Pan-cancer Transcriptomic Predictors of Perineural Invasion Improve Occult Histopathologic Detection

HHS Public Access

Author manuscript

Clin Cancer Res. Author manuscript; available in PMC 2021 November 15.

Published in final edited form as:

Clin Cancer Res. 2021 May 15; 27(10): 2807–2815. doi:10.1158/1078-0432.CCR-20-4382.

Pan-cancer transcriptomic predictors of perineural invasion improve occult histopathological detection

Jimmy A. Guo[1,2,3,4,5,6,†], **Hannah I. Hoffman**[1,2,3,†], **Stuti Shroff**[7], **Peter Chen**[8], **Peter G. Hwang**[9], **Daniel Y. Kim**[10], **Daniel W. Kim**[3], **Stephanie W. Cheng**[11], **Daniel Zhao**[12], **Brandon A. Mahal**[13], **Mohammed Alshalalfa**[14], **Andrzej Niemierko**[3], **Jennifer Y. Wo**[3], **Jay S. Loeffler**[3], **Carlos Fernandez-del Castillo**[15], **Tyler Jacks**[2], **Andrew J. Aguirre**[1,6], **Theodore S. Hong**[3], **Mari Mino-Kenudson**[7], **William L. Hwang**[1,2,3]

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[2]Koch Institute for Integrative Cancer Research and Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA.

[3]Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA, USA.

[4]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, USA.

[5]School of Medicine, University of California, San Francisco, San Francisco, CA, USA.

[6]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA.

[7]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA.

[8]Raytheon Technologies, Brooklyn, NY, USA.

[9]Massachusetts Institute of Technology, Cambridge, MA, USA.

[10]Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA, USA.

[11]Harvard University, Cambridge, MA, USA.

[12]New York Medical College, Valhalla, NY, USA.

[13]Department of Radiation Oncology, Miller School of Medicine, Miami, FL, USA.

[14]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA.

[15]Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts.

Correspondence: William L. Hwang, MD, PhD, Department of Radiation Oncology, Massachusetts General Hospital, 100 Blossom Street, MA 02114, whwang1@mgh.harvard.edu, Tel: 617-726-6050, Fax: 617-726-3603.

[†]These authors contributed equally: Jimmy A. Guo, Hannah I. Hoffman.

Author contributions

J.A.G., H.I.H., and W.L.H. developed the concept and designed the study. J.A.G., H.I.H, and D.Y.K. collected histopathological and clinical data from the TCGA GDC Data Portal and cBioPortal. J.A.G., H.I.H., W.L.H. analyzed the RNA-seq data and classifier gene lists with computational assistance from P.C. and P.G.H. P.C. and P.G.H. used machine learning approaches to develop classifiers for PNI. S.S. re-reviewed pathology slides for occult PNI analysis. H.I.H., B.A.M., M.A., A.N., and W.L.H. guided and performed statistical analyses for survival. J.A.G., H.I.H., P.C., P.G.H., S.W.C., and D.Z. generated figures and tables. D.W.K., B.A.M., J.Y.W., J.S.L., C.F.C., T.J., A.J.A., T.S.H., and M.M.K. provided scientific, clinical, and/or pathological input. W.L.H. supervised the research. J.A.G., H.I.H., and W.L.H. wrote the manuscript. All authors reviewed the manuscript.

Abstract

Purpose: Perineural invasion (PNI) is associated with aggressive tumor behavior, recurrence, and metastasis, and can influence the administration of adjuvant treatment. However, standard histopathological examination has limited sensitivity in detecting PNI and does not provide insights into its mechanistic underpinnings.

Experimental Design: A multi-variate Cox regression was performed to validate associations between PNI and survival in 2029 patients across 12 cancer types. Differential expression and gene set enrichment analysis were used to learn PNI-associated programs. Machine learning models were applied to build a PNI gene expression classifier. A blinded re-review of H&E slides by a board-certified pathologist helped determine whether the classifier could improve occult histopathological detection of PNI.

Results: PNI associated with both poor OS (hazard ratio, 1.73; 95% CI, 1.27–2.36; $P < 0.001$) and DFS (hazard ratio, 1.79; 95% CI, 1.38–2.32; $P < 0.001$). Neural-like, pro-survival, and invasive programs were enriched in PNI-positive tumors ($P_{adj} < 0.001$). Although PNI-associated features likely reflect in part the increased presence of nerves, many differentially-expressed genes mapped specifically to malignant cells from single-cell atlases. A PNI gene expression classifier was derived using random forest and evaluated as a tool for occult histopathological detection. On a blinded H&E re-review of sections initially described as PNI-negative, more specimens were re-annotated as PNI-positive in the high classifier score cohort compared to the low-scoring cohort ($P = 0.03$, Fisher's exact test).

Conclusions: This study provides salient biological insights regarding PNI and demonstrates a role for gene expression classifiers to augment detection of histopathological features.

Keywords

Perineural invasion; cancer; occult detection; machine learning; transcriptomic classifier

Introduction

While hematogenous and lymphatic routes have long been recognized as conduits for tumor dissemination, the complicity of nerves in promoting malignant phenotypes has come into focus more recently (1–4). It is now known that tumor initiation and development are often accompanied by an increase in intra- and peri-tumoral nerve density and that disruption of crosstalk between nerves and cancer cells can impede tumor progression and extend overall survival in animal models (3,5). Perineural invasion (PNI) is a manifestation of this dynamic interplay, defined by a reciprocal tropism that results in neoplastic infiltration in, around, or through nerve bundles (6). Its presence has been noted during routine histological examination in numerous cancers, including head & neck, pancreatic, colorectal, prostate, and breast, suggesting that PNI may be a conserved phenomenon with benefits to tumor fitness. Indeed, prior literature has shown that PNI in many cancer types is associated with aggressive tumor behavior, higher rates of post-treatment recurrence, metastases, pain, and reduced survival (6).

Despite these clinical ramifications, biological mechanisms contributing to and resulting from PNI are incompletely understood, hindering the rational development of therapeutic interventions. Prior studies exploring PNI-associated genes in pancreatic and prostate cancer (7–10) have pinpointed those encoding neurotrophic factors (e.g. *BDNF*, *GDNF*, *NT1/2*, *NGF*) and chemokines (e.g. *CX3CL1*, *SEMA3A*, *CXCL16*), among others, but many of these studies were underpowered by small patient cohorts of individual cancer types. Moreover, PNI is often overlooked during histopathological analyses due to tissue sampling bias, suboptimal staining techniques, and lack of standardized reporting guidelines for certain cancer types (6,11,12). The high frequency of occult cases has been demonstrated in prior studies; for example, review of head & neck cancer specimens using nerve-specific antibodies instead of H&E increased the number of PNI positive cases by nearly three-fold (6,12). While nerve-specific stains may be helpful, they depend on antigen availability and the sensitivity and specificity of the antibodies used. In addition, it would be resource-intensive to perform the multitude of antigen-specific assays that would be necessary to extensively profile the various tumor features that are clinically relevant. In light of these challenges, global mRNA profiling through RNA-seq or microarray technologies should have a versatile role in clinical oncology within the next decade (13,14) and may guide the development of personalized therapeutics, drive real-time clinical decision making, and assist with the identification of histological features or micrometastatic lesions that are otherwise difficult to detect by visual inspection alone.

We created a custom cohort of 2029 patients across 12 cancer types from The Cancer Genome Atlas (TCGA) and leveraged surgical pathology reports and bulk transcriptomic data to identify the fundamental gene expression programs that associate with, differentiate, and predict PNI, irrespective of tumor origin. We found numerous genes and gene sets up- or down-regulated in patients with PNI, many of which had not been previously identified. We then overlaid these genes onto single-cell atlases of 29 patients with head & neck and colorectal cancers (15,16) to identify their likely cellular sources. Furthermore, to determine the utility of differentially-expressed genes for prediction of PNI status, we leveraged machine learning approaches to develop a novel gene expression classifier. We performed a blinded re-review of tissue sections initially annotated as PNI-negative and found that our classifier could indeed improve the detection of occult cases. Taken together, these findings may inform clinically-deployable diagnostic strategies and also guide therapeutic approaches targeted at PNI.

Materials and Methods

Data collection and usage

Data on PNI and LVI, gene expression, demographics and survival from a total of 2029 patients across 12 cancer sites (bile duct, bladder, breast, colorectal, esophageal, head & neck, liver, pancreas, prostate, stomach, thyroid, uterus) were collected. PNI and LVI annotations were extracted from surgical pathology reports provided by The Cancer Genome Atlas (TCGA) provisional studies on cBioPortal (17,18), while data on patient characteristics (age, cancer type, race, sex), clinical outcomes (disease-free survival from initial treatment [DFS], overall survival from diagnosis [OS]), and bulk transcriptomics

(HTSeq-Counts) were collected from the National Cancer Institute's GDC Data Portal. Single-cell RNA-seq data of head & neck and colorectal cancer (15,16) were obtained using accession IDs GSE103322 and GSE81861, respectively. The data used were formerly obtained with written and informed consent from patients; those studies had been approved by an institutional review board and conducted in accordance with ethical guidelines (U.S. Common Rule).

Patients with missing OS or DFS data were excluded from univariate and multivariate survival analyses. Patients were included in the differential expression analyses only if they had both PNI annotations and RNA-sequencing data. All patients with annotated PNI statuses and RNA-sequencing, irrespective of survival data availability, were included for classifier development. Classifier scores were additionally generated for patients with RNA-seq data but without PNI annotations. No data imputation was performed.

Differential gene expression analyses

Differential gene expression analyses using DESeq2 (Bioconductor) were performed on the following subsets of patients: PNI-positive *vs.* PNI-negative and $PNI^{+}LVI^{-}$ *vs.* $PNI^{-}LVI^{+}$, while adjusting for age, sex, race, and cancer type. Trimmed means were used to account for outliers. Each gene from DEseq2 was assigned a $\log_2$fold change value as well as a *P*-value adjusted by a Benjamini-Hochberg correction to account for the false discovery rate. Genes with an adjusted *P*-value of less than 0.001 for PNI-positive *vs.* PNI-negative and less than 0.005 for $PNI^{+}LVI^{-}$ *vs.* $PNI^{-}LVI^{+}$ were inputted into the Molecular Signatures Database (MSigDB) v7.1 for gene set enrichment analyses (GSEA) using an FDR q-value threshold of 0.05.

Single-cell analysis

Single cell expression of genes for head & neck cancer was quantified as $E_{i,j}=log_2(TPM_{i,j}/10+1)$, in accord with the prior study (15), in which $TPM_{i,j}$ represents transcripts-per-million for gene i in sample j as calculated by RSEM. Fragments per kilobase of transcript, per million mapped reads (FPKM) was used as the normalization method for colorectal cancer, in accord with the original study (16). Cell type assignments were maintained from the prior studies. A gene was considered specific to a cell type if its expression in that cell type was at least one standard deviation higher than its average expression across all cell types. Minimum cutoffs of 0.1 for $E_{i,j}$ and 1 for FPKM were used to preclude cell type assignments for insignificantly expressed genes.

Machine learning classification of perineural invasion

The expression (transcripts per million) of significant genes ($P_{adj} < 0.001$, 1030 features) identified by DESeq2 was used in a grid search workflow to screen eight different models (random forest [RF], logistic regression, XGBoost, decision tree, k-nearest neighbors, extra random trees, support vector machine, stochastic gradient descent) and various hyperparameters for their performance on PNI classification, as determined by the area under the curve (AUC) of the receiver operating characteristic (ROC). We tuned the number of trees, maximum depth of tree, and minimum samples per leaf for RF; the parameter 'C' (regularization constant) for logistic regression; the maximum depth of tree, learning rate,

and L2 regularization magnitude for XGBoost; maximum depth for decision tree; the parameter 'C' (regularization constant) for support vector machine; 'alpha' (regularization constant) for stochastic gradient descent; number of neighbors for k-nearest neighbors; and the number of trees, maximum depth of tree, and minimum samples per leaf for extra random trees. The RF model was further tuned following the initial grid search. We also iterated on the adjusted *p*-value cutoff from DESeq2 to determine if including a greater number of differentially-expressed genes ($P_{adj} > 0.001$) could improve classification. A random 80–20 split of our patient dataset was used for training and validation, respectively, when generating the RF classifier. To validate that the initial model generated from a random split was representative, 10 additional train-test splits were performed and assessed. A cutoff score for differentiating between PNI-positive and PNI-negative predictions was determined by optimizing for the maximum F_1 score.

Detection of occult perineural invasion

To assess the ability of our PNI random forest classifier to identify occult cases among the PNI-negative subset, we performed a power calculation to estimate the required sample size using $(1-\beta) = 0.9$ and $\alpha = 0.05$. Recall and precision parameters were used to compute occult rates while assuming large sample size to estimate prevalence. Approximations for expected histopathological review "true positive" rates in the high classifier score group were extracted from relevant literature. A review "false positive" rate of 10% was assumed for the low classifier score group to provide a conservative sample size estimate.

Following estimation of the required sample size, PNI scores were assigned to all GI (colorectal, esophageal, pancreatic, gastric) cancer patients initially annotated as PNI-negative. Scores were generated by outputting the probability of PNI as predicted by the random forest classifier. The 33 lowest-scoring patients with at least two tissue sections were selected to comprise the low-scoring group (scores < 0.4), and conversely, the 33 highest-scoring patients with at least two tissue slides were selected to comprise the high-scoring group (scores > 0.6).

Finally, raw, unprocessed tissue slide images from the 66 GI cancer patients were re-reviewed by a board-certified GI pathologist blinded to the PNI scores (range 0–1). This was performed by systematically reviewing each scanned slide in a grid-like fashion at high magnification and focusing on the pancreatic parenchyma, peripancreatic soft tissue and the bowel wall. In the bowel wall, the pathologist focused on the submucosa and the myenteric plexus between two layers of muscularis propria to assess for PNI. Within these foci and other foci, PNI was deemed to be present if neoplastic glands were found to be in direct contact with the neural axons (i.e., within the perineurium, or if there was any intraneural invasion). If there were any stromal fibroblasts between the nerve and tumor cells, those foci were not regarded as PNI.

Statistical analyses

Multivariate Cox proportional hazards models were used to estimate the associations between PNI with OS and DFS, stratifying by cancer type and including age, sex, and race as co-variates. We conducted one-way ANOVA followed by Tukey's HSD post-hoc

comparisons on RF classifier scores – defined as the probability that the classifier would assign a given patient as PNI-positive – for unannotated, histologically-negative, and histologically-positive patients. For analysis of the blinded histological re-review, a two-sided Fisher's exact test was used to evaluate statistical differences in the proportion of patients in the high-scoring group *vs.* the low-scoring group with occult PNI.

Results

Perineural invasion portends poor survival

PNI has been associated with reduced survival in many, but not all, cancer types (6,19,20). To confirm the prognostic significance of PNI in our cohort, we first performed a pan-cancer multivariate Cox regression analysis of patients with and without PNI for overall survival (OS) and disease-free survival (DFS), adjusting for age, sex, cancer type, and race as covariates. Of the 1334 PNI-annotated patients examined, 1320 had accessible data on OS and 1174 had accessible data on DFS (Table 1; Fig. 1). As expected, PNI-positive patients exhibited lower OS and DFS (Supplementary Fig. 1). To account for the possibility that PNI may have cancer-type-dependent prognostic ramifications, we additionally performed univariate and multivariate survival analyses of individual cancer types and of pooled neoplasms within an anatomical system (Supplementary Table 1, Supplementary Fig. 2), though many of these analyses were markedly less powered than that performed across the pan-cancer cohort. Nonetheless, these subset analyses revealed associations (log-rank $P < 0.05$) between PNI-positivity and poorer OS in bladder, head & neck, and genitourinary cancers, as well as PNI-positivity and poorer DFS in bladder, esophageal, head & neck, stomach, gastrointestinal, and genitourinary cancers (Supplementary Table 1), even after adjustments for co-variates.

Characterization of gene expression programs for perineural invasion

To elucidate the fundamental genetic underpinnings of PNI across numerous cancer types, we performed differential gene expression analyses on RNA-seq data comparing 845 PNI-positive to 489 PNI-negative patients ($P_{adj} < 0.001$) (Fig. 2A). We next performed gene set enrichment analyses (GSEA) of significant features to extract recurring biological pathways, yielding significant overlap with signatures for synapses, glioblastoma subtypes, cell projection organization, and transcription factor activity (e.g., *SOX6*) (Fig. 2B), as well as others such as cell motility and matrisomal interactions that are consistent with known contributors to invasion (21). In addition to GSEA, we examined individual genes within our differential expression results to facilitate biological interpretation. Interestingly, we identified pathways associated with epithelial-mesenchymal transition (EMT) (e.g., *VIM*, *MET*, *MEST*), heat shock molecular chaperones (e.g., *HSPD1*, *DNAJB11*, *HSPA8*), and neurogenesis/axon guidance (e.g., *GDNF*, *NEUROG3*, *ROBO3*, *SEMA4B*). However, many of these PNI-associated genes are not exclusive to one cell type, and it is unknown how their expression is distributed among malignant cells and accompanying cells in the microenvironment. Indeed, these findings could derive in part from greater nerve density – and thus Schwann cell density – in PNI-positive tumors. Furthermore, it is unclear whether these associated programs are specific to PNI or also present in LVI, an orthogonal route of tumor spread.

PNI can occur without vascular or lymphatic invasion, and is thought to represent the initial steps for one distinct route of metastasis (6). To identify genes that differentiate PNI from LVI, we performed a differential expression analysis between $PNI^{+}LVI^{-}$ and $LVI^{+}PNI^{-}$ patients, unveiling 228 and 1755 genes uniquely associated with PNI and LVI, respectively ($P_{adj} < 0.005$) (Fig. 2C). $PNI^{+}LVI^{-}$ tumors were enriched in neural-like signatures such as synaptic signaling, neuron differentiation, neurogenesis, post-synaptic response, as well as cytoskeletal organization, while $LVI^{+}PNI^{-}$ tumors were enriched in gene signatures for secretion, response to oxygen, interleukin production, and ion transport (Fig. 2D), reflecting the existence of biological programs unique to each of PNI and LVI.

We additionally identified differentially expressed genes for select individual cancer types ($n > 10$ for each group) in our cohort (Supplementary Fig. 3), given that PNI may be driven by or result in distinct pathobiological mechanisms at different neoplastic sites. Notably, *AVEN*, a negative regulator of cell death, is the most upregulated gene ($P_{adj} = 1.73 \times 10^{-6}$) in PNI-positive prostate cancer and is consistent with prior studies that have similarly reported enrichment of anti-apoptotic genes in co-cultures of mouse dorsal root ganglia and human cancer cells (8,9). In head & neck cancer, we note overlaps with gene ontology terms for apoptosis-related (GO: apoptotic process, $P_{adj} = 2.86 \times 10^{-8}$) and metabolic processes (GO: macromolecule catabolic process, proteolysis, organophosphate biosynthetic process; $P_{adj} < 10^{-7}$ for all), suggesting there may be alterations to survival and metabolic capabilities associated with PNI. To further power our differential expression results for cancers with smaller sample sizes ($n < 10$ in either the PNI-positive or PNI-negative category), we also repeated these analyses for separately pooled gastrointestinal (GI) cancers and genitourinary (GU) cancers (Supplementary Fig. 3).

Neoplastic programs may drive perineural invasion in concert with other cell compartments

Many differentially-expressed genes and gene programs from our PNI-positive *vs.* PNI-negative analysis are neural-like, but it is unclear which cell types express them. To address the range of possibilities, we first examined these genes in the context of atlases that contain cell-type-specific information. As such, we evaluated the proportion of genes that were specific to various cell types captured by single-cell RNA-seq studies of head & neck squamous cell carcinoma and colorectal cancer (Fig. 2E) (15,16), two representative neoplasms that contain moderate to high levels of PNI within the broader otolaryngologic and gastrointestinal systems, respectively. In head & neck cancer, the proportions of differentially-expressed genes that were assigned to a specific cell type were as follows: 16.2% malignant cells, 6.3% myocytes, 4.8% endothelial cells, 4.2% fibroblasts, 4.0% dendritic cells, 3.0% macrophages, 2.8% B cells, 2.5% T cells, and 1.3% mast cells (Fig. 2E). However, the majority of differentially-expressed genes were either expressed in more than one cell type (30.7%) or were not detected in the single-cell RNA-seq data / could not be mapped to a cell type (24.2%) (Fig. 2E). Comparable proportions were observed in colorectal cancer (Fig. 2E). Importantly, more genes mapped to malignant cells than to any other cell type, suggesting a central role for neoplastic-intrinsic transcriptional programs in facilitating PNI. However, our data also reflect that an ensemble of cells in the microenvironment may participate in this invasive process.

Although mapping genes in our PNI signature to prior atlases may suggest plausible cell types of origin, it still does not fully elucidate whether the observed cell-type-specific expression patterns actively contribute to PNI, non-functionally demarcate cell populations in PNI-positive tumors, or exist as mere artifacts of the analysis. Indeed, it is possible they encapsulate contributions from alternative processes that are associated with but nonetheless distinct from PNI, such as heightened intra-tumoral nerve/glial-cell density and/or neurogenic programs in reactive stroma. To address this outstanding question, we cross-referenced our PNI signature against previously described expression profiles from dorsal root ganglia/cancer cell co-cultures and protein-level validation studies of perineural malignant cells (8,9,22). Interestingly, this revealed shared upregulation (e.g., pro-survival, anti-apoptotic) and downregulation (e.g. mitochondrial-ribosomal, prostaglandin-associated) of genes in related pathways. Furthermore, in concordance with a study (9) of human prostate specimens that had identified upregulation of NF-κB in perineural malignant cells, we also found upregulation of *NFKB1* (P_{adj} = 0.003), suggesting that many components of our signature describe *de facto* PNI biology.

Machine learning of gene expression data enables classification of histological outcomes

We next sought to determine whether differentially-expressed genes from the pan-cancer PNI-positive *vs.* PNI-negative analysis could predict histological outcomes. To this end, we conducted a grid search of eight machine learning classification models and found that the random forest (RF) approach was the best performing (Fig. 3A). Among the features weighted most strongly by the model (≥90th percentile), we noted associations with the NF-κB pathway (e.g., *APOBEC3B*, *SRGN*, *PDLIM2*, *NOS2*) that are congruent with observations from prior *in vitro* studies (9). In addition, the model considered genes involved in axon guidance (e.g., *SEMA4B*) and p75 neurotrophin receptor-mediated signaling (e.g., *FAM13A*) to be of high importance (≥90th percentile), which may reflect neuro-epithelial crosstalk that facilitates PNI or enhanced nerve density in PNI-positive tumors.

Our PNI-RF model exhibited an area under the curve (AUC) of 0.80 (Fig. 3B). To confirm that the model derived from a randomly generated 80% training/20% validation data split was representative of overall RF performance on the dataset, we performed 10 additional random splits that yielded a mean AUC of 0.78 ± 0.02. Recall exceeded precision by a notable margin (91% *vs.* 72%) when using a PNI cutoff score (0.48) that optimized for F_1 (0.80), revealing that the classifier identified ‘true positive’ cases with high accuracy but also called a substantial number of ‘false positives.’ While this result may reflect limitations of our classifier, we posit that the prevalence of ‘false positives’ may actually represent the presence of histologically occult cases. This possibility is consistent with previous findings that PNI is often overlooked during H&E examination (11,12). We similarly generated and evaluated system-specific classification models (Supplementary Fig. 4).

Gene expression classifiers improve occult histopathological detection

We then examined whether our PNI-RF classifier could enhance histopathological detection of occult cases, given that low detection sensitivity remains a unique challenge for PNI due to suboptimal identification with routine histological staining and a lack of standardized guidelines for synoptic reporting in many cancer types. We designed a blinded review of

tissue sections from patients initially determined to be histologically PNI-negative, split into two equal cohorts with high (>0.6) or low (<0.4) classifier scores. Based on a recall of 91% and precision of 72%, the occult PNI rate in our cohort was estimated to be 15%. In a prior study examining pancreatic cancer sections for PNI, 19 of 59 specimens were initially annotated as PNI^+ based on H&E (23). The number of PNI^+ cases increased to 33 on H&E review and further increased to 49 after S100 nerve-specific staining. Taking S100 staining as the histological "gold standard," there were 30 occult PNI^+ cases after initial H&E examination. H&E review correctly overturned 14 of these 30 occult cases (change from PNI-negative to PNI-positive), which corresponds to a true positive rate of 0.47. Assuming all our high classifier score specimens harbor occult PNI, we estimated that H&E review will detect these cases at a rate of 0.47. The low classifier score cohort should have a low probability of occult PNI so we conservatively estimated a detection rate of 0.1 on H&E review. Taking the difference between these rates in the high and low classifier score groups yielded an estimated effect size of 0.37. For a two-sided Fisher's exact test with a power of 90% and $\alpha = 0.05$, the total sample size requirement is approximately 66 specimens (33 high PNI score, 33 low PNI score). In addition, since patients with prostate cancer comprised a sizeable portion (43.8%) of our PNI-positive cohort, we specifically evaluated whether our pan-cancer classifier provides value for non-prostatic malignancies. We therefore selected 66 patients with GI cancers (colorectal, esophageal, pancreatic, gastric) who were initially annotated as PNI-negative; of these, patients were evenly split by whether our PNI classifier predicted them to be PNI-positive (mean score = 0.69, standard deviation = 0.046, range = 0.61–0.77) or -negative (mean score = 0.30, standard deviation = 0.042, range = 0.21–0.36). On average, there were three slides available for review per patient in both the high- and low-scoring cohorts. A board-certified GI pathologist (S.S.) blinded to the PNI classifier scores then reviewed the slides. The pathologist overturned significantly more high-scoring cases (15 of 33, 45.4%) than low-scoring ones (6 of 33, 18.1%) (Fisher's exact test, $P = 0.0332$) (Fig. 3C), suggesting that our gene expression classifier facilitates occult histopathological detection and thereby addresses some of the challenges of H&E evaluation. From the pathology review, we were able to identify various patterns of occult PNI including total encirclement in which the malignant cells wrap themselves around the entire circumference of the nerve sheath, partial encirclement (>33% of the circumference of the nerve sheath), and intraneural ductal formation wherein the malignant cells form aberrant duct-like structures within one of the three layers of the nerve sheath (Fig. 3D). Overall, however, we did not observe any marked associations between the pattern of invasion and PNI score, suggesting that our classifier likely distinguishes between presence and absence of PNI, irrespective of how it manifests histologically.

Unannotated patients may represent a mixture of true negatives and occult cases

Despite the poor clinical outcomes associated with PNI, there is a lack of standardized guidelines for reporting in many cancer types, leading to underreporting in certain contexts. For example, there are a considerable number of patients without PNI annotations in neoplasms such as non-small cell lung cancer even though the prevalence of neural invasion has been reported to be over 25% (24). Alternatively, we posit that PNI was likely not observed if a pathologist did not provide a specific annotation. To assess the hypothesis that unannotated patients represent a combination of occult and true negative cases, we compared

PNI classifier scores across unannotated, histologically-negative, and histologically-positive patients. Indeed, we found significant differences among each of the three groups: the histologically-positive cohort had the highest mean PNI score, followed by the unannotated and then the histologically-negative cohorts (one-way ANOVA, $P < 0.01$; Tukey's HSD, $P < 0.01$) (Fig. 3E). We anticipate that our pan-cancer PNI classifier may help nominate additional cancer types that have a sufficient burden of occult PNI to warrant routine annotation.

Discussion

Although nerve- and tumor-specific stains can be used to enhance histological identification of PNI, antibody-based strategies are limited by sensitivity, specificity, and scalability. Conversely, as full transcriptome technologies near widespread clinical adoption within the next decade (13,25,26), it will be feasible to obtain detailed molecular portraits of tumors and perform focused *in silico* analyses (e.g., prediction of PNI) based on subsets of relevant genes therein. Indeed, in this study, we provide a proof-of-principle demonstration that RNA-seq gene expression data can serve as a valuable adjunct to histopathology. Importantly, we underscore that our gene expression classifier should not be used as a substitute for histopathological inspection, but as a potential way to nominate previously-annotated, PNI-negative cases that could be re-reviewed for occult PNI. One strength of our classifier in this regard is that it is optimized for high sensitivity and is thereby well-suited for flagging false-negative cases. However, as with other machine learning methods that maximize model performance, the random forest approach used in this study runs the risk of overfitting the data. It may therefore be useful for subsequent investigations to first evaluate the PNI-RF model against another large, independent multi-cancer dataset with both PNI annotations and mRNA information prior to adoption as a histopathological assistive tool. In light of tissue sampling bias, future work with this model would additionally benefit from exploring the optimal number of re-reviewed sections needed for occult detection, which was not feasible in our current study given the limited number of tissue slide images per patient. Moreover, further investigation aimed at quantifying the extent and morphological subtypes of PNI may prove to be more clinically impactful than a binary separation of tumors into PNI-positive *vs.* -negative categories.

Beyond clinicopathological applications, the PNI classifier provides biological insights into this specific invasive phenotype as we were able to both capture genes previously reported for PNI in individual cancers (e.g., *CXCL16, NFKB1*) (9,10) and nominate new candidate genes for functional investigation. Interestingly, genes differentially expressed in PNI-positive tumors overlapped a gene set for the mesenchymal subtype of glioblastoma, which is associated with the EMT phenotype and the tumor necrosis factor (TNF) pathway (27). Concordantly, our PNI classifier included a number of EMT-related (e.g., *VIM*, *MET*, *MEST*) and TNF-induced genes (e.g., *TNFAIP8L2*, *TNFAIP3*), which could be functionally explored in future studies. Prior to doing so, it will be crucial to validate that a tissue of interest expresses a given signature gene, given that the signatures were derived from a pan-cancer cohort. Furthermore, it will be necessary to identify the cell type of origin for each gene, as PNI is known to involve crosstalk among numerous cell types (3,7) and primary analyses were performed on bulk RNA-seq data. We addressed this challenge by

demonstrating the utility of mapping salient signature genes onto single-cell RNA-seq data, which revealed that more genes mapped to malignant cells than to any other cell type. This finding indicated that cancer cell-intrinsic transcriptional programs likely play a critical role in promoting PNI in collaboration with other cell types in the microenvironment.

It is worth noting some of the underlying limitations of the study datasets and discussing future directions that can help overcome them. First, in light of the predicted abundance of occult PNI cases (estimated at 15%), one may deduce that the PNI-negative group is fundamentally impure and would therefore weaken the survival, gene signature, and classifier analyses. Next, transcriptomic profiling by RNA-seq or other bulk-level dissociative techniques neither provide information about each gene's cell-of-origin nor preserve the spatial context of nerves and Schwann cells relative to malignant and other stromal cells. As a result, it remains necessary to validate that each component of the signature identified is central to PNI and not a byproduct of an associative artifact of PNI-positive tumors (e.g., reactive stroma, alternative mechanisms along the neuro-epithelial interaction spectrum). Further insights into the underlying molecular mechanisms driving PNI should therefore be explored using either spatial transcriptomics or higher-plex protein-based assays, particularly because foci of PNI often do not occupy the majority of tumor volume but associate with poor clinical outcomes nonetheless. These approaches could also be beneficial for elucidating whether distinct mechanisms underlie the various patterns of PNI. Finally, functional validation of a subset of PNI classifier genes would provide a novel opportunity to therapeutically intervene upon this pervasive avenue of tumor invasion and dissemination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Patrick Yu, Karthik Jagadeesh, and Zackery Ely for helpful discussions and advice regarding computational resources. We thank the UCSF Dean's Yearlong Fellowship (J.A.G.), Chinese American Medical Society Summer Research Fellowship (J.A.G.), Lustgarten Foundation (T.J.), American Society for Clinical Oncology/Conquer Cancer Foundation Young Investigator Award (W.L.H), Hopper-Belmont Foundation Inspiration Award (W.L.H.) and American Cancer Society/Massachusetts General Hospital Institutional Research Grant (W.L.H.). T.J. is an investigator of the Howard Hughes Medical Institute. W.L.H. is an Andrew L. Warshaw, M.D. Institute for Pancreatic Cancer Research Fellow.

Conflicts of Interest

M.M.K. has served as a compensated consultant for H3 Biomedicine and AstraZeneca and received a research grant (to institution) from Novartis that is not related to this work. P.C. is an employee of Raytheon Technologies. No Raytheon intellectual property, equipment, or funding was used in this research. D.Y.K. is a paid consultant at Verve Therapeutics, unrelated to this research. D.Y.K.'s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. T.J. is a member of the Board of Directors of Amgen and Thermo Fisher Scientific. He is also a co-Founder of Dragonfly Therapeutics and T2 Biosystems. T.J. serves on the Scientific Advisory Board of Dragonfly Therapeutics, SQZ Biotech, and Skyhawk Therapeutics. T.J. laboratory currently also receives funding from the Johnson & Johnson Lung Cancer Initiative, but this funding did not support the research described in this manuscript. None of these affiliations represent a conflict of interest with respect to the design or execution of this study or interpretation of data presented in this manuscript. All other authors declare no competing interests.

References

1. Zeng Q, Michael IP, Zhang P, Saghafinia S, Knott G, Jiao W, et al. Synaptic proximity enables NMDAR signalling to promote brain metastasis. Nature 2019; 573:526–531. [PubMed: 31534217]
2. Venkataramani V, Tanev DI, Strahle C, Studier-Fischer A, Fankhauser L, Kessler T, et al. Glutamatergic synaptic input to glioma cells drives brain tumour progression. Nature 2019; 573:532–538. [PubMed: 31534219]
3. Zahalka AH, Frenette PS. Nerves in cancer. Nat Rev Cancer 2020; 20:143–157. [PubMed: 31974491]
4. Renz BW, Tanaka T, Sunagawa M, Takahashi R, Jiang Z, Macchini M, et al. Cholinergic signaling via muscarinic receptors directly and indirectly suppresses pancreatic tumorigenesis and cancer stemness. Cancer Discov 2018; 11:1458–1473.
5. Renz BW, Takahashi R, Tanaka T, Macchini M, Hayakawa Y, Dantes Z, et al. β2 Adrenergic-Neurotrophin Feedforward Loop Promotes Pancreatic Cancer. Cancer Cell 2018; 34:863–867. [PubMed: 30423300]
6. Liebig C, Ayala G, Wilks JA, Berger DH, Albo D. Perineural invasion in cancer: A review of the literature. Cancer 2009; 115:3379–91. [PubMed: 19484787]
7. Bapat AA, Hostetter G, Von Hoff DD, Han H. Perineural invasion and associated pain in pancreatic cancer. Nat Rev Cancer 2011; 11:695–707. [PubMed: 21941281]
8. Dai H, Li R, Wheeler T, Ozen M, Ittmann M, Anderson M, et al. Enhanced survival in perineural invasion of pancreatic cancer: an in vitro approach. Hum Pathol 2007; 38:299–307. [PubMed: 17097719]
9. Ayala GE, Dai H, Ittmann M, Li R, Powell M, Frolov A, et al. Growth and survival mechanism associated with perineural invasion in prostate cancer. Cancer Res 2004; 64:6082–90. [PubMed: 15342391]
10. Ha HK, Lee W, Park HJ, Lee SD, Lee JZ, Chung MK. Clinical significance of CXCL16/CXCR6 expression in patients with prostate cancer. Mol Med Rep 2011; 4:419–24. [PubMed: 21468586]
11. Conte GA, Qari O, Fasano GA, Guinto RK, Palo L, Parker GS, et al. S100 Staining Adds to the Prognostic Significance of the Combination of Perineural Invasion and Lymphovascular Invasion in Colorectal Cancer. Appl Immunohistochem Mol Morphol 2018; 28:354–359.
12. Kurtz KA, Hoffman HT, Zimmerman MB, Robinson RA. Perineural and vascular invasion in oral cavity squamous carcinoma: Increased incidence on re-review of slides and by using immunohistochemical enhancement. Arch Pathol Lab Med 2005;129:354–9. [PubMed: 15737030]
13. Shalek AK, Benson M. Single-cell analyses to tailor treatments. Sci Transl Med 2017; 9.
14. Roychowdhury S, Chinnaiyan AM. Translating cancer genomes and transcriptomes for precision oncology. CA Cancer J Clin 2016; 66:75–88. [PubMed: 26528881]
15. Puram S V, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer Cell 2017; 171:1611–1624.
16. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat Genet 2017;49:708–718. [PubMed: 28319088]
17. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401–4. [PubMed: 22588877]
18. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013; 6:pl1.
19. Jardim JF, Francisco ALN, Gondak R, Damascena A, Kowalski LP. Prognostic impact of perineural invasion and lymphovascular invasion in advanced stage oral squamous cell carcinoma. Int J Oral Maxillofac Surg 2015; 44:23–8. [PubMed: 25457832]
20. Maru N, Ohori M, Kattan MW, Scardino PT, Wheeler TM. Prognostic significance of the diameter of perineural invasion in radical prostatectomy specimens. Hum Pathol 2001; 32:828–33. [PubMed: 11521227]

21. Woodhouse EC, Chuaqui RF, Liotta LA. General mechanisms of metastasis. Cancer 1997;80:1529–37. [PubMed: 9362419]

22. Prueitt RL, Yi M, Hudson RS, Wallace TA, Howe TM, Yfantis HG, et al. Expression of microRNAs and protein-coding genes associated with perineural invasion in prostate cancer. Prostate 2008; 68:1152–64. [PubMed: 18459106]

23. Zeng L, Guo Y, Liang J, Chen S, Peng P, Zhang Q, et al. Perineural invasion and TAMs in pancreatic ductal adenocarcinomas: Review of the original pathology reports using immunohistochemical enhancement and relationships with clinicopathological features. J Cancer 2014; 5:754–60. [PubMed: 25368675]

24. Yilmaz A, Duyar SS, Cakir E, Aydin E, Demirag F, Karakaya J, et al. Clinical impact of visceral pleural, lymphovascular and perineural invasion in completely resected non-small cell lung cancer. Eur J Cardio-thoracic Surg 2011;40:664–70.

25. Cieślik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet 2018;19:93–109. [PubMed: 29279605]

26. McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. N Engl J Med 2011;364:340–50. [PubMed: 21268726]

27. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010;17:98–110. [PubMed: 20129251]

Statement of Translational Relevance

Perineural invasion (PNI) is a pervasive mechanism of cancer invasion, recurrence, and metastasis but the underlying mechanisms are poorly understood, and standard histopathological examination has limited detection sensitivity. Using a custom cohort of 2029 patients across 12 cancer types with bulk transcriptomic data and surgical pathology annotation, we identified the fundamental gene expression programs associated with PNI. Single-cell RNA-seq atlases were leveraged to infer the cell type sources of these differentially-expressed genes, which provides a timely blueprint for functional investigations aimed at elucidating and therapeutically disrupting aberrant tumor-nerve crosstalk. We then harnessed machine learning approaches to develop a novel gene expression classifier that predicts the presence of PNI. A blinded pathology review of tissue sections demonstrated that the PNI classifier facilitates the detection of occult cases.

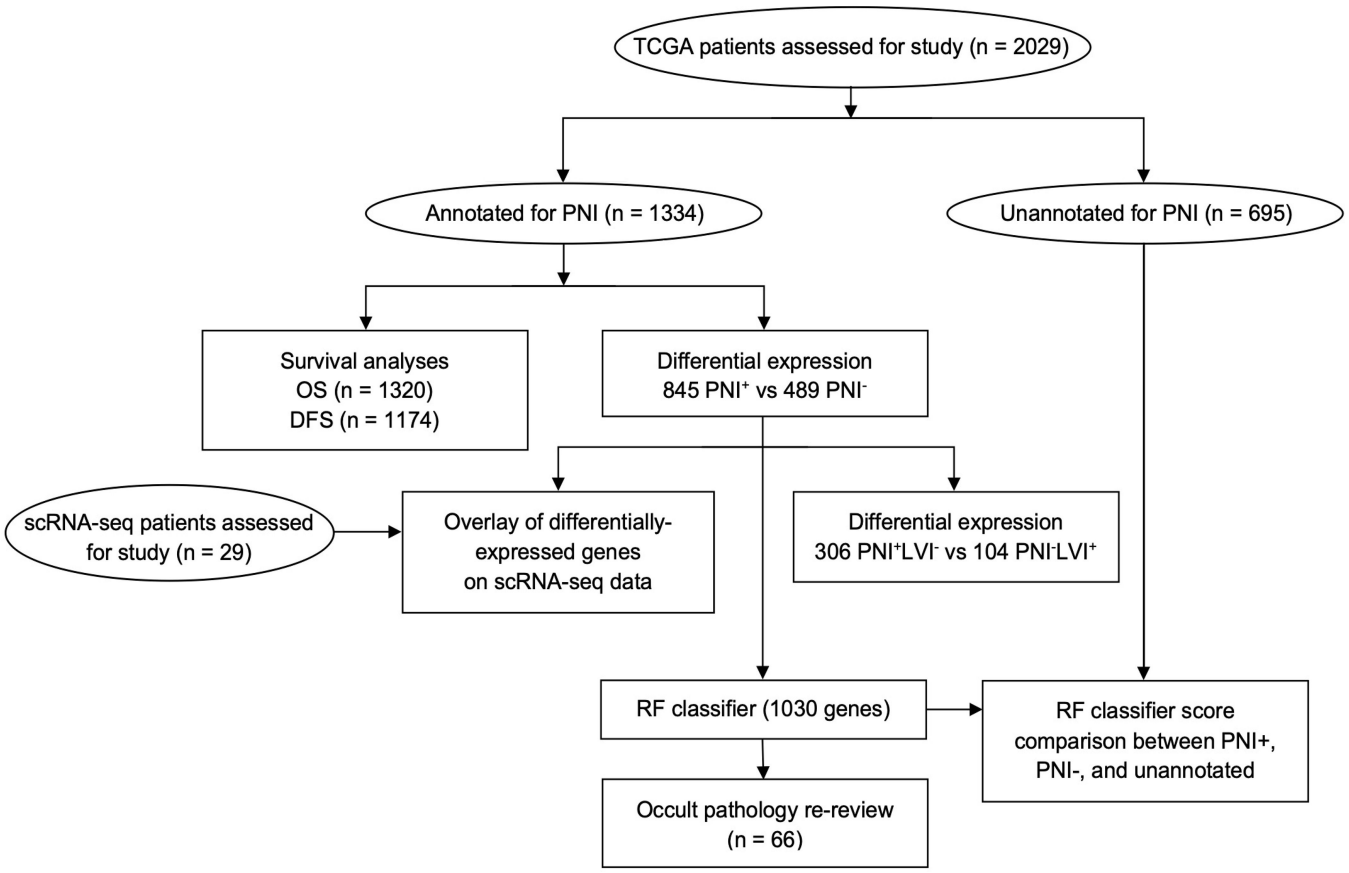

Figure 1. Study design flow diagram.

Ovals denote data inputs for analyses, while rectangles denote analyses conducted.

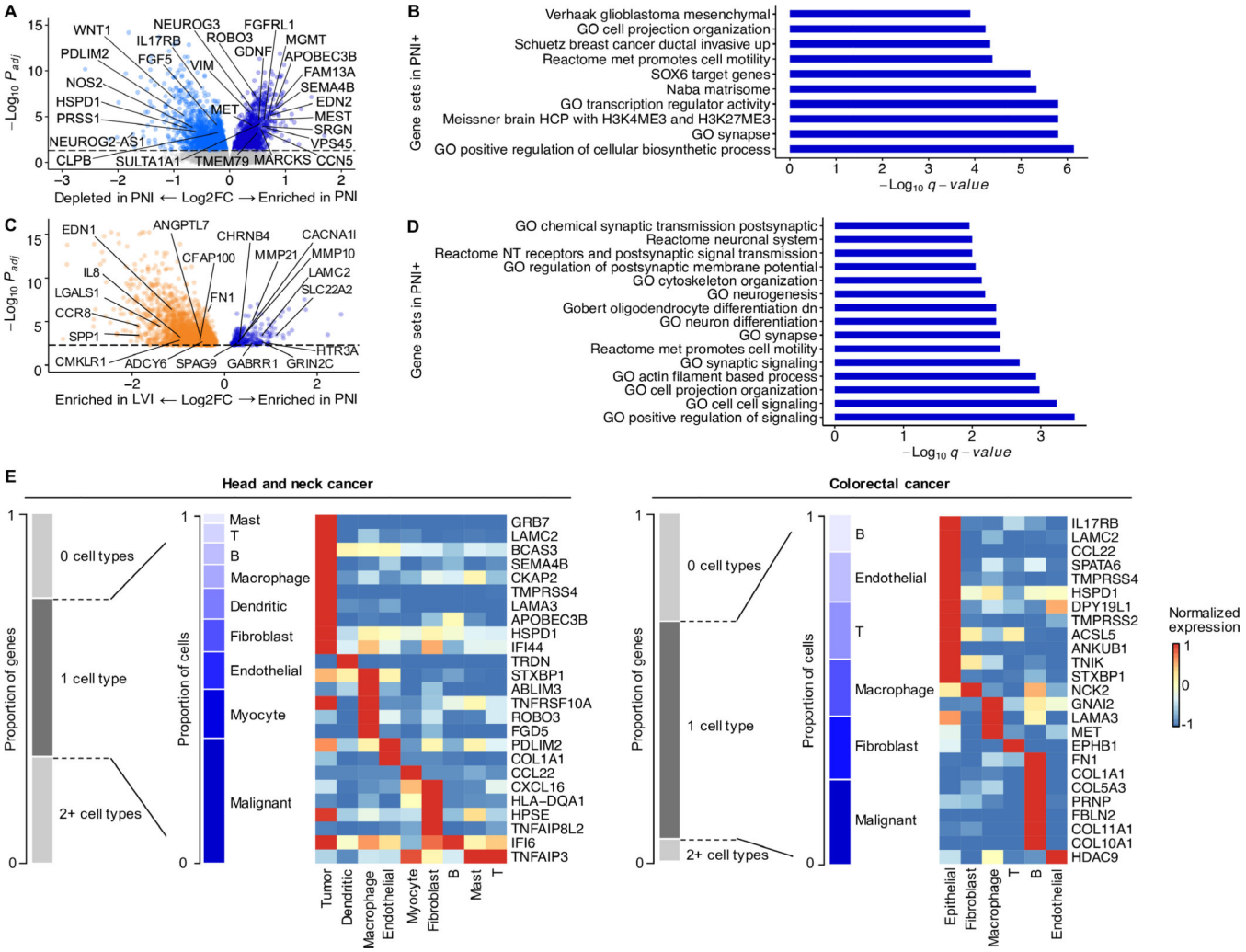

Figure 2. Distinct genes and gene programs associate with PNI, which receive contributions from multiple cell compartments.

A, Volcano plot of genes enriched and depleted in PNI-positive patients with select genes highlighted from the PNI classifier. **B,** Gene set enrichment analysis of features in PNI-positive patients. **C,** Volcano plot of genes enriched in $PNI^{+}LVI^{-}$ relative to $PNI^{-}LVI^{+}$ patients, or vice versa, with select genes labeled. **D,** Gene set enrichment analysis of features in $PNI^{+}LVI^{-}$ patients. **E,** Left-sided stacked barplots show proportions of differentially-expressed genes in PNI that are specific to zero, one, or more than one cell type identified in the single-cell atlases of head and neck cancer and colorectal cancer. Right-sided stacked barplots show the proportion of signature genes that mapped exclusively to one cell type. Heat maps depict select genes that are specific to different cell types identified from single-cell RNA-seq studies of head and neck cancer and colorectal cancer.

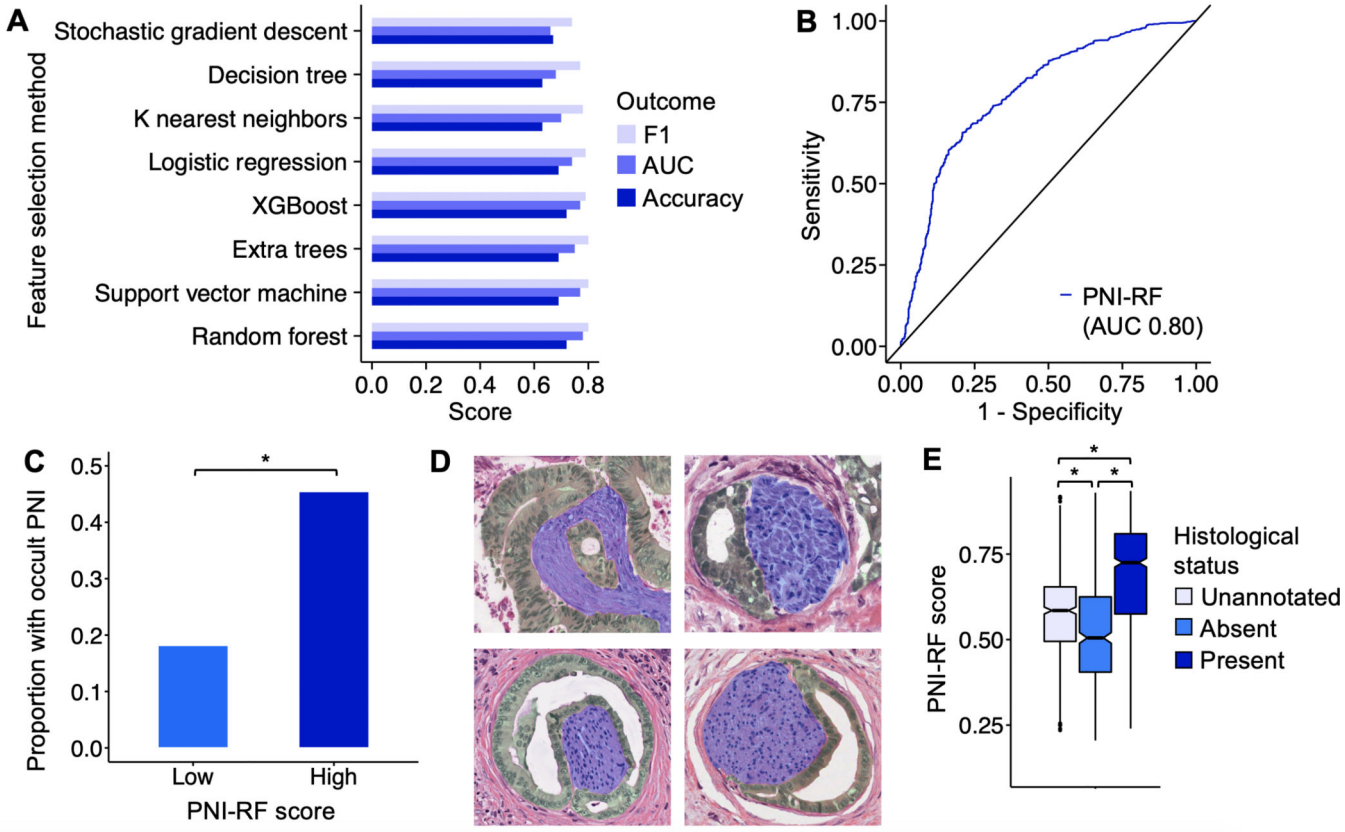

Figure 3. Gene expression classifiers derived using machine learning algorithms predict PNI and improve occult histopathological detection.

A, Bar plot of performance (F1, accuracy, AUC) of various machine learning models using a grid search workflow. **B,** ROC curve of optimized RF classifier for PNI, with an AUC of 0.80. **C,** Bar plot representing the proportion of GI cancer patients with occult PNI in the high-scoring group *vs.* the low-scoring group, as determined by blinded histological re-review of H&E slides. * $P = 0.0332$. **D,** Four representative H&E examples of various patterns of occult PNI for GI cancers that were missed on initial review, with green masks denoting the malignant glands or ducts and purple masks denoting nerves. Top left shows complete encirclement as well as intraneural glandular formation; bottom left shows complete encirclement by malignant ducts; top and bottom right show malignant ducts in the perineural space. **E,** Comparison of classifier scores among patients who were unannotated, histologically-negative, or histologically-positive for PNI. Error bars represent range of scores. * $P < 0.01$.

Table 1.

Demographics and clinicopathological characteristics of TCGA patient cohort used in study.

Sex, age, race, PNI status, LVI status, and survival information for study cohort across 12 cancer types.

	BLCA (n = 103)	BRCA (n = 290)	CHOL (n = 27)	COAD (n = 177)	ESCA (n = 75)	HNSC (n = 366)	LIHC (n = 100)	PAAD (n = 163)	PRAD (n = 461)	STAD (n = 144)	THCA (n = 70)	UCS (n = 53)
Sex												
Men	78	3	11	84	66	271	65	90	462	107	23	NA
Women	25	287	16	91	9	95	35	73	NA	37	47	53
Age, y	68 (10)	57 (12)	63 (14)	64 (14)	59 (11)	61 (2)	58 (15)	65 (11)	61 (7)	67 (10)	50 (15)	70 (10)
Race												
White	85	181	22	123	38	309	53	144	387	81	52	41
B/AA	7	96	2	52	2	35	7	5	45	9	7	9
Asian	8	8	3	0	33	8	36	11	11	41	9	3
AI/AN	0	1	0	0	0	1	0	0	1	0	1	0
NH/PI	0	0	0	0	0	0	0	0	0	1	0	0
PNI												
Yes	36	6	7	39	12	175	1	134	370	62	3	0
No	33	11	7	107	15	160	20	25	46	65	0	0
LVI												
Yes	54	88	7	60	23	129	32	99	102	101	23	34
No	49	126	18	102	50	195	67	53	321	42	46	19
Survival data												
OS	103	215	27	175	75	364	99	154	461	144	70	53
DFS	84	207	24	157	65	284	88	128	460	127	70	26

Abbreviations: BLCA = bladder cancer, BRCA = breast cancer, CHOL = bile duct cancer, COAD = colon cancer, ESCA = esophageal cancer, HNSC = head and neck cancer, LIHC = liver cancer, PAAD = pancreas cancer, PRAD = prostate cancer, STAD = stomach cancer, THCA = thyroid cancer, UCS = uterus cancer, B/AA = Black or African American, AI/AN = American Indian or Alaska native, NH/PI = Native Hawaiian or other Pacific Islander, PNI = perineural invasion, LVI = lymphovascular invasion, OS = overall survival, DFS = disease free survival, NA = not applicable
Categorical data presented as n, continuous data as mean (SD).