Integrated transcriptomic and epigenetic profiling reveals a conserved differentiation state in neuroendocrine neoplasms

Timothy Khumpan [1,2] *, Namrata Vijayvergia [3] * †, Kasonde Chewe[10] *, Eric Eleam[1], Juan Nicolas Quiñones-Romero[1], Suraj Peri [3,4], Gong Yulan [5], Paul F. Engstrom [3], Cai Qi [5], Jiangming Pei [6], Gulnaz Alekbaeva[1], Kerry Campbell [7], Johnathan Whetstine [1], Igor Astsaturov [3], Hayan Lee [1,8,9†]

[1] Cancer Epigenetics Institute, Nuclear Dynamics and Cancer Program, Fox Chase Cancer Center

[2] Department of Computer and Information Sciences, College of Science and Technology, Temple University

[3] Department of Hematology/Oncology, Fox Chase Cancer Center

[4] Bioinformatics Oncology Target Discovery, Merck

[5] Histopathology Facility, Fox Chase Cancer Center

[6] Clinical Genomics, Fox Chase Cancer Center

[7] Cancer Signaling and Microenvironment Program, Fox Chase Cancer Center

[8] Department of Cancer and Cellular Biology, Lewis Katz School of Medicine, Temple University

[9] Department of Medical Genetics and Molecular Biochemistry, Lewis Katz School of Medicine, Temple University

[10] School of Biomedical Engineering, Science, Health Systems, Drexel University

timothy.khumpan@temple.edu

namrata.vijayvergia@fccc.edu

kasonde.chewe@drexel.edu

eric.eleam@temple.edu

suraj.peri@fccc.edu

yulan.gong@fccc.edu

pfengstrom@gmail.com

qi.cai@fccc.edu

jianming.pei@fccc.edu

gulnaz.alekbaeva@fccc.edu

kerry.campbell@fccc.edu

johnathan.whetstine@fccc.edu

igor.astsaturov@fccc.edu
hayan.lee@fccc.edu
* Contributed equally to this work
† Corresponding authors

Abstract

Background

Neuroendocrine neoplasms (NENs) comprise biologically heterogeneous tumors whose classification into well-differentiated neuroendocrine tumors (WD-NETs) and poorly differentiated neuroendocrine carcinomas (PD-NECs) remains clinically challenging. Morphology and proliferation indices alone often fail to resolve high-grade cases, motivating molecular frameworks that capture differentiation state across tissues.

Results

We profiled 36 FFPE NENs from 11 anatomical sites using a targeted 784-gene transcriptomic panel and developed interpretable machine-learning classifiers to model histologic differentiation. Across cross-validation, models achieved stable discrimination between WD-NETs and PD-NECs (87–93% accuracy). Feature attribution identified a conserved molecular axis separating lineage-preserved WD-NETs from proliferative, replication stress–enriched PD-NECs. WD-NETs retained neuroendocrine and neuronal signaling programs, whereas PD-NECs exhibited activation of cell-cycle, DNA damage response, and chromatin-regulatory pathways.

Orthogonal validation confirmed these findings at multiple levels: subtype-specific protein expression by immunohistochemistry (e.g., EZH2, PAK3) and independent DNA methylation profiling from an external lung NEN cohort demonstrated concordant stratification and increased epigenetic stemness in PD-NECs. Together, transcriptomic and epigenetic analyses reveal a conserved differentiation state that transcends anatomical origin.

Conclusions

Integrated multi-omic profiling identifies a tissue-agnostic differentiation axis that robustly stratifies neuroendocrine neoplasms and provides biologically interpretable biomarkers for resolving histologic ambiguity. These findings establish differentiation state as a unifying molecular framework for NEN classification and suggest a foundation for future diagnostic and therapeutic development.

Keywords

Neuroendocrine neoplasm, neuroendocrine carcinoma (NEC), neuroendocrine tumor (NET), transcriptomics, machine learning, biomarker, epigenetics, DNA methylation, cell cycle, mitotic age

Background

Neuroendocrine neoplasms (NENs) arise from specialized epithelial cells that secrete amine or peptide hormones in response to neural and endocrine stimuli. These cells are most abundant in the gastrointestinal (GI) tract, pancreas, and lungs but occur across many organ systems [1, 2, 3, 4, 5, 6]. Although NENs remain less common than many solid tumors, incidence has risen substantially. In the United States, rates increased from 1.09 per 100,000 in 1973 to 6.98 per 100,000 in 2012 (6.4-fold) [7], with parallel rises in the United Kingdom and Taiwan [8].

Accurate histological classification is essential for prognosis and therapy yet remains challenging because key features overlap across subtypes [9]. Differentiation status, assessed by morphology and proliferation markers, is the strongest prognostic determinant [10], but tumors with similar histology can follow divergent clinical courses. PD-NECs are highly aggressive with rapid progression and limited treatment durability [11]; pancreatic PD-NECs have median overall survival of ~11–12 months [12, 13]. In contrast, WD-NETs are typically indolent with median survival of 5 to >10 years depending on site and grade; for example, pancreatic WD-NETs show 8–12 years and small-intestinal NETs often exceed 65–75% 5-year survival [14, 8, 10].

The World Health Organization (WHO) framework [15, 16] relies on histomorphology and proliferation thresholds (mitotic count >20 per 2mm^2 and Ki-67 >20%) to classify PD-NECs as high-grade tumors [10]. While clinically essential, this scheme has limitations. WD-NETs span a broad grade spectrum (G1–G3), and high-grade WD-NETs can be morphologically indistinguishable from PD-NECs. Immunohistochemical markers such as chromogranin A and synaptophysin, and aberrant p53/RB1 patterns, are helpful but not definitive and frequently overlap in high-grade tumors [15, 17].

A recent multi-institutional study reported poor interobserver agreement in high-grade NEN classification, with experts failing to reach consensus in nearly two-thirds of cases [15]. This highlights the limitations of morphology-only approaches and motivates quantitative molecular classifiers.

Next-generation sequencing (NGS) has advanced understanding of the molecular landscape [11, 18], but molecular profiling has also revealed overlap between WD-NETs and PD-NECs. Some WD-NETs exhibit high proliferation and TP53 alterations typical of PD-NECs, limiting diagnostic specificity of these markers [15, 19]. PD-NEC drivers remain incompletely characterized due to rarity and substantial intra- and inter-tumoral heterogeneity [20, 21, 22]. NENs also have low tumor mutational burden (TMB) compared

with many solid tumors [6, 20], constraining genomics-only approaches. In a cohort of 85 metastatic/advanced NENs, only 16 (18.82%) were PD-NECs [20], limiting power to define robust subtype drivers.

Several molecular studies have sought to refine NEN classification beyond morphology. Protein expression analyses show loss of *DAXX*/*ATRX* in WD-NETs and aberrant p53/*RB1* in PD-NECs [17]. Epigenetic profiling further supports distinct DNA methylation landscapes [23], but many studies are limited to pancreatic NENs or emphasize tissue-of-origin rather than histologic differentiation. Large methylation studies can infer tissue site across organs [24, 25, 26], yet often do not resolve the PD-NEC versus WD-NET distinction. Existing approaches can be constrained by anatomical specificity, reliance on single markers, or unsupervised clustering with limited clinical interpretability and technical variability. Transcriptomic profiling offers a complementary framework by capturing active cellular programs (lineage identity, proliferation, stress responses). Whether differentiation state represents a conserved, tissue-agnostic molecular axis in NENs remains unclear. Transcriptomic signatures can also be confounded by tissue-of-origin and microenvironmental composition, underscoring the need for interpretable models and orthogonal validation.

To address these limitations, we developed an interpretable transcriptome-based machine learning framework to distinguish PD-NECs from WD-NETs across diverse anatomical sites. Using targeted transcriptomic profiling, we trained supervised classifiers (random forest and logistic regression) that explicitly model differentiation while enabling direct interrogation of feature importance. This approach identified a conserved molecular axis: activation of cell-cycle, replication stress, and DNA damage response programs in PD-NECs versus preservation of neuroendocrine lineage and neuronal signaling programs in WD-NETs. We validated these signatures using orthogonal immunohistochemical markers and independent DNA methylation datasets, demonstrating concordant stratification and increased epigenetic stemness in PD-NECs. Together, this framework provides a quantitative, biologically interpretable basis for resolving histologic ambiguity and highlights transcriptomic state as a robust indicator of differentiation-related aggressiveness.

Results

Principal component analysis reveals a conserved differentiation axis separating PD-NEC and WD-NET

We profiled 36 neuroendocrine neoplasms (NENs) spanning 11 primary anatomical sites (esophagus, lung, breast, stomach, pancreas, small bowel, cecum, colon, rectum, ileum) along with five tumors of unknown origin (Fig. 1A; Supplementary Table S1). Independent pathological review classified 21 tumors (58%) as well-differentiated neuroendocrine tumors (WD-NETs) and 15 (42%) as poorly differentiated neuroendocrine carcinomas (PD-NECs). Transcriptomic profiling was performed using the NanoString nCounter

PanCancer Immune Profiling Panel with Panel Plus customization (784 genes), capturing cell-cycle, DNA damage response, chromatin regulation, signaling, and immune-related programs.[27, 28]

To determine whether global transcriptional variation reflected differentiation state rather than tissue of origin, we performed principal component analysis (PCA) on normalized log-transformed expression data. Unsupervised PCA revealed clear separation of PD-NECs and WD-NETs along principal component 1 (Fig. 1B). Permutational multivariate analysis of variance (PERMANOVA) confirmed that histologic subtype explained a significant proportion of transcriptomic variance ($R^2 = 0.17$, $F = 6.9$, $p = 0.001$), supporting differentiation status as a dominant molecular dimension across anatomically diverse tumors.

Importantly, separation was observed despite substantial heterogeneity in primary site, suggesting the presence of a conserved, tissue-agnostic transcriptional program linked to differentiation state. Prior transcriptomic studies have identified subtype-associated gene expression differences within restricted anatomical contexts; however, whether a unified differentiation-linked program persists across tissues has remained unclear.[29, 30, 31, 22] Our findings indicate that such a conserved axis is detectable even within a heterogeneous multi-site cohort.

To define genes contributing to this separation, differential expression analysis was performed ($|\log2FC| > 1$, FDR < 0.05). PD-NECs exhibited increased expression of genes involved in replication licensing and checkpoint control, including *CDC6*, *CHEK1*, *SFN*, and the epigenetic regulator *EZH2* (Fig. 1C). These genes collectively reflect a replication-competent, checkpoint-activated state consistent with high mitotic flux and replication stress adaptation. In contrast, WD-NETs showed higher expression of genes associated with neuronal signaling and regulated secretory function, including *PAK3*, *CAMK2B*, *GRIA3*, and *RASA4*, indicating preservation of neuroendocrine lineage identity.

Pathway enrichment analysis mirrored these findings (Fig. 1D). PD-NECs were enriched for cell-cycle progression, homologous recombination, and DNA damage response pathways, whereas WD-NETs were enriched for neuronal, synaptic, and cell–cell signaling programs. Together, these results support a model in which PD-NECs and WD-NETs occupy distinct positions along a conserved differentiation axis characterized by proliferative reprogramming versus lineage preservation.

Given the modest cohort size, these findings should be interpreted as strong but preliminary signals. Nevertheless, the consistency of subtype separation across anatomically diverse tumors suggests that differentiation state represents a fundamental organizing principle of NEN biology.

Unsupervised clustering reveals molecular substructure and an intermediate differentiation state

To determine whether transcriptomic variation resolved tumors into molecular subgroups beyond the binary WD-NET/PD-NEC classification, we performed unsupervised hierarchical clustering of normalized expression profiles. Clustering identified three reproducible groups (Fig. 2A), suggesting structured molecular organization rather than a simple two-class division.

Group 1 (n = 20) was composed predominantly of WD-NETs and was characterized by elevated expression of genes associated with neuroendocrine lineage preservation and calcium-dependent signaling, including *CACNB2*, *ZBTB16*, *GRIA3*, and *WNT4* (Supplementary Fig. 1a–h). These features reinforce a transcriptional program consistent with regulated secretory identity and neuronal differentiation.

Group 2 (n = 12) was enriched for PD-NECs and exhibited increased expression of genes involved in cell-cycle regulation, chromatin remodeling, and oncogenic signaling, including *CDKN2A*, *EZH2*, *MYB*, *IL8*, *CTNNB1*, and *NOTCH1*. The coordinated upregulation of cell-cycle and epigenetic regulators suggests a proliferative and transcriptionally reprogrammed state aligned with poorly differentiated carcinoma biology.

Group 3 (n = 4) comprised tumors of mixed histology (2 WD-NETs, 2 PD-NECs) and displayed attenuated expression of both Group 1 and Group 2 marker sets, along with increased *GATA3*. Two cases harbored *TP53* mutations without concurrent *RB1* alteration. These tumors were more dispersed in PCA space (Fig. 2B), indicating greater transcriptional heterogeneity. Although limited in size, this subgroup may represent an intermediate or transitional differentiation state, potentially reflecting partial lineage erosion or evolving tumor plasticity. This observation is exploratory but raises the possibility that differentiation state exists along a continuum rather than as a strict binary classification.

Together, unsupervised clustering supports the existence of structured transcriptional states within NENs and provides additional evidence for a conserved differentiation axis that stratifies tumors independent of anatomical origin.

Orthogonal validation of differentiation-linked markers by immunohistochemistry

To validate whether transcriptomic differences translated to protein-level stratification, we performed immunohistochemistry (IHC) on an independent tissue microarray. Based on differential expression and biological relevance, we selected *EZH2* as a PD-NEC–associated marker and *PAK3* as a WD-NET–associated marker.

Consistent with transcriptomic findings, PD-NECs demonstrated significantly higher nuclear expression of EZH2 and its associated repressive chromatin mark H3K27me3 (Fig. 3A–B; Supplementary Fig. 3). EZH2 was selected due to robust mRNA upregulation (fold change = 4.2, adjusted p = 9.38 $\times 10^{-}5$) and its established role in Polycomb-mediated

epigenetic repression and proliferative signaling. Increased H3K27me3 further supports the presence of a chromatin-repressed, dedifferentiated transcriptional state in PD-NECs.

In contrast, WD-NETs exhibited significantly higher PAK3 expression (fold change = 7.7, adjusted p = 2.67 $\times 10^{-}8$), consistent with retention of neuronal and neuroendocrine lineage programs. Nuclear H-scores (0–300) confirmed subtype-specific protein expression patterns, demonstrating concordance between RNA and protein levels.

Importantly, antibody specificity for EZH2 was validated by depletion experiments in retinal pigment epithelial (RPE) cells (Supplementary Fig. 4A–C), strengthening confidence in the observed staining patterns.

These findings provide orthogonal protein-level validation of the transcriptomic differentiation axis and demonstrate that molecular subtype signatures are reflected in stable chromatin and lineage-associated protein states.

Interpretable machine learning resolves a conserved differentiation axis

Distinguishing poorly differentiated NECs from high-grade WD-NETs remains a major diagnostic challenge, particularly in morphologically ambiguous tumors. To test whether transcriptional state alone could robustly encode histologic identity, we trained two interpretable classifiers—logistic regression (LR) and random forest (RF)—on normalized NanoString expression profiles (~800 genes).

Both models achieved stable discrimination under stratified cross-validation (accuracy 87–93%), indicating that differentiation state is strongly encoded within the transcriptome (Fig. 4A). Importantly, these models were not black boxes: feature weights revealed coherent biological programs underlying classification.

As an external sanity check, we applied the trained classifier to an independent pancreatic WD-NET dataset (GSE118014). All samples were predicted as WD-NET, consistent with known histology, supporting model specificity rather than overfitting.

Rather than merely separating classes, the models exposed a biologically coherent transcriptional axis that recapitulates the proliferative-versus-lineage-preserved dichotomy observed in the unsupervised analyses (Fig. 1B).

Convergent feature attribution identifies core subtype-defining genes

Across modeling strategies, a striking convergence emerged.

PD-NEC–associated features. Positive LR coefficients were dominated by regulators of DNA replication (*CDC6*, *MCM* family), checkpoint control (*CHEK1*, *SFN*), chromatin regulation (*HIST1H3B*), Fanconi anemia/DNA repair (*FANCA*, *BRIP1*), and mitotic kinase signaling (*TTK*). *SFN* showed the strongest LR weight ($\beta = +0.0815$), with *CHEK1*, *E2F1*, and *CDC6* closely following. These genes anchor the G1/S transition and replication stress response network (Fig. 4H).

Random forest Gini importance independently highlighted *FANCA*, *CAMK2B*, *CACNA1H*, *BRIP1*, *PAK3*, *CACNB2*, *HIST1H3B*, *TTK*, *MCM2*, and *GNG7*. SHAP analysis confirmed that *FANCA*, *HIST1H3B*, *CDC6*, *TTK*, and *SFN* consistently drove PD predictions.

The intersection of LR, Gini, and SHAP metrics identified eight robust biomarkers: *FANCA*, *CDC6*, *HIST1H3B*, *BRIP1*, *DKK1*, *CACNB2*, *PAK3*, *CAMK2B*. This convergence across independent attribution strategies strengthens confidence that these genes reflect biological signal rather than model artifact.

WD-NET–associated features. Negative LR coefficients were led by *CAMK2B* ($\beta = -0.0962$), followed by *PAK3*, *GRIA3*, *FGF14*, *CACNA1D*, *MAPK10*, and *ZBTB16*. These genes are enriched in neuronal and synaptic signaling pathways and are highly expressed in brain tissues (Fig. 4E–F), consistent with preservation of neuroendocrine lineage identity. Notably, *PAK3* demonstrated concordant RNA and protein enrichment (Fig. 3), reinforcing biological validity.

Pathway architecture reveals a replication–lineage polarity

Pathway enrichment of PD-weighted genes revealed overwhelming dominance of cell-cycle programs (WP179; FDR ~$10^{-}31$), G1/S and G2/M transitions, ATR checkpoint signaling, Fanconi anemia, PI3K–Akt–mTOR signaling, and RB pathway dysregulation (Table 1; Fig. 4H). The core module centers on *CHEK1*, *SFN*, *CDC6*, *E2F1*, *TTK*, and *MCM* family members, forming a tightly connected replication stress and mitotic checkpoint axis.

In contrast, WD-NET–associated genes mapped to neuronal development and synaptic signaling programs, including calcium channel subunits (*CACNB2*, *CACNA1D*) and glutamatergic signaling genes (*GRIA3*). Together, these results define a replication-driven PD-NEC state versus a lineage-preserved WD-NET state, paralleling programs observed in small-cell lung carcinoma and other high-grade neuroendocrine carcinomas.

Transcriptome-guided inference of tumor origin

Given the low tumor mutational burden of many NENs, genomic profiling alone is often insufficient for origin inference. Using Pearson correlation–based KNN (K=1), we evaluated five tumors lacking annotated primary sites.

PD_0009 showed extremely high similarity to a small bowel NEC (r = 0.92). PD_0014 and PD_0015 matched pancreatic NEC (r > 0.9). PD_0001 demonstrated weaker similarity (r = 0.48), highlighting limitations of nearest-neighbor approaches in heterogeneous NECs. Correlations were generally higher among WD-NETs (e.g., r = 0.97 between WD_0006 and WD_0017), consistent with their tighter clustering in PCA space (Fig. 2B). While exploratory, these findings suggest transcriptomic proximity may assist in origin inference when appropriate reference cohorts exist.

Orthogonal validation via DNA methylation profiling

To test whether the transcriptional differentiation axis is reflected at the epigenetic level, we analyzed an independent lung NEN cohort profiled by EPIC methylation array (GSE211483). Principal component analysis of β-values separated PD-NECs from WD-NETs, with PC1 explaining ~19.2% of variance (Fig. 6A). This mirrors transcriptomic separation (Fig. 1B), suggesting cross-modal coherence.

PD-NECs exhibit promoter hypermethylation and Polycomb-associated repression. Differential methylation analysis (FDR < 0.05, $|\Delta\beta| > 0.2$) identified >2,000 DMPs enriched in PD-NECs. Promoter- and CpG island–associated hypermethylation was prominent in NECs, including developmental loci (e.g., *HOXA2*, *HOXA3*), and showed significant promoter/island bias (Fisher's exact p < $10^{-}3$). This pattern is consistent with a CpG island hypermethylator-like phenotype.

Methylation–expression concordance reinforces classifier biology. Several classifier-prioritized genes showed methylation–expression coupling: *MKI67* displayed higher expression and promoter hypomethylation in NECs (ΔExpr = 1.17; $\Delta\beta = -0.231$), *SFN* and *FANCA* were hypomethylated in NECs, and *CAMK2B*/*FGF14* showed NET-biased methylation patterns (Fig. 6C–F, J–N). Wilcoxon tests confirmed significance for multiple loci (e.g., *EZH2* p.adj = 0.0017; *GRIA3* p.adj = 0.0006).

Increased epigenetic stemness in PD-NECs. pcgtAge scores were highest in NECs, intermediate in NETs, and lowest in normal lung (Fig. 6I), paralleling the transcriptional proliferation axis and indicating a more dedifferentiated epigenetic state.

Integrated interpretation

Across transcriptomic profiling, interpretable machine learning, protein validation, and independent methylation datasets, a coherent model emerges. WD-NETs retain neuronal/neuroendocrine lineage programs, exhibit tighter molecular clustering, and display lower epigenetic stemness, whereas PD-NECs activate replication stress and mitotic checkpoint programs, upregulate DNA repair and chromatin modifiers, show promoter hypermethylation, and converge on a conserved proliferative carcinoma state. Together, these data support a conserved differentiation-linked axis that transcends tissue of origin.

Discussion

We show that transcriptomic profiles can support classification of NEN subtypes using interpretable machine learning. Logistic regression and random forest models achieved ~87–93% accuracy in cross-validation for PD-NEC versus WD-NET, and performed as expected on an external WD-NET dataset (GSE118014). This exceeds earlier limited panel approaches (e.g., ~71–78% accuracy in small intestinal NETs [29]), though differences in cohorts and assay design preclude direct benchmarking. We also observed a small cluster

with mixed features; given its size, this should be viewed as hypothesis-generating rather than a defined subtype.

Together, these results support a differentiation-linked axis. WD-NETs retain neuroendocrine lineage programs, whereas PD-NECs adopt a proliferative, epigenetically repressed state. The axis is detectable across transcriptomic and epigenetic layers and appears largely independent of tissue of origin.

PD-NECs were characterized by upregulation of cell-cycle and replication stress regulators (*SFN*, *CHEK1*, *E2F1*, *CDC6*, *TTK*), consistent with high proliferation [21]. WD-NETs showed enrichment of neuroendocrine lineage and synaptic genes (*CAMK2B*, *GRIA3*, *PAK3*, *FGF14*), aligning with a more differentiated phenotype. These results are concordant with observations that high-grade NECs share programs with small-cell lung cancer [21]. While promising, clinical deployment would require harmonized assays, calibrated thresholds, and prospective validation.

Many PD-NEC-associated features regulate DNA replication, checkpoint control, and chromatin state, suggesting a shared proliferative program, while WD-NET features reflect lineage and neuronal signaling programs.

We validated subtype-specific protein expression by IHC. *EZH2* and *H3K27me3* were higher in PD-NECs, while *PAK3* was higher in WD-NETs. External methylation data from lung NENs supported an epigenetic signature of dedifferentiation in NECs (promoter hypermethylation, higher pcgtAge, and methylation–expression shifts at *MKI67* and *SFN*). Recent reports also show broad expression differences between G3 NETs and NECs [22]. WD-NET markers in our study clustered in neuronal/secretory programs (*CAMK2B*, *GRIA3*, *BAIAP3*, *CACNA1D*), suggesting candidate biomarkers, though functional validation is needed to link these genes to NET biology.

To date, WD-NETs and PD-NECs have been considered distinct entities with different cells-of-origin and genetic drivers [19]. Genomic studies supporting this distinction highlight that WD PanNETs frequently harbor *MEN1*, *DAXX*/*ATRX*, and *mTOR* pathway mutations, whereas PD-NECs (in pancreas and elsewhere) often contain disruptive *TP53* and *RB1* alterations [19]. When examining mutation patterns we find similar trends where *TP53* mutation occurring at a higher frequency compared (Fig. 2a). A recent 2020 methylation analysis of pancreatic NENs by Simon et al. showed that PanNECs form a completely separate epigenetic cluster from PanNETs (including G3 well-differentiated NETs), implying an exocrine lineage origin for PanNECs as opposed to an endocrine islet cell origin for PanNETs [23]. This “cell-of-origin” divergence is further reflected in the transcriptomes, where we observed that NET samples (even high-grade) were uniformly classified by our model as NET-like, distinct from NECs. Nevertheless, the spectrum from NET to NEC may not be entirely clear. Rare cases of tumor progression from a low-grade NET to a high-grade, NEC-like phenotype have been documented. For instance, Joseph *et al.* (2024) reported a series of G1/G2 NETs that evolved into G3 neoplasms with acquired *TP53*/*RB1* co-mutations, blurring the line between G3 NET and true NEC [19]. In

their cohort, some progressed tumors even retained well-differentiated morphology despite harboring genetic hallmarks of NEC, illustrating how morphological and molecular changes can decouple in the evolution of NENs. Such cases are exceedingly uncommon (progression from NET to NEC is considered very rare [19], but they highlight that subtype transitions, while infrequent, are possible. More commonly, what appears as "NET to NEC" transformation in the clinic may turn out to be a misclassification or a mixed tumor scenario. For example, in gastroenteropancreatic tumors, one can encounter mixed neuroendocrine-nonneuroendocrine neoplasms (MiNENs) where a NET and a carcinoma co-exist [21]. Additionally, outside the GI/Lung context, some non-neuroendocrine carcinomas can transdifferentiate into a neuroendocrine phenotype under therapy pressure – a phenomenon known as neuroendocrine lineage plasticity [39]. This is well documented in prostate cancer treated with androgen blockade, where emergent neuroendocrine prostate cancer (NEPC) shares molecular programs with small-cell lung cancer [40].

The existence of NE lineage plasticity in other epithelial cancers suggests that an epithelial tumor cell can switch on a neuroendocrine program (often via epigenetic reprogramming [41], effectively *mimicking* a PD-NEC. In the context of NENs, however, most PD-NECs appear to arise *de novo* rather than from pre-existing NETs [21]. Our data support this paradigm where the clear bifurcation in gene expression and DNA methylation profiles between WD and PD classes implies separate evolutionary trajectories. Going forward, it will be important to explore whether the few bona fide NET-to-NEC progression cases exhibit intermediate "warning" molecular signatures – potentially something our classifier (or its underlying feature genes) could detect. Monitoring high-grade NETs for acquisition of NEC-like molecular traits (such as a surge in cell-cycle gene expression or loss of endocrine markers) might help flag patients at risk of aggressive transformation.

Several limitations should be noted. The discovery cohort is modest in size and heterogeneous in primary site, and the models were trained on a targeted NanoString panel rather than whole-transcriptome profiling. While we used cross-validation and external datasets, the analysis remains retrospective and subject to selection and batch effects. Larger, prospectively collected cohorts with harmonized assays will be required to define clinically actionable thresholds, evaluate generalizability across sites, and test whether classifier-guided decisions improve outcomes.

More broadly, this framework illustrates how interpretable transcriptomic models can expose conserved cellular states that cut across anatomical context. Similar differentiation–proliferation axes have been reported in other epithelial cancers with lineage plasticity, suggesting these approaches may generalize beyond NENs.

Conclusions

This study demonstrates that neuroendocrine neoplasms (NENs) can be robustly stratified into well-differentiated NETs and poorly differentiated NECs using transcriptomic data and interpretable machine learning. By integrating logistic regression, random forest classifiers, and SHAP-based feature attribution, we identified reproducible gene signatures, including *SFN*, *FANCA*, *HIST1H3B*, *CHEK1*, and *CAMK2B*, that discriminate histologic subtypes with high accuracy. These markers reflect underlying biological differences, such as cell cycle activation in PD-NECs and neuronal identity in WD-NETs.

Importantly, we validated these findings across multiple levels: (1) orthogonal protein expression by immunohistochemistry, (2) inference of unknown tumor primaries via transcriptomic proximity, and (3) independent external validation using DNA methylation data from a separate lung NEN cohort. The concordant separation of subtypes across expression, methylation, and machine learning predictions underscores the robustness of these molecular signatures.

By combining predictive performance with biological interpretability, our approach offers both diagnostic utility and mechanistic insight. These results highlight the potential for AI-guided classifiers to augment histopathological diagnosis, uncover new biomarkers, and resolve ambiguous or mixed NEN cases. More broadly, this framework can be adapted to other cancer types where histologic grading is subjective, incomplete, or difficult to resolve.

Methods

Study design and oversight

This was a single-center, retrospective study conducted at Fox Chase Cancer Center (FCCC) to develop and validate molecular classifiers of neuroendocrine neoplasm (NEN) histology. Formalin-fixed, paraffin-embedded (FFPE) tumor specimens were profiled by targeted transcriptomics; selected markers were validated by immunohistochemistry (IHC) on tissue microarrays (TMAs). External orthogonal validation was performed using public DNA methylation datasets. All procedures were approved by the FCCC Institutional Review Board, and written informed consent was obtained from all participants. The study was exploratory in nature; sample size was determined by tissue availability for the rare poorly differentiated (PD) NEC subtype rather than a priori power calculations.

Cohort and tissue microarrays

Tumor specimens were obtained from patients with well-differentiated (WD) neuroendocrine tumors (NETs) and PD neuroendocrine carcinomas (NECs) from multiple primary sites. Clinicopathologic data were de-identified prior to analysis. TMAs were constructed

from 36 FFPE blocks (21 WD, 15 PD). Hematoxylin-and-eosin–stained sections were reviewed by a board-certified pathologist to mark representative tumor regions. Duplicate cores (diameter 1 mm) were taken per case using a precision arrayer (Beecher Instruments, Silver Spring, MD, USA) and placed on separate recipient blocks to mitigate positional bias. Tissue had been fixed in 10% phosphate-buffered formaldehyde (24-48 h), processed, and embedded per standard protocols. Each TMA core was scored individually; duplicate core values were averaged per case.

Immunohistochemistry and quantitative image analysis

IHC was performed on 5 µm TMA sections. Slides were deparaffinized, rehydrated, and subjected to heat-induced epitope retrieval in 0.01 M citrate buffer (pH 6.0) for EZH2, H3K27me3, PAK3, and pCHK1, or EDTA buffer (for pSTAT3). Endogenous peroxidase activity was quenched in 3% H_2O_2. Sections were incubated overnight at 4 °C with primary antibodies: EZH2 (D2C9, rabbit, 1:50; Cell Signaling Technology, #5246), Tri-methyl-Histone H3 (Lys27) (C36B11, rabbit, 1:50; Cell Signaling, #9733), PAK3 (A-20, rabbit, 1:500; Invitrogen, PA5-15118), phospho-Chk1 (Ser345, rabbit, 1:500; Thermo Fisher, PA5-34625), and phospho-STAT3 (Tyr705, D3A7, rabbit, 1:30; Cell Signaling, #9145). Detection used the EnVision+ polymer system (Dako/Agilent) with 3,3′-diaminobenzidine (DAB) chromogen. Slides were counterstained with hematoxylin, dehydrated, cleared in xylene, and mounted. Negative controls used matched isotype control IgG.

Whole-slide images were acquired on a PerkinElmer CRi Vectra 2 platform. Multispectral images were unmixed and analyzed in inForm (Akoya Biosciences) using a trainable workflow: (i) tissue segmentation to identify neuroendocrine tumor regions and (ii) cell segmentation to delineate hematoxylin-positive nuclei and DAB-positive subcellular compartments. Staining intensity was binned (0/1+/2+/3+), and H-scores were computed as 1×(%1+) + 2×(%2+) + 3×(%3+) (range 0–300). Nuclear and cytoplasmic H-scores were derived as appropriate. Antibody specificity for EZH2 was confirmed by loss of signal following EZH2 depletion in retinal pigment epithelial (RPE) cells.

Transcriptomic profiling (NanoString)

FFPE sections were reviewed to enrich for invasive tumor areas. Total RNA was extracted using the High Pure FFPET RNA Isolation Kit (Roche) following the manufacturer's protocol. RNA was hybridized overnight to probes of the NanoString nCounter PanCancer Immune Profiling Panel with Panel Plus customization (784 genes spanning canonical cancer pathways, including MAPK, STAT, PI3K, RAS, Cell Cycle, Apoptosis, Hedgehog, Wnt, DNA Damage Control, Transcriptional Regulation, Chromatin Modification, and TGF-β). [27, 28] purified, and digitally counted on the nCounter platform (NanoString Technologies). Manufacturer controls were used to assess RNA quality and run performance. Raw Reporter Code Count (RCC) files were used for downstream analysis.

Pre-processing and differential expression analysis

Raw counts were normalized and transformed in R. Background correction and positive/negative control normalization were applied as per NanoString recommendations. Library size normalization and variance modeling were performed with voom (limma), and log-counts per million (logCPM) values were computed (edgeR). Differential expression was assessed with linear modeling and empirical Bayes moderation (limma). Multiple testing correction used the Benjamini–Hochberg false discovery rate (FDR); unless stated otherwise, significance thresholds were FDR < 0.05 and $|\log2FC| \geq 1$. Principal component analysis (PCA) was performed on scaled logCPM matrices. Unsupervised hierarchical clustering used Euclidean distances and Ward's linkage. All scripts are available in the GitHub project repository.

Machine-learning classification of histology

Histologic class (PD vs WD) was modeled using logistic regression (LR) and a random forest classifier (RFC) implemented in scikit-learn (Python). Features comprised normalized transcript levels (~800 genes). Data were stratified to preserve class balance. Model evaluation used stratified 5-fold cross-validation, reporting accuracy, balanced accuracy, precision/recall, F1, and ROC-AUC. Hyperparameters were tuned via RandomizedSearchCV with internal cross-validation. Confusion matrices and ROC/PR curves were generated for interpretability.

Feature importance and model interpretability

Global and local feature attributions were derived from: (i) LR coefficients (L1/L2/elastic-net regularization), (ii) RFC Gini importance, and (iii) SHAP (Shapley Additive exPlanations) values computed on the RFC using mean absolute class-specific SHAP. Robust markers were identified by intersecting features meeting pre-specified thresholds ($|\beta| \geq 0.04$ for LR; Gini ≥ 0.001; mean SHAP (PD class) ≥ 0.001). Gene-level summaries and per-sample SHAP plots are provided in Supplementary materials.

Pathway enrichment analysis

Pathway analysis was performed in Cytoscape (v3.10.1) using gene sets from LR (positive/negative coefficients under L1/L2/elastic-net) and from RFC SHAP/Gini (PD- and WD-enriched features analyzed separately). Enrichment used pathway databases (e.g., WikiPathways/Reactome) and network summarization (e.g., EnrichmentMap) with similarity metrics to cluster related pathways. Reported statistics include similarity scores and FDR-adjusted p-values. Cell-cycle modules (WP179, G1→S, G2/M, ATR) and PI3K–Akt/RB pathways were prioritized based on effect size and recurrence across models.

Primary-site inference for tumors of unknown origin

For five tumors lacking annotated primaries, site inference used k-nearest neighbors (K = 1) on the full logCPM feature space. Similarity was quantified by Pearson correlation; distance was defined as $1 - r$. Each unknown was assigned the primary site of its nearest labeled neighbor. Pairwise correlations were visualized in a heatmap with nearest-neighbor links highlighted. For selected pairs, quantile–quantile (QQ) plots visualized concordance. Sensitivity analyses varying K and correlation/distance metrics are provided in Supplementary materials.

External validation using DNA methylation data

Independent validation was performed on public lung NEN methylation data (Illumina Infinium MethylationEPIC; GSE211483). Raw IDAT files were processed in R with minfi/ChAMP. QC included control probes and total signal; probes with detection $p > 0.01$ in any sample, sex chromosome probes, and probes with known SNPs at the CpG or extension site were removed. Background correction and dye-bias normalization used noob; BMIQ harmonized type I/II probe distributions. β-values were used for visualization; M-values [$\log2(\beta/(1-\beta))$] were used for statistical testing where appropriate.

Unsupervised analyses (PCA and hierarchical clustering) were performed on filtered β-matrices without variance preselection for the primary figures; sensitivity analyses using variable CpGs were also evaluated. Differentially methylated positions (DMPs) were identified using linear models with empirical Bayes moderation and FDR control (typically FDR < 0.05 and $|\Delta\beta| \geq 0.2$). Genomic context enrichment (promoters, CpG islands, shores/shelves) was evaluated, and Polycomb target–based stemness (pcgtAge) scores were computed. Where matched RNA data (GSE211486) were available, methylation–expression relationships were assessed by Wilcoxon tests with Benjamini–Hochberg adjustment.

Statistical analyses

All statistical analyses were performed in R (version ≥ 4.1) and Python (version ≥ 3.10). Unless otherwise noted, two-sided tests were used with FDR control by Benjamini–Hochberg. Continuous variables are summarized as median (IQR) unless specified. Group comparisons for IHC H-scores used Wilcoxon rank-sum tests with FDR adjustment; correlations used Pearson's r. Plots were generated using ggplot2, ComplexHeatmap, matplotlib, and seaborn. Reproducibility was supported by scripted analyses, version-controlled code, and fixed random seeds where applicable.

Software and versions

Key software included: R (≥ 4.1), Bioconductor (minfi, ChAMP, limma, edgeR), Python (≥ 3.10), scikit-learn (≥ 1.3) for LR/RFC/CV, shap (≥ 0.43) for SHAP, and Cytoscape (v3.10.1). Exact package versions and command-line invocations are documented in the repository.

Competing Interests

The authors declare that they have no competing interests.

Funding

This research was supported by new faculty funds from Fox Chase Cancer Center and the Cancer Center Support Grant P30CA006927.

Authors Contributions

H.L. and N.V. conceptualized the study. P.F.E. designed the transcriptome study. S.P. and J.M. performed transcriptome experiments. G.Y. and C.Q. conducted histological analysis. T.K., K.C and H.L. performed computational analyses and developed machine learning models. H.L. implemented the visualization application. T.K., H.L., S.P., K.C., J.W., I.A., and N.V. contributed to manuscript writing. H.L. and N.V. supervised the study. All authors read and approved the final manuscript.

Data Availability

Formalin-fixed, paraffin-embedded (FFPE) tumor tissues were collected under IRB-approved protocols at Fox Chase Cancer Center. Additional access may be available upon reasonable request, subject to institutional review and material transfer agreement. Gene expression data generated in this study are provided in Supplementary Data 1. Public datasets used include GSE118014 [30], methylation data from GSE211483 [23], and matched transcriptome data from GSE211486 [23]. Analysis scripts, trained models, and visualization tools are available in the project repository (link in Supplementary Materials).

References

[1] AM Schmitt et al. “New Genetics and Genomic Data on Pancreatic Neuroendocrine Tumors: Implications for Diagnosis, Treatment, and Targeted Therapies”. In: *Endocr Pathol* 27 (2016), pp. 200–4.

[2] A Scarpa et al. “Whole-genome landscape of pancreatic neuroendocrine tumours”. In: *Nature* 543 (2017), pp. 65–71.

[3] SV Laddha et al. “Integrative Genomic Characterization Identifies Molecular Subtypes of Lung Carcinoids”. In: *Cancer Res* 79 (2019), pp. 4339–47.

[4] C Marchiò et al. “Distinctive pathological and clinical features of lung carcinoids with high proliferation index”. In: *Virchows Arch* 471 (2017), pp. 713–20.

[5] A Mafficini and A Scarpa. “Genetics and Epigenetics of Gastroenteropancreatic Neuroendocrine Neoplasms”. In: *Endocr Rev* 40 (2019), pp. 506–36.

[6] A Puccini et al. “Comprehensive Genomic Profiling of Gastroenteropancreatic Neuroendocrine Neoplasms (GEP-NENs)”. In: *Clin Cancer Res* 26 (2020), pp. 5943–51.

[7] 9. “Das S, Dasari A”. In: *Epidemiology, Incidence, and Prevalence of Neuroendocrine Neoplasms: Are There Global Differences? Curr Oncol Rep* 23 (2021), p. 43.

[8] BE White et al. “Incidence and survival of neuroendocrine neoplasia in England 1995-2018: A retrospective, population-based study”. In: *Lancet Reg Health Eur* 23 (2022), p. 100510.

[9] JY Zhang and PL Kunz. “Making Sense of a Complex Disease: A Practical Approach to Managing Neuroendocrine Tumors”. In: *JCO Oncology Practice* 18 (2022), pp. 258–64.

[10] MH Shah et al. “Neuroendocrine and Adrenal Tumors, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology”. In: *J Natl Compr Canc Netw* 19 (2021), pp. 839–68.

[11] 15. “Klimstra DS, Beltran H, Lilenbaum R, Bergsland E”. In: *The Spectrum of Neuroendocrine Tumors: Histologic Classification, Unique Features and Areas of Overlap* American Society of Clinical Oncology Educational Book (0).

[12] O Basturk et al. “Poorly differentiated neuroendocrine carcinomas of the pancreas: a clinicopathologic analysis of 44 cases”. In: *Am J Surg Pathol* 38 (2014), pp. 437–47.

[13] JM Zuetenhorst and BG Taal. “Metastatic carcinoid tumors: a clinical review”. In: *Oncologist* 10 (2005), pp. 123–31.

[14] A Dasari et al. “Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States”. In: *JAMA Oncol* 3 (2017), pp. 1335–42.

[15] LH Tang et al. “A Practical Approach to the Classification of WHO Grade 3 (G3) Well-differentiated Neuroendocrine Tumor (WD-NET) and Poorly Differentiated Neuroendocrine Carcinoma (PD-NEC) of the Pancreas”. In: *Am J Surg Pathol* 40 (2016), pp. 1192–202.

[16] ID Nagtegaal et al. “The 2019 WHO classification of tumours of the digestive system”. In: *Histopathology* 76 (2020), pp. 182–8.

[17] MS Banck et al. “The genomic landscape of small intestine neuroendocrine tumors”. In: *J Clin Invest* 123 (2013), pp. 2502–8.

[18] 20. “Couvelard A, Cazes A, Cros J”. In: *Updates in histopathological classification and tissue biomarkers of digestive neuroendocrine neoplasms: What the clinician should know* Best Pract Res Clin Endocrinol Metab (0).

[19] NM Joseph et al. “Progression of Low-Grade Neuroendocrine Tumors (NET) to High-Grade Neoplasms Harboring the NEC-Like Co-alteration of RB1 and TP53”. In: *Endocr Pathol* 35 (2024), pp. 325–37.

[20] J van Riet et al. “The genomic landscape of 85 advanced neuroendocrine neoplasms reveals subtype-heterogeneity and potential therapeutic targets”. In: *Nat Commun* 12 (2021), p. 4612.

[21] M Frizziero et al. “Expanding Therapeutic Opportunities for Extrapulmonary Neuroendocrine Carcinoma”. In: *Clin Cancer Res* 28 (2022), pp. 1999–2019.

[22] V Angerilli et al. “Comprehensive genomic and transcriptomic characterization of high-grade gastro-entero-pancreatic neoplasms”. In: *Br J Cancer* 131 (2024), pp. 159–70.

[23] T Simon et al. “DNA methylation reveals distinct cells of origin for pancreatic neuroendocrine carcinomas and pancreatic neuroendocrine tumors”. In: *Genome Med* 14 (2022), p. 24.

[24] WM Hackeng et al. “Genome Methylation Accurately Predicts Neuroendocrine Tumor Origin: An Online Tool”. In: *Clin Cancer Res* 27 (2021), pp. 1341–50.

[25] A Tirosh et al. “Distinct DNA Methylation Signatures in Neuroendocrine Tumors Specific for Primary Site and Inherited Predisposition”. In: *J Clin Endocrinol Metab* 105 (2020), pp. 3285–94.

[26] A Di Domenico et al. “Epigenetic landscape of pancreatic neuroendocrine tumours reveals distinct cells of origin and means of tumour progression”. In: *Communications Biology* 3 (2020), pp. 1–11.

[27] Alessandra Cesano. “nCounter® PanCancer Immune Profiling Panel (NanoString Technologies, Inc., Seattle, WA)”. In: *Journal for ImmunoTherapy of Cancer* 3 (2015), p. 42.

[28] NanoString Technologies. *nCounter® PanCancer Immune Profiling Panel*. 2026. URL: `https://nanostring.com/products/ncounter-assays-panels/ncounter-rna-assays/oncology/pancancer-immune-profiling/` (visited on 02/04/2026).

[29] I Drozdov et al. “Predicting neuroendocrine tumor (carcinoid) neoplasia using gene expression profiling and supervised machine learning”. In: *Cancer* 115 (2009), pp. 1638–50.

[30] CS Chan et al. “ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup”. In: *Nat Commun* 9 (2018), p. 4158.

[31] V Melone et al. “Identification of functional pathways and molecular signatures in neuroendocrine neoplasms by multi-omics analysis”. In: *J Transl Med* 20 (2022), p. 306.

[32] JY Jung et al. “14-3-3 Sigma Protein Contributes to Hepatocyte Growth Factor-mediated Cell Proliferation and Invasion Matrix Metalloproteinase-1 Regulation in Human Gastric Cancer”. In: *Anticancer Res* 42 (2022), pp. 519–30.

[33] H Huang et al. “TTK regulates proliferation and apoptosis of gastric cancer cells through the Akt-mTOR pathway”. In: *FEBS Open Bio* 10 (2020), pp. 1542–9.

[34] J Krushkal et al. “Epigenome-wide DNA methylation analysis of small cell lung cancer cell lines suggests potential chemotherapy targets”. In: *Clin Epigenetics* 12 (2020), p. 93.

[35] LS Pongor et al. “Integrative epigenomic analyses of small cell lung cancer cells demonstrates the clinical translational relevance of gene body methylation”. In: *iScience* 25 (2022), p. 105338.

[36] P Shannon et al. “Cytoscape: a software environment for integrated models of biomolecular interaction networks”. In: *Genome Res* 13 (2003), pp. 2498–504.

[37] AM van Harten et al. “Targeting the cell cycle in head and neck cancer by Chk1 inhibition: a novel concept of bimodal cell death”. In: *Oncogenesis* 8 (2019), pp. 1–16.

[38] Y Jiao et al. “DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors”. In: *Science* 331 (2011), pp. 1199–203.

[39] A Sreekumar and S Saini. “Role of transcription factors and chromatin modifiers in driving lineage reprogramming in treatment-induced neuroendocrine prostate cancer”. In: *Front Cell Dev Biol* 11 (2023), p. 1075707.

[40] P Cejas et al. “Subtype heterogeneity and epigenetic convergence in neuroendocrine prostate cancer”. In: *Nat Commun* 12 (2021), p. 5775.

[41] 52. “Colapietra F, Della Monica P, Di Napoli R, França Vieira E Silva F, Settembre G, Marino MM, et al”. In: *Epigenetic Modifications as Novel Biomarkers for Diagnosis, Prognosis, and Therapeutic Targeting in Thyroid, Pancreas, and Lung Neuroendocrine Tumors* J Clin Med [Internet] (0).