	521	502	479	436
	MSK	375	358	330	303
	VICC	55	52	48	37
Diffuse Glioma (DIFG)	DFCI	400	390	383	361
	MSK	214	204	187	168
	VICC	11	10	8	4
Prostate Adenocarcinoma (PRAD)	DFCI	126	118	98	67
	MSK	300	280	233	163
	VICC	16	10	6	3
Pancreatic Adenocarcinoma (PAAD)	DFCI	136	125	104	71
	MSK	233	216	187	154
	VICC	10	8	6	1
Ovarian Epithelial Tumor (OVT)	DFCI	257	229	184	112
	MSK	100	60	38	10
	VICC	12	9	5	2
Esophagogastric Adenocarcinoma (EGC)	DFCI	171	153	114	66
	MSK	82	70	44	24
	VICC	11	8	7	2
Endometrial Carcinoma (UCEC)	DFCI	123	116	95	73
	MSK	105	100	91	70
	VICC	7	6	6	2
Melanoma (MEL)	DFCI	134	127	115	103
	MSK	108	103	98	92
	VICC	24	24	23	23
Bladder Urothelial Carcinoma (BLCA)	DFCI	86	81	67	52
	MSK	93	84	78	65
	VICC	4	4	4	3

		Minimum $p_{\max}$ threshold			
		0.0	0.5	0.7	0.9
Renal Cell Carcinoma (RCC)	DFCI	79	71	61	50
	MSK	85	75	68	56
	VICC	6	5	4	3
Head and Neck Squamous Cell Carcinoma (HNSCC)	DFCI	55	50	39	28
	MSK	27	18	12	5
	VICC	.	.	.	.
Cholangiocarcinoma (CHOL)	DFCI	18	12	10	7
	MSK	40	31	24	16
	VICC	1	.	.	.
Gastrointestinal Stromal Tumor (GIST)	DFCI	47	46	43	40
	MSK	34	33	31	30
	VICC	.	.	.	.
Well-Differentiated Thyroid Cancer (WDTC)	DFCI	17	15	14	9
	MSK	31	31	29	25
	VICC	1	1	1	.
Pleural Mesothelioma (PLMESO)	DFCI	24	21	14	10
	MSK	18	17	10	6
	VICC	5	3	2	1
Meningothelial Tumor (MNGT)	DFCI	27	25	23	20
	MSK	3	3	1	.
	VICC	1	1	1	1
Gastrointestinal Neuroendocrine Tumors (GINET)	DFCI	20	17	16	11
	MSK	3	3	2	.
	VICC	.	.	.	.
Pancreatic Neuroendocrine Tumor (PANET)	DFCI	15	14	13	8
	MSK	24	22	19	15
	VICC	.	.	.	.
Acute Myeloid Leukemia (AML)	DFCI	15	11	10	6
	MSK	.	.	.	.
	VICC	.	.	.	.
Non-Hodgkin Lymphoma (NHL)	DFCI	8	8	7	6
	MSK	12	11	8	6
	VICC	.	.	.	.

References

[1] K. Oien and J. Dennis, "Diagnostic work-up of carcinoma of unknown primary: From immunohistochemistry to molecular profiling," *Annals of Oncology*, vol. 23, pp. x271–x277, 2012.

[2] G. R. Varadhachary and M. N. Raber, "Cancer of unknown primary site," *New England Journal of Medicine*, vol. 371, no. 8, pp. 757–765, 2014.

[3] E. Shadmi, N. Flaks-Manov, M. Hoshen, O. Goldman, H. Bitterman, and R. D. Balicer, "Predicting 30-day readmissions with preadmission electronic health record data," *Medical care*, vol. 53, no. 3, pp. 283–289, 2015.

[4] E. Özyılmaz *et al.*, "Worse pre-admission quality of life is a strong predictor of mortality in critically ill patients," *Turkish Journal of Physical Medicine and Rehabilitation*, vol. 68, no. 1, p. 19, 2022.

[5] D. A. Vyas, L. G. Eisenstein, and D. S. Jones, *Hidden in plain sight—reconsidering the use of race correction in clinical algorithms*, 2020.

[6] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable ai for trees," *Nature machine intelligence*, vol. 2, no. 1, pp. 56–67, 2020.

[7] R. W. Davies, J. Flint, S. Myers, and R. Mott, "Rapid genotype imputation from sequence without reference panels," *Nature genetics*, vol. 48, no. 8, pp. 965–969, 2016.

[8] A. Gusev, S. Groha, K. Taraszka, Y. R. Semenov, and N. Zaitlen, "Constructing germline research cohorts from the discarded reads of clinical tumor sequences," *Genome medicine*, vol. 13, pp. 1–14, 2021.

[9] B. K. Bulik-Sullivan *et al.*, "Ld score regression distinguishes confounding from polygenicity in genome-wide association studies," *Nature genetics*, vol. 47, no. 3, pp. 291–295, 2015.

[10] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.

[11] A. M. Conway, C. Mitchell, E. Kilgour, G. Brady, C. Dive, and N. Cook, "Br J CancerMolecular characterisation and liquid biomarkers in Carcinoma of Unknown Primary (CUP): taking the 'U' out of 'CUP'," *Br J Cancer*, vol. 120, no. 2, pp. 141–153, Jan. 2019.

[12] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972.

[13] C. Davidson-Pilon, "Lifelines: Survival analysis in python," *Journal of Open Source Software*, vol. 4, no. 40, p. 1317, 2019.

[14] A. J. Schoenfeld *et al.*, "The genomic landscape of smarca4 alterations and associations with outcomes in patients with lung cancersmarca4 alterations in lung cancer," *Clinical Cancer Research*, vol. 26, no. 21, pp. 5701–5708, 2020.

[15] S. Papillon-Cavanagh, P. Doshi, R. Dobrin, J. Szustakowski, and A. M. Walsh, "Stk11 and keap1 mutations as prognostic biomarkers in an observational real-world lung adenocarcinoma cohort," *ESMO open*, vol. 5, no. 2, e000706, 2020.

[16] T. Takahashi *et al.*, "Mutations in keap1 are a potential prognostic factor in resected non-small cell lung cancer," *Journal of surgical oncology*, vol. 101, no. 6, pp. 500–506, 2010.

[17] T. Berghmans *et al.*, "Thyroid transcription factor 1—a new prognostic factor in lung cancer: A meta-analysis," *Annals of oncology*, vol. 17, no. 11, pp. 1673–1676, 2006.

[18] J. G. Tate *et al.*, "Cosmic: The catalogue of somatic mutations in cancer," *Nucleic acids research*, vol. 47, no. D1, pp. D941–D947, 2019.

[19] K. Fizazi, F. Greco, N. Pavlidis, G. Daugaard, K. Oien, and G. Pentheroudakis, "Cancers of unknown primary site: Esmo clinical practice guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 26, pp. v133–v138, 2015.

[20] D. Chakravarty *et al.*, "Oncokb: A precision oncology knowledge base," *JCO precision oncology*, vol. 1, pp. 1–16, 2017.

[21] A. Posner *et al.*, "A comparison of dna sequencing and gene expression profiling to assist tissue of origin diagnosis in cancer of unknown primary," *The Journal of Pathology*, vol. 259, no. 1, pp. 81–92, 2023.

[22] W. Jiao *et al.*, "A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns," *Nature communications*, vol. 11, no. 1, p. 728, 2020.

[23] G. G. Anderson and L. M. Weiss, "Determining tissue of origin for metastatic cancers: Meta-analysis and literature review of immunohistochemistry performance," *Applied Immunohistochemistry & Molecular Morphology*, vol. 18, no. 1, pp. 3–8, 2010.