Artificial intelligence-based histopathology image analysis identifies a novel subset of endometrial cancers with distinct genomic features and unfavourable outcome

Amirali Darbandsari[1]*, Hossein Farahani[2,3]*, Matthew Wiens[2], Dawn Cochrane[4], Maryam Asadi[2], Ali Khajegili Mirabadi[2], Amy Jamieson[5], David Farnell[3,6], Pouya Ahmadvand[2], Maxwell Douglas[4], Samuel Leung[4], Purang Abolmaesumi[1], Steven JM Jones[7], Aline Talhouk[5], Stefan Kommoss[8], C Blake Gilks[3,6], David G. Huntsman[3,4]‡, Naveena Singh[3,6]‡, Jessica N. McAlpine[5]‡, Ali Bashashati[2,3]‡†

* equal contribution

‡ These authors jointly supervised this work

† Corresponding author

Affiliations:

[1]Department of Electrical and Computer Engineering, University of British Columbia; Vancouver, BC, Canada.

[2]School of Biomedical Engineering, University of British Columbia; Vancouver, BC, Canada.

[3]Department of Pathology and Laboratory Medicine, University of British Columbia; Vancouver, BC, Canada.

[4]Department of Molecular Oncology, British Columbia Cancer Research Center; Vancouver, BC, Canada.

[5]Department of Gynaecology and Obstetrics and Division of Gynaecologic Oncology, University of British Columbia; Vancouver, BC, Canada.

[6]Vancouver General Hospital; Vancouver, BC, Canada.

[7]Michael Smith Genome Sciences Center, British Columbia Cancer Research Center; Vancouver, BC, Canada.

[8]Department of Women's Health, Tübingen University Hospital; Tübingen, Germany.

Correspondence: Ali Bashashati. Email: ali.bashashati@ubc.ca. Mailing address: 2222 Health Sciences Mall, Vancouver, BC V6T 1Z3 Canada

Abstract: Endometrial cancer (EC) has four molecular subtypes with strong prognostic value and therapeutic implications. The most common subtype (NSMP; No Specific Molecular Profile) is assigned after exclusion of the defining features of the other three molecular subtypes and includes patients with heterogeneous clinical outcomes. In this study, we employed artificial intelligence (AI)-powered histopathology image analysis to differentiate between p53abn and NSMP EC subtypes and consequently identified a novel sub-group of NSMP EC patients that had markedly inferior progression-free and disease-specific survival (termed 'p53abn-like NSMP'), in a discovery cohort of 368 patients and an independent validation cohort of 290 patients from another center. Shallow whole genome sequencing revealed a higher burden of copy number abnormalities in the 'p53abn-like NSMP' group compared to NSMP, suggesting that this new group is biologically distinct compared to other NSMP ECs. Our work demonstrates the power of AI to detect prognostically different and otherwise unrecognizable subsets of EC where conventional and standard molecular or pathologic criteria fall short, refining image-based tumor classification.

Keywords: endometrial cancer, computational pathology, digital pathology, molecular classification, Artificial intelligence

Introduction

The clinicopathological parameters used for decades to classify endometrial cancers (EC) and guide management have been sub-optimally reproducible, particularly in high-grade tumors[1,2]. Specifically, inconsistency in grade and histotype assignment has yielded inaccurate assessment of the risk of disease recurrence and death. As a result, many women affected by EC may be over-treated or are not directed to treatment that might have reduced their risk of recurrence. In 2013, the Cancer Genome Atlas (TCGA) project demonstrated that endometrial cancers could be stratified into four distinct prognostic groups using a combination of whole genome and exome sequencing, microsatellite instability (MSI) assays, and copy number analysis[3]. These subtypes were labelled according to dominant genomic abnormalities and included 'ultra-mutated' ECs harboring *POLE* mutations, 'hypermutated' identified to have microsatellite instability, copy-number low, and copy-number high endometrial cancers.

Inspired by this initial discovery, our team and a group from the Netherlands independently and concurrently developed a pragmatic, clinically applicable molecular classification system that classifies ECs into : (i) *POLE* mutant (*POLE*mut) with pathogenic mutations in the exonuclease domain of *POLE* (DNA polymerase epsilon, involved in DNA proofreading repair), (ii) mismatch repair deficient (MMRd) diagnosed based on the absence of key mismatch repair proteins on immunohistochemistry (IHC), (iii) p53 abnormal (p53abn) as assessed by IHC, and (iv) NSMP (No Specific Molecular Profile), lacking any of the defining features of the other three subtypes[4,5]. Categorization of ECs into these subtypes recapitulates the survival curves/prognostic value of the four TCGA molecular subgroups and enhances histopathological evaluation, offering an objective and reproducible classification system with strong prognostic value and therapeutic implications. In 2020, the World Health Organization (WHO) recommended integrating these key molecular features into standard pathological reporting of ECs when available[6].

*POLE*mut endometrial cancers have highly favorable outcomes with almost no deaths due to disease. While the three other molecular subtypes are associated with more variable outcomes (MMRd and NSMP are considered 'intermediate risk' and p53abn ECs have the worst prognosis), within each subtype there are clinical and prognostic outliers[7–10]. This is particularly true within the largest subtype, NSMP (representing ~50% of ECs). The majority of NSMP tumors are early stage, low grade, estrogen driven tumors likely cured by surgery alone. However, a subset of patients with NSMP EC experience a very aggressive disease course, comparable to what is observed in patients with p53abn ECs. At present, limited tools exist to identify these aggressive outliers and current clinical guidelines do not stratify or

direct treatment within NSMP EC beyond using pathologic features[11,12]. Thus, for half of diagnosed endometrial cancers, i.e., NSMP EC, assumption of indolence is inappropriate and clinicians need tools for accurate risk stratification of individual patients when making treatment decisions.

With the rise of artificial intelligence (AI) in the past decade, deep learning methods (e.g., deep convolutional neural networks and their extensions) have shown impressive results in processing text and image data[13]. The paradigm shifting ability of these models to learn predictive features from raw data presents exciting opportunities with medical images, including digitized histopathology slides[14–17]. In recent years, these models have been deployed to reproduce or improve pathology diagnosis in various disease conditions (e.g.,[18–20]), explore the potential link between histopathologic features and molecular markers in different cancers including EC[17,21–24] , and directly link histopathology to clinical outcomes[25–28]. More specifically, two recent studies have reported promising results in the application of deep learning-based models to identify the four molecular subtypes of EC from histopathology images.

Building on a recent study reporting morphological heterogeneity in NSMP ECs and nuclear features typical of p53abn ECs in this subtype[29], we built a deep learning-based image classifier to differentiate between the NSMP and p53abn ECs. We then hypothesized that within the NSMP molecular subtype of endometrial cancer, there is a subset of patients with aggressive disease whose tumors have histological features similar to p53abn EC and that these tumors can be identified by deep learning models applied to hematoxylin & eosin (H&E)-stained slides. Our results show that these cases (referred to as *p53abn-like NSMP*) have inferior outcomes compared to the other NSMP ECs, similar to that of p53abn EC, in two independent cohorts. Furthermore, shallow whole genome sequencing studies suggested that the genomic architecture of the *p53abn-like NSMP* differs from other NSMP ECs, showing increased copy number abnormalities, a characteristic of p53abn EC.

Results

Patient cohort selection and description. 1,678 H&E-stained hysterectomy tissue sections from 658 patients with histologically confirmed endometrial carcinoma of NSMP or p53abn subtypes were included in this study[3–5]. Our discovery cohort included 155 whole section slides (WSI) from 146 patients from TCGA[3] and 431 WSIs (222 patients) from another center[5]. Our validation set included tissue microarray (TMA) data corresponding to 290 patients from our own center[4]. Tables 1 and 2 show the clinicopathological features of the discovery and validation cohorts.

Histopathology-based machine learning classifier to differentiate NSMP and p53abn ECs. Fig. 1 depicts our AI-based histopathology image analysis pipeline. A subset of 27 whole section H&E slides from the TCGA cohort were annotated by a board-certified pathologist (DF) using a custom in-house histopathology slide viewer (*cPathPortal*) to identify areas containing tumor and stromal cells. A deep convolutional neural network (CNN)-based classifier was then trained to acquire pseudo-tumor and benign annotations for the remaining slides in the discovery cohort. The identified tumor regions were then divided into 512x512 pixel patches at 20x objective magnification. The number of extracted patches from each subtype and performance measure for the tumor stroma classifier can be found in Supplementary Tables 1 and 2. To address variability in slide staining due to differences in staining protocols across different centres, and inter-patient variability, we utilized the Vahadane color normalization technique[30]. We then trained a VarMIL model[31] based on multiple instance learning (MIL) to differentiate H&E image patches associated with p53abn and NSMP ECs.

In a "group 10-fold" cross-validation strategy, the patients in our discovery cohort were divided into 10 groups and in various combinations, 60% were used for training, 20% for validation, and 20% for testing; resulting in 10 different binary p53abn vs. NSMP classifiers. These 10 classifiers were then used to label the cases as p53abn or NSMP and their consensus was used to come up with a label for a given case. For patients with multiple slides, to prevent data leakage between training, validation, and test sets, we assigned slides from each patient to only one of these sets.

Fig. 2A and Supplementary Table 3 show the receiver operating characteristics (ROC) and precision/recall curves as well as performance metrics of the resulting classifiers for the discovery and validation sets, respectively. These results suggest that our p53abn vs. NSMP classifier achieves 89·4% and 79·8% mean balanced accuracy (across the 10 classifiers) and area under the curve (AUC) of 0.95 and 0.88 in both the discovery and validation sets, respectively (for details see Supplementary Tables 3 & 4 and Supplementary Fig. 1).

Identification of a subset of NSMP ECs with inferior survival. Our proposed ML-based models classified 17·65% and 20% of NSMPs as p53abn for the discovery and validation cohorts, respectively (Supplementary Table 5). These cases (referred to as *p53abn-like NSMP* group) presumably show p53abn histological features in the assessment of H&E images even though immunohistochemistry did not show mutant-pattern p53 expression and these were therefore classified as NSMP by the molecular classifier. We hypothesized that such cases may in fact exhibit similar clinical behavior as p53abn ECs.

Fig. 2B,C show the progression free survival (PFS) and disease specific survival (DSS) of the discovery and validation sets. Compared to the rest of the NSMP cases, *p53abn-like NSMP*s had markedly inferior PFS (10-year PFS 55·7% vs. 89·6% (p < 2.7e-7)) and DSS (10-year DSS 62·6% vs. 93·7% (p < 1·8e-7)) in our discovery cohort. These findings were confirmed in the validation cohort, with 20% of the 195 patients categorized as p53abn-like tumors, showing 10-year PFS of 65·4% vs. 91·2% (p < 1·1e-4) and DSS of 58·3% vs. 84·3% (p < 5·3e-5). Additionally, comparison of the PFS and DSS between *p53abn-like NSMP* and p53abn ECs revealed a trend, though not statistically significant, in which *p53abn-like NSMP*s had better outcome compared to p53abn ECs in both the discovery and validation cohorts (Supplementary Fig. 2A,B).

Of note, our model also identified a subset of p53abn ECs (representing 20%; referred to as *NSMP-like p53abn*) with resemblance to NSMP as assessed by H&E staining. While we observed marginally superior disease-specific survival in the identified cases compared to the rest of the p53abn group both in the discovery and validation cohorts, progression free survival was not significantly different between the groups (Supplementary Fig. 3A,B).

Robustness of *p53abn-like NSMP* subtype. Our proposed deep learning-based model was built to differentiate between NSMP and p53abn EC subtypes. Given that these subtypes are determined based on molecular assays, their accurate identification from routine H&E-stained slides would have removed the need to perform molecular testing that might only be available in specialized centers. However, our observation of imperfect results and characterization of discordant cases as *p53abn-like NSMP* required further investigation to rule out the possibility of a more superior deep learning model that could result in a better performance in differentiating p53abn and NSPMP molecular subtypes. Therefore, we implemented five other deep learning-based image analysis strategies to test the stability of the identified classes (see Methods section for further details). Our results showed that these models also achieve balanced accuracies ranging from 83.5-88.6% and 77.3-80.2% and AUCs ranging from 0.88-0.95 and 0.8-0.88 in both the discovery and validation sets, respectively (Supplementary Fig. 4 and Supplementary Tables 6-7). Furthermore, Kaplan-Meier survival analysis of the so-called *p53abn-like NSMP* group identified by these models also corroborated with our initial findings in which this new subgroup had statistically significant inferior survival compared to the rest of the patients (Supplementary Fig. 5). These results suggest that the choice of the algorithm did not substantially affect the findings and outcome of our study. To further investigate the robustness of our results, we utilized an unsupervised approach in which we extracted histopathological features from the slides in our validation cohort utilizing KimiaNet[32] feature representation. Our results suggested that p53abn-like NSMP and the rest of the NSMP cases

constitute two separate clusters with no overlap (Figure 3A) suggesting that our findings could also be achieved with unsupervised approaches. It is noteworthy to mention that we utilized the original KimiaNet weights for feature extraction without any finetuning the model on our datasets.

Comparison of NSMP and *p53abn-like NSMP*. To further investigate the differences between NSMP and *p53abn-like NSMP* cases, we compared various clinical, pathological, and molecular variables (Supplementary Tables 8 and 9). Our analysis showed an enrichment of *p53abn-like NSMP* cases with higher grade and higher stage tumors (p < 1·4e-25; p < 2·4e-4, respectively). In a multi-variate Cox regression analysis, the association between *p53abn-like NSMP* and progression free survival remained significant in the presence of grade, stage, and histology (p = 0·01 and Hazard Ratio = 2·5; Supplementary Table 10). Furthermore, Fig. 3B shows an enrichment for estrogen receptor (ER) and progesterone receptor (PR) positive cases in the *p53abn-like NSMP*s in the subset of the cohort that the status of these markers were available (p < 5·2e-3 and p < 2·3e-4, respectively).

Independent pathology review of selected NSMP cases. Two expert gynepathologists (NS, CBG) independently reviewed whole section slides of a subset of NSMP cases including the *p53abn-like NSMP* subtype. They specifically assessed whether tumors showed nuclear features that have been previously described as being associated with *TP53* mutation/mutant pattern p53 expression in endometrial carcinoma[29]. The *p53abn-like NSMP* cases were enriched with tumors showing increased nuclear atypia, as assessed by altered chromatin pattern, nucleolar features, pleomorphism, atypical mitoses, or giant tumor cells (p < 0·00005 for both reviewers).

Genomic characterization of *p53abn-like NSMP* cases. We next sought to investigate the molecular profiles of *p53abn-like NSMP* cases in our validation set for which we had access to tissue material. Targeted sequencing of exonic regions in a number of genes revealed enrichment of *p53abn-like NSMP* cases with *CTNNB1* mutations (Fisher's exact test p-value = 0·01; Fig. 3B). However, exonic point mutation data for the TCGA subset of the discovery cohort were available and suggested a lack of enrichment of the *p53abn-like NSMP* group with specific gene mutations including *CTNNB1* (Fig. 3C).

We next selected representative samples of NSMP, p53abn, and *p53abn-like NSMP* cases in our validation cohort and performed shallow whole genome sequencing (sWGS). Overall, copy number profile analysis of these cases revealed that *p53abn-like NSMP* cases harbor a higher fraction of altered genome compared to NSMP cases but still lower than what we observe in p53abn cases (Fig. 4A; p < 0·035). These findings were further validated in the TCGA cohort (Fig. 4B; p < 5·46e-5).

We next investigated the gene expression profiles associated with the *p53abn-like NSMP*, NSMP, and p53abn tumors within the TCGA cohort. Unsupervised clustering of patients based on gene expression profiles of their tumors showed that eleven of the 21 *p53abn-like NSMPs* had similar expression profiles to p53abn tumors, while the remaining 10 cases clustered together with the NSMP group (Fig. 4C). While p53abn and NSMP groups were clustered separately, unsupervised analysis of the gene expression profiles did not reveal any differences between *p53abn-like NSMP* group and other subtypes, i.e., they did not have a unique gene expression profile but instead clustered with one of the known molecular subtypes. We then performed pairwise differential expression analysis and pathway analysis, separately comparing *p53abn-like NSMP* and p53abn groups against NSMP cases. These results suggested the upregulation of PI3k-Akt, Wnt, and Cadherin signaling pathways both in *p53abn-like NSMP* and p53abn groups (compared to NSMP). Interestingly, while these pathways were up-regulated in both groups, we found little to no overlap between the specific down- and up-regulated genes in the *p53abn-like NSMP* and p53abn groups (compared to NSMP) suggesting that the molecular mechanisms associated with p53abn and p53abn-like tumors might be different even though p53abn and *p53abn-like NSMP* groups had similar histopathological profiles as assessed based on H&E slides.

Discussion

Although many patients with endometrial carcinoma may be cured by surgery alone, about 1 in 5 patients have more aggressive disease and/or have the disease spread beyond the uterus at the time of diagnosis. Identifying these at-risk individuals remains a challenge, with current tools lacking precision. Molecular classification offers an objective and reproducible classification system that has strong prognostic value; improving the ability to discriminate outcomes compared to conventional pathology-based risk stratification criteria. However, it has become apparent that within molecular subtypes and most profoundly within NSMP ECs, there are clinical outcome outliers. The current study addresses this diversity by employing AI-powered histopathology image analysis, in an attempt to identify clinical outcome outliers within the most common molecular subtype of endometrial cancer (Fig. 5). Our results have several clinical and biological implications.

To be clear, for some molecular subtypes, such as *POLE*mut endometrial cancers with almost uniformly favourable outcomes, no further stratification, at least within Stage I-II disease (encompassing >90% of *POLE*mut ECs), is needed. Multiple studies, as well as meta-analyses[33], have shown that in patients with *POLE*mut endometrial cancers, additional pathological or molecular features are not associated with outcomes, i.e., are not prognostic, as *POLE* is the overriding feature that determines survival. However,

for NSMP endometrial cancers, additional stratification tools are greatly needed. Designation of NSMP is the last step in molecular classification, only defined by what molecular features it does ***not*** have; that is without pathogenic *POLE* mutations, without mismatch repair deficiency or p53 abnormalities as assessed by IHC. This leaves a large group of pathologically and molecularly diverse tumors with markedly varied clinical outcomes.

Our AI-based histomorphological image analysis model identified a subset of NSMP endometrial cancers with inferior survival. This subset of patients encompassed approximately 20% of NSMP tumors which are the most common molecular subtype, representing half of endometrial cancers diagnosed in the general population, and thus account for 10% of all ECs. Our results suggest that clinicopathological, IHC, gene expression profiles, or NGS molecular markers (except for copy number burden to some extent) may not be able to identify these p53abn-like outliers. Of note, our results corroborated with a recent report that identified a similar subset of NSMP cases with higher nuclear atypia in 3% of NSMP cases (n = 4 out of 120) with poor outcome, although this difference was not statistically significant likely due to a small sample size and differences in the image analysis models (p = 0·13)[23]. Taken together, AI applied to histomorphological images of routinely generated H&E slides appears to enable a more encompassing and easily implementable stratification of NSMP tumors and provides greater value than any single or combined pathological/molecular profile could achieve.

Molecular characterization of the identified subtype using sWGS suggests that these cases harbor an unstable genome with a higher fraction of altered genome, similar to the p53abn group but with lesser degree of instability. These results suggest that the identified subgroup based on histopathology images is biologically distinct. Furthermore, in spite of the fact that similar gene expression pathways were implicated in both groups and H&E images of both groups as assessed by AI had resemblances, expression data analysis revealed minimal overlap between the differentially expressed genes in both p53abn and p53abn-like EC compared to NSMP cases. This suggests that they may have different etiologies and warrants further biological interrogation of these groups in future studies.

Certainly, others have attempted to refine stratification within early-stage endometrial cancers, including within the molecularly defined NSMP subset. PORTEC4a used a combination of pathologic and molecular features (MMRd, L1CAM overexpression, *POLE*, *CTNNB1* status) to identify low, intermediate, and high risk individuals assigned to favourable, intermediate, and unfavourable risk groups which then determined observations vs. treatment[34]. TAPER/EN.10 also stratifies early-stage NSMP tumors by pathological (e.g., histotype, grade, LVI status) and molecular features (*TP53,* ER status) to identify those

individuals appropriate for de-escalated therapy[35]. In retrospective series, key parameters of ER and grade have been suggested to discern outcomes within NSMP. ER status was also demonstrated to stratify outcomes in patients with NSMP ECs enrolled in clinical trials[36]. However, even in-depth profiling of apparent low risk ECs has failed to find pathogenomic features that would discern individuals who develop recurrence from other apparent indolent tumors[37]. Stasenko et al.[37] assessed a series of 486 cases of 'ultra-low risk' endometrial cancers defined as stage 1A with no myoinvasion, no LVI, grade 1 of which 2.9% developed recurrence with no identifiable associated clinical, pathological or molecular features[37]. Current treatment guidelines, even where molecular features are incorporated, offer little in terms of directing management within NSMP endometrial cancers beyond consideration of pathological features, leaving clinicians to struggle with optimal management[12]. A more comprehensive stratification tool within NSMP endometrial cancers would be of tremendous value, and AI discernment from histopathological images as a tool that can be readily applied to H&E slides that are routinely generated as part of the practice is appealing.

Our proposed AI model also identified a subset of p53abn ECs with marginally superior DSS and resemblance to NSMP (*NSMP-like p53abn*) as assessed by H&E staining. Further investigation of the identified groups and deep molecular and omics characterization of this subset of p53abn ECs may in fact aid us in refining this subtype and identifying a subset of p53abn cases with statistically superior outcomes.

This study is the first to consider the application of AI in refining endometrial cancer molecular subtypes. In general, such studies to generate new knowledge using AI in histopathology are extremely sparse as a majority of the effort has focused on recapitulating the existing body of knowledge (e.g., to diagnose cancer, to identify histological subtypes, to identify known molecular subtypes). This study moves beyond the mainstream AI applications within the current context of standard histopathology and molecular classification. This enables us to direct efforts to understand the biological mechanisms of this newly identified subset. This could present an exciting opportunity to utilize the power of AI to inform clinical trials and deep biological interrogation by adding more precision in patient stratification and selection.

AI histopathologic imaging-based application within NSMP enables discernment of outcomes within the largest endometrial cancer molecular subtype. It can be easily added to clinical algorithms after performing hysterectomy, identifying some patients (*p53abn-like NSMP*) as candidates for treatment analogous to what is given in p53abn tumors. Furthermore, the proposed AI model can be easier to implement in practice (for example, in a cloud-based environment), leading to greater impact on patient

management and even more equitable cancer care if confirmed in diagnostic biopsies. If, from diagnostic office biopsy or surgical curettage endometrial cancers could be classified as NSMP tumors and then AI stratification applied, we would have the opportunity to guide therapeutic decision making as well as surgical management, potentially directing individuals at very low risk of metastases to simple hysterectomy in the community and more aggressive *p53abn-like NSMP* to cancer centers for lymph node assessment, omental sampling and directed biopsies given a higher likelihood of upstaging.

MATERIALS AND METHODS

Histopathology slide digitization

Histopathology slide images associated with the TCGA cohort were acquired from the TCGA GDC portal (https://portal.gdc.cancer.gov). Histopathology slides associated with the Vancouver cohort as well as the Tübingen University Women's Hospital were scanned using an Aperio AT2 scanner.

AI tumor-normal classifier and automatic annotation

The downstream tumor subtype classifier relies on the tumor areas of the tissues. Given that the manual annotation of all slides by pathologists is tedious and time-consuming, we first trained a deep learning model to identify the tumor areas of the slides automatically (Supplementary Fig. 6**)**. To train the model, we utilized 27 slides that were annotated by a board-certified pathologist. First, we split the slides into training (51·8%), validation (22·2%), and testing (26%) sets. To identify the tumor regions of WSIs, we divided them into smaller tiles referred to as patches and extracted 5,091 (2,167 tumor, 2,924 stroma) non-overlapping patches. A maximum of 200 patches with the size of 512×512 pixels at 20x objective magnification were extracted from the annotated regions of each slide. As the baseline architecture for our classifier, we exploited ResNet18[38], a simple and effective residual network, with the pretrained ImageNet[39] weights. We trained the model with the learning rate and weight decay of 1e-4 for five epochs using the Adam optimizer[40]. As the amount of tumor and stroma patches were not equal, we used a balanced sampler with a batch size of 150 which meant that in each batch, the model was trained using 75 tumor patches and 75 stroma patches. The resulting classifier achieved 99·76% balanced accuracy on the testing set, indicating the outstanding performance of this tumor/non-tumor model (Supplementary Table 2). The trained model was then applied to detect tumor regions on the rest of the WSIs. To that end, we extracted patches with identical size and magnification to the training phase. To achieve smoother boundaries for the predicted tumor areas we enforced a 60% overlap between neighbouring patches. In addition, to reduce false positives we used a minimum threshold probability of 90% for tumor patches. Finally, for consistency, we applied the trained model on the discovery set, including the cases that were manually annotated by a pathologist.

Deep learning models for tumor subtype classification

Due to the lack of pixel-wise annotations, we employed variability-aware multiple instance learning (VarMIL)[31] that utilizes the multiple instance learning technique in which an image is modeled as an instance containing a bag of unlabelled patches or tiles. Algorithm 1 elaborates on the prediction mechanism of VarMIL in detail. VarMIL consists of three sections: a feature extractor network ($\mathcal{F}_{\mathcal{C}onv2\mathcal{D}}$), attention layers, and classification layers ($\mathcal{F}_{\mathcal{FC}}$). First, the feature extractor network computes feature embeddings ($z_j \in \mathbb{R}^d$) for the extracted patches of an instance (i.e., image), where d is the dimension of the embeddings. Second, given that patches of a given image are not necessarily equally important in subtype prediction, an attention mechanism calculates the contribution of each patch ($a_j \in \mathbb{R}$) based on its embedding. Subsequently, VarMIL computes the image's representation ($z \in \mathbb{R}^{2d}$) by taking the weighted variance of patches ($z_\sigma \in \mathbb{R}^d$) into account alongside their weighted average ($\bar{z} \in \mathbb{R}^d$). Finally, the model feeds the derived representation as the input of the classification section to predict the subtype. To avoid over-fitting, we employed a variety of augmentation methods including horizontal and vertical flipping, color jitter, size jitter, random rotation, and Cutout[41]. Furthermore, we utilized early stopping[42] as an additional form of regularization in training, and if the validation loss did not decrease after five epochs, we decreased the learning rate. Furthermore, we stopped the training if the validation loss did not decrease after 10 consecutive epochs. We devised a two-step training procedure for the proposed network, in which the feature extractor network was trained independently from the attention and classification layers. First, we trained the feature extractor, ResNet34[38], using patches as inputs for 30 epochs with the learning rate and weight decay of 1e-4 and 1e-5, respectively. To accomplish this, we assigned the label of its corresponding slide to each patch. After optimizing the network, we employed its convolutional layers as the feature extractor ($d = 512$). For the attention and classification layers, we selected a multilayer perceptron (MLP) with a single hidden layer with 128 nodes (q = 128). We trained these layers with the same number of epochs and weight decay as before but with the learning rate of 1e-5. Models were trained using a single dgxV100 GPU with 32GB RAM. The programming language was PyTorch[43], and we selected the hyperparameters experimentally.

Algorithm 1 subtype prediction

Input: $P_i = \{p_{i1}, .., p_{ik_i}\}$: a set of extracted patches from the ith image, i.e., I_i

for $j \leftarrow 1$ to k_i do
    $z_j \leftarrow \mathcal{F}_{\mathcal{C}onv2\mathcal{D}}(p_{ij})$
end for
Let $W \in \mathbb{R}^q$ and $V \in \mathbb{R}^{q \times d}$ be the attention parameters
for $j \leftarrow 1$ to k_i do

$a_j \leftarrow \exp\left(W^\top \tanh\left(Vz_j\right)\right) / \sum_{t=1}^{K_i} \exp(W^\top \tanh(Vz_t))$

end for

$\bar{z} \leftarrow \sum_{j=1}^{k_i} a_j z_j$

$z_\sigma \leftarrow \frac{k_i}{k_i - 1} \sum_{j=1}^{k_i} a_j \left(z_j - \bar{z}\right)^2$

$z \leftarrow \bar{z} \oplus z_\sigma$

$Y_i \leftarrow \mathcal{F}_{\mathcal{FC}}(z)$

Output: Y_i: the predicted subtype of I_i

We further assessed the robustness of our findings with five other models formulated on distinct concepts: (1) Vanilla[44,45], (2) Histogram-Based[46], (3) Iterative Draw and Rank Sampling (IDaRS)[47], (4) Attention-based[48,49], and (5) Vector of Locally Aggregated Descriptor (VLAD)[50,51].

(1) Vanilla is a simple and frequently used concept in digital pathology[44,52] . In this setting, we train a DL model on the extracted patches from a histopathology slide in a fully supervised manner. Here, each patch's label corresponds to the subtype of its corresponding histopathology slide. The process involves passing patches through convolutional layers and feeding the generated feature maps into fully connected layers. The model is trained using the cross-entropy loss function[53], similar to standard classification tasks.

(2) IDaRS shares similar assumptions with Vanilla, involving training a model on image patches in a fully supervised manner and assigning the image's label to its patches[47]. However, unlike Vanilla, where all extracted patches are used in training, IDaRS employs a selection procedure. Only informative patches that contribute to the image's subtype are included during training. The selection algorithm utilizes the Monte-Carlo[54] sampling approach.

(3) The Histogram-Based concept[46] addresses the task of identifying a slide's subtype, similar to IDaRS and Vanilla, by transforming a weakly supervised problem into a fully supervised one. A key distinction of this concept is the integration of a histogram and a classification module, instead of relying on majority voting. This modification improves the model's interpretability without significantly increasing the parameter count.

(4) DeepMIL[55] combines the concepts of MIL and attention. It leverages MIL techniques, treating an image as a collection(bag) of unlabelled patches, while the attention-based approach maintains the nature of the weakly supervised task, in contrast to the previously mentioned concepts. This perspective removes the need to assign labels to individual patches within an image. Moreover, it recognizes that patches within an image have varying degrees of importance to its subtype, and their contributions are calculated using an attention mechanism.

(5) VLAD, a family of algorithms, considers Histopathology images as Bag of Words (BoWs), where extracted patches serve as the words. Due to its favorable performance in large-scale databases, surpassing other BoWs methods, we adopt VLAD as a technique to construct slide representation[50].

Identification of *p53abn-like NSMP*s

The initial hypothesis was that NSMP cases with a poor prognosis resemble p53abn morphologically. Assuming the hypothesis is correct, subtype classifiers should label cases in this group as p53abn. Using the same rationale, we partitioned the NSMP subtype into two subgroups: *p53abn-like NSMP* and the remaining NSMP cases. To this end, we devised a voting system based on the classifiers' consensus. If the fraction of classifiers predicting an NSMP case as p53abn exceeded a specified confidence threshold, the image was labelled as *p53abn-like NSMP*; otherwise, the image was labelled as NSMP. In this work, we labeled a sample as *p53abn-like NSMP* when an NSMP sample, based on ProMisE, was classified as p53abn in more than seven out of the 10 cross-validation classifiers.

Unsupervised clustering of NSMP patch representations

To investigate the robustness of our results in identifying p53abn-like NSMPs and visualize the distribution of the patch representations, we employed a two-step approach. In the first step, we applied KimiaNet[32] to the patches that were extracted from the histopathology slides associated with the NSMP EC cases. KimiaNet is a deep model trained on a large set of histopathology data, to encode each patch with dimensions of 512´512 pixels into a compact 1024´1 vector. By leveraging the embeddings from KimiaNet's last pooling layer, we condensed the essential features of each patch into a representative vector. In the second step, we applied Uniform Manifold Approximation and Projection (UMAP)[56], a dimensionality reduction technique, to project the encoded vectors of all the patches within the NSMP and p53abn-like NSMP onto a two-dimensional space. UMAP excels at preserving both local and global structures of high-dimensional data, enabling us to visualize the relationships and patterns within the encoded patches in a more interpretable manner.

Targeted point mutation profiling

The exon capture libraries were sequenced using the Illumina Genome Analyzer (GAIIx) as 76bp pair-end reads as described in [57]. The reads were aligned to the human genome using the BWA aligner version v0·5 ·9. and SNVs were called by a combination of binomial exact test and MutationSeq as previously described[58,59] . To remove the germline mutations, the predicted SNVs were filtered through dbSNP, 1000 Genome (http://www.1000genomes.org/) and the control normals. All SNVs were profiled by mutationassessor[60] for functional impact of the missense mutations. snpEff

(http://snpeff.sourceforge.net/) was used to find splice site mutations. All silent mutations were removed. The indels were filtered by the control normals and then profiled by Oncotator (http://www.broadinstitute.org/oncotator/).

Survival analysis

We assessed the significance of subgroups using the Kaplan-Meier (KM) estimator on two survival endpoints: Disease Specific Survival (DSS) and Progression Free Survival (PFS). Survival outcomes were not accessible for four (1·47%; three NSMP and one *p53abn-like NSMP*) and two (1·03%; one NSMP and one *p53abn-like NSMP*) patients in the discovery and validation sets, respectively. In some individuals, clinical data were partially available (for example, survival data of a patient only contained DSS while PFS was unknown), explaining why the number of cases varies among KM curves for the same set. In addition, given that the TCGA survival data lacked DSS, the German cohort served as the discovery set for the DSS KM curves.

Shallow whole genome sequencing (cohort and experiments)

DNA was extracted (GeneRead FFPE DNA kit from Qiagen) from FFPE core tumor samples and was sheared to 200bp using a Covaris S220. Libraries were constructed using the ThruPlex DNA-seq kit (Takara) with seven cycles of amplification (library prep strategy from Brenton Lab similar to the one published in 2018)[61]. Library quality was assessed using the Agilent High Sensitivity DNA kit (Agilent Technologies), and pooled libraries were run on the Illumina NovaSeq at the Michael Smith Genome Sciences Centre targeting 600M reads per pooled batch. The sWGS data was run through basic processing which includes trimming with Trimmomatic[62], alignment with bwa-mem2[63], duplicate removal with Picard[64], and sorting with samtools[65]. Sequencing coverage and quality were evaluated using fastQC[66] and samtools. If acceptable, the data was passed along to the next step of determining genomic copy numbers (QDNAseq[67] + rascal[68]) and signature calling. The signature calling step uses techniques including mixture modelling and non-negative matrix factorization and is composed mostly of software from the CN-Signatures[61] package with a few in-house modifications and additions. Interim data munging and ETL (extract, transform, load) are done primarily in bash and R (tidyverse), while visualization and plotting is performed mostly just in R using ggplot2 and pheatmap.

Gene expression analysis

For expression profiling, we used RNA-seq profiles obtained from the TCGA-UCEC cohort[3]. Specifically, we used the GDC data portal[69] to download primary tumors sequenced on the Illumina Genome Analyzer platform with patient IDs matching those used in our study. Raw, un-normalized counts were used. DeSeq2[70] was used to process the raw count matrix and perform differential expression analysis (DEA)

and hierarchical clustering. Samples were categorized as NSMP, *p53abn-like NSMP*, and p53abn. Genes with total count five or less were removed. Counts were normalized using DeSeq2's variance-stabilizing transform tool. The 500 most variable genes based on DEA were kept for hierarchical clustering. Per-gene Z-scaling was applied to normalize the clustering features. Finally, the complete-linkage method was used for both gene-clustering and sample-clustering. Subsequent pathway analysis on the list of differentially expressed genes was performed using Reactome[71] FI plugin in Cytoscape[72].

Statistical assessment

Log-rank test was utilized to assess the significance of the difference between KM curves for the identified patient groups. In addition, the significance of groups for enrichment of specific genomic or molecular features was assessed using the Fisher's exact test and the Mann-Whitney U rank test for discrete and continuous data, respectively. Throughout all experiments, $p < 0{\cdot}05$ was regarded as the significance level.

References:

1 Gilks CB, Oliva E, Soslow RA. Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma. *The American journal of surgical pathology* 2013; **37**: 874–81.

2 Hoang LN, McConechy MK, Köbel M, *et al.* Histotype-genotype correlation in 36 high-grade endometrial carcinomas. *The American journal of surgical pathology* 2013; **37**: 1421–32.

3 Levine DA, Getz G, Gabriel SB, *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* 2013; **497**: 67–73.

4 Talhouk A, McConechy MK, Leung S, *et al.* Confirmation of ProMisE: A simple, genomics-based clinical classifier for endometrial cancer. *Cancer* 2017; **123**: 802–13.

5 Kommoss S, McConechy MK, Kommoss F, *et al.* Final validation of the ProMisE molecular classifier for endometrial carcinoma in a large population-based case series. *Annals of Oncology* 2018; **29**: 1180–8.

6 Editorial Board WC of T. WHO Classification of Tumours Female Genital Tumours, 5th edn. International Agency for Research on Cancer, 2020.

7 Kasius JC, Pijnenborg J, Lindemann K, *et al.* Risk Stratification of Endometrial Cancer Patients: FIGO Stage, Biomarkers and Molecular Classification. *Cancers* 2021; **13**: 5848.

8 Thompson E, Huvila J, Chiu D, *et al.* Further stratification of no specific molecular profile (NSMP/P53WT) endometrial carcinomas to refine prognosis and identify therapeutic opportunities. *International Journal of Gynecologic Cancer* 2021; **31**: A17–A17.

9 Leo AD, Biase D de, Lenzi J, *et al.* ARID1A and CTNNB1/β-Catenin Molecular Status Affects the Clinicopathologic Features and Prognosis of Endometrial Carcinoma: Implications for an Improved Surrogate Molecular Classification. *Cancers* 2021; **13**: 950.

10 Kolehmainen A, Pasanen A, Tuomi T, Koivisto-Korander R, Bützow R, Loukovaara M. Clinical factors as prognostic variables among molecular subgroups of endometrial cancer. *PLoS One* 2020; **15**: e0242733.

11 Prakasan AM, James FV, Jagathnath Krishna KM, *et al.* The Pattern of Recurrence in Carcinoma Endometrium. *Indian Journal of Gynecologic Oncology* 2022; **20**: 1–7.

12 National Comprehensive Cancer Network (NCCN). Uterine Neoplasms NCCN Guidelines Version 4.2021. 2021. https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1473.

13 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–44.

14 Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* 2019; **16**: 703–15.

15 Srinidhi CL, Ciga O, Martel AL. Deep neural network models for computational histopathology: A survey. 2019; published online Dec 28. https://arxiv.org/abs/1912.12378v1 (accessed April 20, 2020).

16 Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Cancer* 2022; **3**: 1026–38.

17 Kather JN, Heij LR, Grabsch HI, *et al.* Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Cancer* 2020; **1**: 789–99.

18 Bejnordi BE, Veta M, Diest PJ van, *et al.* Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* 2017; **318**: 2199–210.

19 Bulten W, Kartasalo K, Chen P-HC, *et al.* Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med* 2022; **28**: 154–63.

20 Campanella G, Hanna MG, Geneslaw L, *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019; **25**: 1301–9.

21 Fu Y, Jung AW, Torne RV, *et al.* Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer* 2020; **1**: 800–10.

22 Hong R, Liu W, DeLair D, Razavian N, Fenyö D. Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models. *Cell Rep Med* 2021; **2**: 100400.

23 Fremond S, Andani S, Wolf JB, *et al.* Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: a combined analysis of the PORTEC randomised trials and clinical cohorts. *The Lancet Digital Health* 2023; **5**: e71–82.

24 Wang T, Lu W, Yang F, *et al.* Microsatellite Instability Prediction of Uterine Corpus Endometrial Carcinoma Based on H&E Histology Whole-Slide Imaging. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). 2020: 1289–92.

25 Wetstein SC, de Jong VMT, Stathonikos N, *et al.* Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images. *Sci Rep* 2022; **12**: 15102.

26 Lee Y, Park JH, Oh S, *et al.* Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. *Nat Biomed Eng* 2022; : 1–15.

27 Zadeh Shirazi A, McDonnell MD, Fornaciari E, *et al.* A deep convolutional neural network for segmentation of whole-slide pathology images identifies novel tumour cell-perivascular niche interactions that are associated with poor survival in glioblastoma. *Br J Cancer* 2021; **125**: 337–50.

28 Jiang S, Zanazzi GJ, Hassanpour S. Predicting prognosis and IDH mutation status for patients with lower-grade gliomas using whole slide images. *Sci Rep* 2021; **11**: 16849.

29 Kang EY, Wiebe NJ, Aubrey C, *et al.* Selection of endometrial carcinomas for p53 immunohistochemistry based on nuclear features. *The Journal of Pathology: Clinical Research* 2022; **8**: 19–32.

30 Vahadane A, Peng T, Sethi A, *et al.* Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Transactions on Medical Imaging* 2016; **35**: 1962–71.

31 Schirris Y, Gavves E, Nederlof I, Horlings HM, Teuwen J. DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. *Medical Image Analysis* 2022; **79**: 102464.

32 Riasatian A, Babaie M, Maleki D, *et al.* Fine-Tuning and training of densenet for histopathology image representation using TCGA diagnostic slides. *Medical Image Analysis* 2021; **70**: 102032.

33 McAlpine JN, Chiu DS, Nout RA, *et al.* Evaluation of treatment effects in patients with endometrial cancer and POLE mutations: An individual patient data meta-analysis. *Cancer* 2021; **127**: 2409–22.

34 Wortman B, Bosse T, Nout R, *et al.* Molecular-integrated risk profile to determine adjuvant radiotherapy in endometrial cancer: evaluation of the pilot phase of the PORTEC-4a trial. *Gynecologic oncology* 2018; **151**: 69–75.

35 ClinicalTrials.gov identifier (NCT number): NCT04705649. Tailored Adjuvant Therapy in POLE-mutated and p53-wildtype Early Stage Endometrial Cancer (TAPER). https://clinicaltrials.gov/ct2/show/NCT04705649 (accessed Feb 2, 2022).

36 Stasenko M, Feit N, Lee SS, *et al.* Clinical patterns and genomic profiling of recurrent 'ultra-low risk'endometrial cancer. *International Journal of Gynecologic Cancer* 2020; **30**.

37 Vermij L, Smit V, Nout R, Bosse T. Incorporation of molecular characteristics into endometrial cancer management. *Histopathology* 2020; **76**: 52–63.

38 He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770–8.

39 Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009: 248–55.

40 Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv:14126980 [cs]* 2017; published online Jan 29. http://arxiv.org/abs/1412.6980 (accessed March 19, 2022).

41 DeVries T, Taylor GW. Improved Regularization of Convolutional Neural Networks with Cutout. *arXiv:170804552 [cs]* 2017; published online Nov 29. http://arxiv.org/abs/1708.04552 (accessed March 17, 2022).

42 Caruana R, Lawrence S, Giles L. Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping. In: Proceedings of the 13th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2000: 381–7.

43 Paszke A, Gross S, Massa F, *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems. Curran Associates, Inc., 2019. https://papers.nips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html (accessed March 17, 2022).

44 McConechy MK, Ding J, Cheang MC, *et al.* Use of mutation profiles to refine the classification of endometrial carcinomas. *J Pathol* 2012; **228**: 20–30.

45 Shah SP, Morin RD, Khattra J, *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 2009; **461**: 809–13.

46 Ding J, Bashashati A, Roth A, *et al.* Feature-based classifiers for somatic mutation detection in tumour–normal paired sequencing data. *Bioinformatics* 2012; **28**: 167–75.

47 Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 2011; **39**: e118.

48 Macintyre G, Goranova TE, De Silva D, *et al.* Copy number signatures and mutational processes in ovarian carcinoma. *Nat Genet* 2018; **50**: 1262–70.

49 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–20.

50 Vasimuddin Md, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). 2019: 314–24.

51 Picard toolkit. *Broad Institute, GitHub Repository* 2019. https://broadinstitute.github.io/picard/.

52 Danecek P, Bonfield JK, Liddle J, *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 2021; **10**: giab008.

53 Andrews S. FastQC: a quality control tool for high throughput sequence data. *abraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom*.

54 Scheinin I, Sie D, Bengtsson H, *et al.* DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res* 2014; **24**: 2022–32.

55 Sauer CM, Eldridge MD, Vias M, *et al.* Absolute copy number fitting from shallow whole genome sequencing data. 2021; : 2021.07.19.452658.

56 Grossman RL, Heath AP, Ferretti V, *et al.* Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* 2016; **375**: 1109–12.

57 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 2014; **15**: 550.

58 Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol* 2010; **11**: R53.

59 Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**: 2498–504.

60 Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A. HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. *JCO Clinical Cancer Informatics* 2019; : 1–7.

Data sharing. Histopathology data associated with the TCGA cohort can be acquired from TCGA GDC portal (https://portal.gdc.cancer.gov). Upon the publication of this manuscript, histopathology slides associated with the validation cohort from Vancouver will be released, and code used in this manuscript will be made available on GitHub.

Acknowledgements. This work was supported by Terry Fox Research Institute, Canadian Institute of Health Research, Natural Sciences and Engineering Research Council of Canada, Michael Smith Foundation for Health Research, OVCARE Carraresi, and VGH UBC Hospital Foundation. The funders had no involvement in study conception, data collection, data analysis, data interpretation, writing of the report, or publication decision.

Author contributions. A.D. and H.F. were the research project leaders and led and designed all data analysis. A.D. implemented all the deep learning pipelines. M.W. and A.B. performed the gene expression analysis. M.A. and A.K. performed image analysis. D.C. and A.J. contributed to clinical review and case selection for molecular profiling. D.F. performed the histopathology slide annotations. P.Ah. contributed to software infrastructure for slide annotation. M.D. and H.F. performed bioinformatics analysis. P.Ab and S.JM.J. provided advice on machine learning analysis and provided computational resources. A.T. and S.L. contributed to clinical informatics and biobanking. C.B.G. and N.S. reviewed all specimens for histological and molecular pathology and contributed to manuscript writing. S.K. were responsible for the specimen and clinical data from Tübingen University. J.N.M. and C.B.G. contributed to cohort construction, tumor banking, and initial draft of the manuscript. A.D., H.F., and A.B. wrote the first draft of the manuscript. D.G.H., N.S., J.N.M. conceived the project, provided oversight, edited the manuscript, and are co-senior authors of the manuscript. A.B. conceived and oversaw the project and is the senior corresponding author. All authors have reviewed and approved the manuscript content.

Declaration of interests. Authors declare no competing interests.

Ethics statement. The Declaration of Helsinki and the International Ethical Guidelines for Biomedical Research Involving Human Subjects were strictly adhered throughout the course of this study. All study protocols have been approved by the University of British Columbia/BC Cancer Research Ethics Board.

List of Figures:

Fig. 1: Workflow of the AI-based histopathology image analysis. First, the quality control framework, HistoQC[73], generates a mask that comprises tissue regions exclusively and removes artifacts. Then, an AI model to identify tumor regions within histopathology slides is trained. Next, images are tessellated into small patches and normalized to remove color variations. The normalized patches are fed to a deep learning model to derive patch-level representations. Finally, a model based on multiple instance learning (VarMIL) was utilized to predict the patient subtype.

Fig. 2: Performance statistics and Kaplan Meier (KM) survival curves for AI-identified EC subtypes. **(A)** AUROC and precision-recall plots of average of 10 splits for p53abn vs. NSMP classifier for discovery and validation sets, **(B)** KM curves associated with PFS and DSS for the discovery set, **(C)** KM curves associated with PFS and DSS (where available) in the validation set.

Fig. 3: (A) Histopathological features from the slides in the validation cohort utilizing KimiaNet feature representation from the slides in the validation cohort demonstrate that p53abn-like NSMP and the rest of the NSMP cases constitute two separate clusters with no overlap, **(B)** Clinicopathological features and point mutation data for the validation cohort, **(C)** Clinicopathological features and point mutation data for the TCGA cohort.

Fig. 4: Molecular profiling of *p53abn-like NSMP* cases. Boxplots of copy number burden (i.e., fraction genome altered) in NSMP, *p53abn-like NSMP*, and p53abn cases in the **(A)** validation (6 NSMP, *7 p53abn-like NSMP*, 5 p53abn) and **(B)** TCGA (69 NSMP, *21 p53abn-like NSMP*, 56 p53abn) cohorts. **(C)** Gene expression profiles associated with the *p53abn-like NSMP* (n=21), NSMP (n=69), and p53abn (n=56) tumors in the TCGA cohort.

Fig. 5: The refined classification scheme that leverages AI screening as a supplementary stratification mechanism within the NSMP molecular subtype.

List of Tables:

Table 1: Clinicopathologic features of the discovery set.

Table 2: Clinicopathologic features of the validation set.

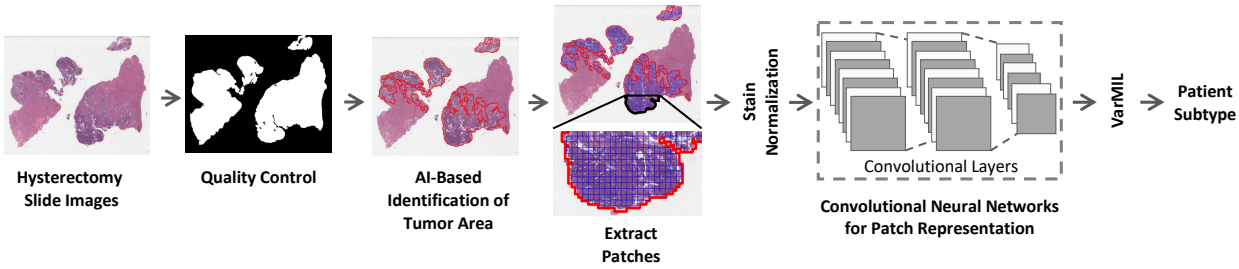

Fig. 1: Workflow of the AI-based histopathology image analysis. First, the quality control framework, HistoQC[73], generates a mask that comprises tissue regions exclusively and removes artifacts. Then, an AI model to identify tumor regions within histopathology slides is trained. Next, images are tessellated into small patches and normalized to remove color variations. The normalized patches are fed to a deep learning model to derive patch-level representations. Finally, a model based on multiple instance learning (VarMIL) was utilized to predict the patient subtype.

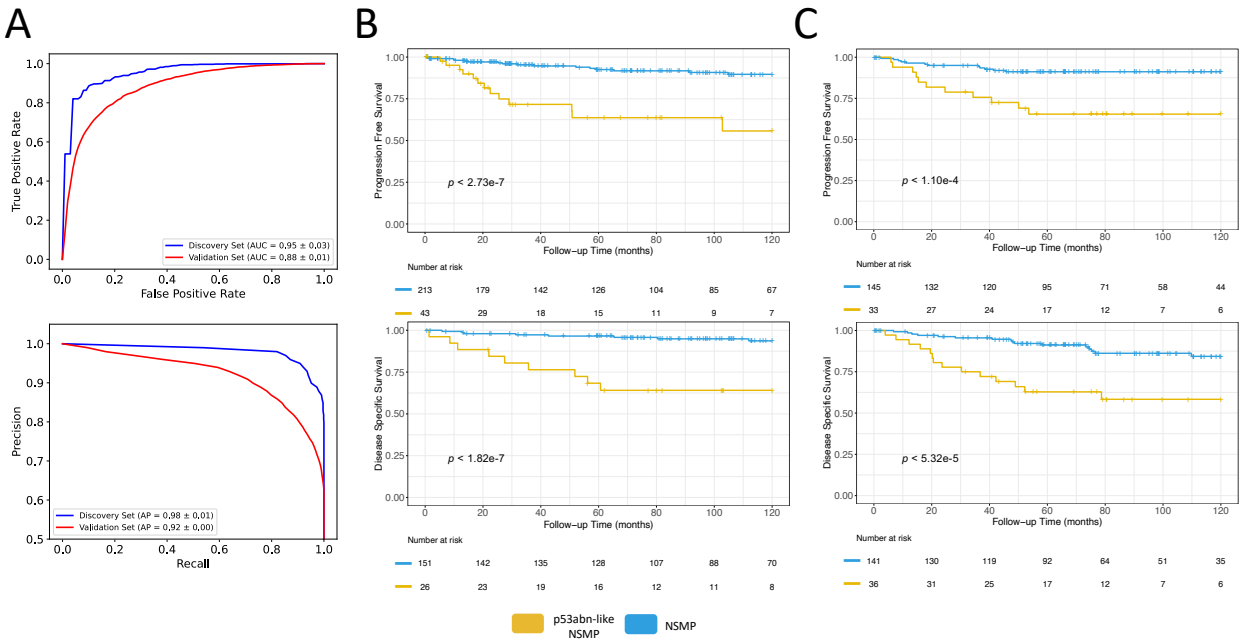

Fig. 2: Performance statistics and Kaplan Meier (KM) survival curves for AI-identified EC subtypes. **(A)** AUROC and precision-recall plots of average of 10 splits for p53abn vs. NSMP classifier for discovery and validation sets, **(B)** KM curves associated with PFS and DSS (where available) for the discovery set, **(C)** KM curves associated with PFS and DSS (where available) in the validation set.

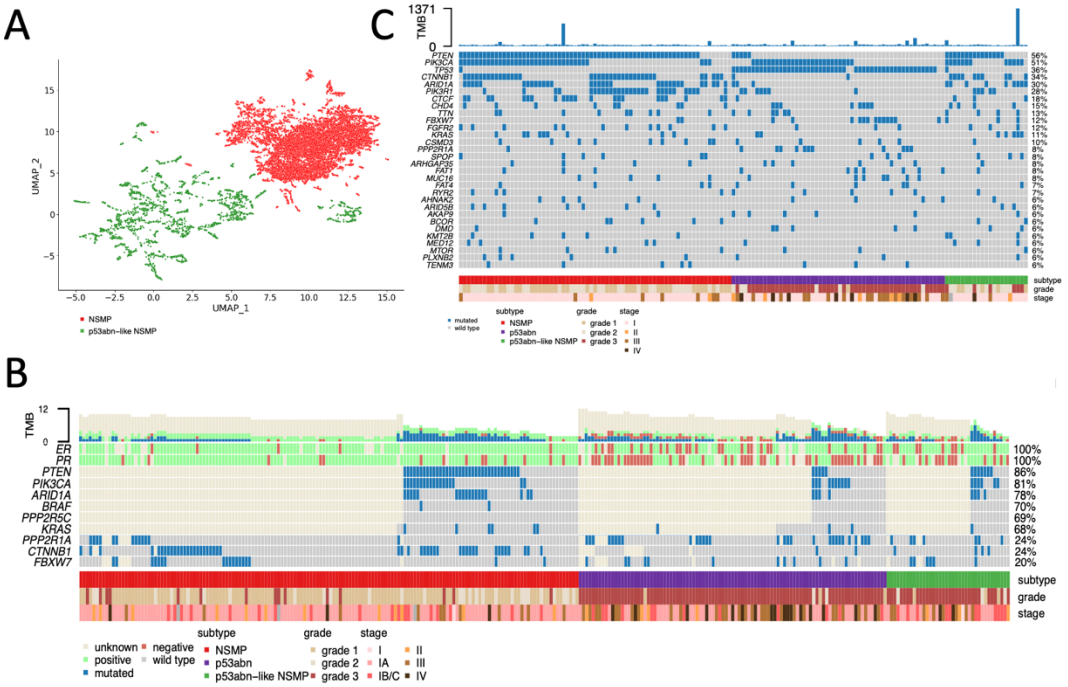

Fig. 3: (A) Histopathological features from the slides in the validation cohort utilizing KimiaNet feature representation from the slides in the validation cohort demonstrate that p53abn-like NSMP and the rest of the NSMP cases constitute two separate clusters with no overlap, **(B)** Clinicopathological features and point mutation data for the validation cohort, **(C)** Clinicopathological features and point mutation data for the TCGA cohort.

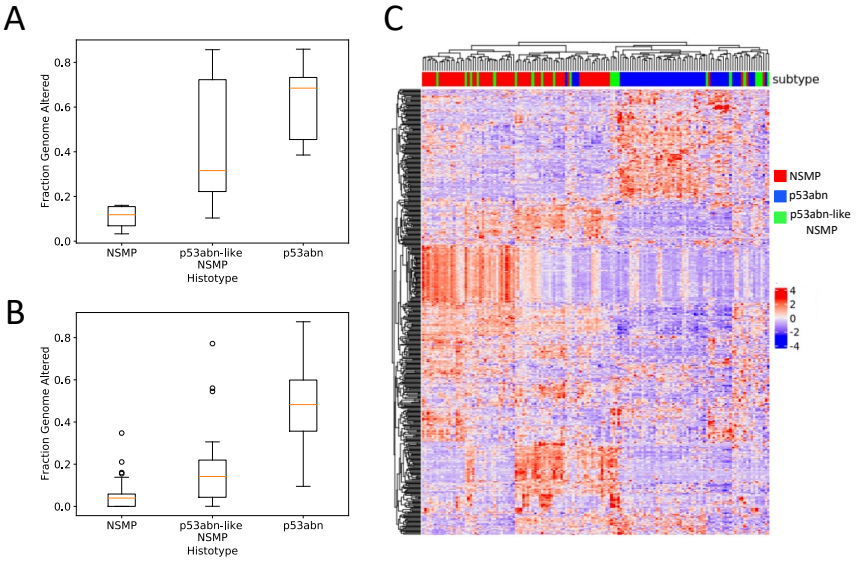

Fig. 4: Molecular profiling of *p53abn-like NSMP* cases. Boxplots of copy number burden (i.e., fraction genome altered) in NSMP, *p53abn-like NSMP*, and p53abn cases in the **(A)** validation (6 NSMP, *7 p53abn-like NSMP*, 5 p53abn) and **(B)** TCGA (69 NSMP, *21 p53abn-like NSMP*, 56 p53abn) cohorts. **(C)** Gene

expression profiles associated with the *p53abn-like NSMP* (n=21), NSMP (n=69), and p53abn (n=56) tumors in the TCGA cohort.

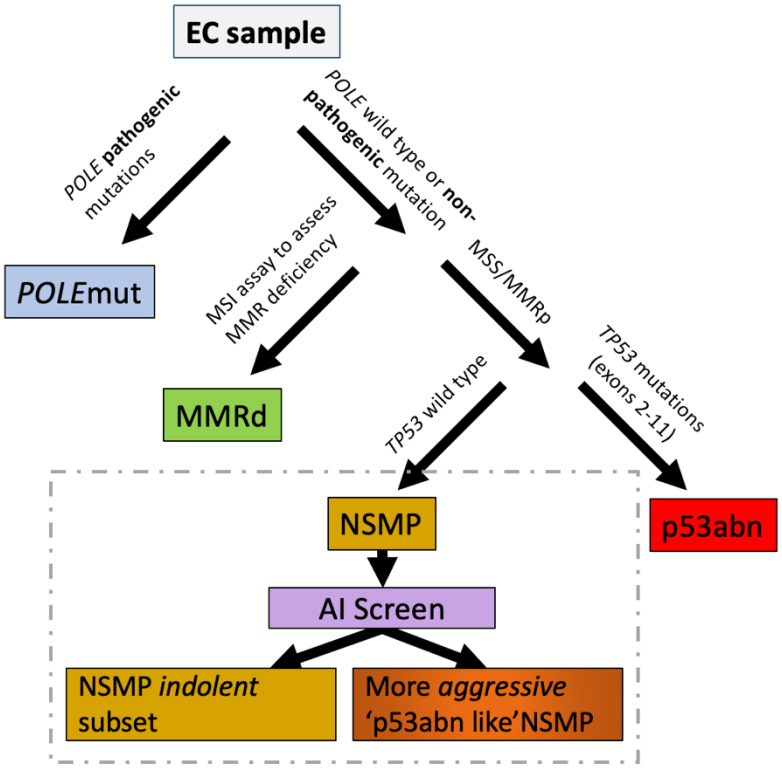

Fig. 5: The refined classification scheme that leverages AI screening as a supplementary stratification mechanism within the NSMP molecular subtype.

Table 1: Clinicopathologic features of the discovery set.

Variable	Total	NSMP	p53abn
Total	363	268 (73·83%)	95 (26·17%)
Age at diagnosis			
<60 yrs	121 (33·33%)	110 (41·04%)	11 (11·58%)
≥60 yrs	242 (66·67%)	158 (58·96%)	84 (88·42%)
Histotype			
Endometrioid	288 (79·34%)	262 (97·76%)	26 (27·37%)
Non-endometrioid	75 (20·66%)	6 (2·24%)	69 (72·63%)
Tumor grade			
Low grade (G1–2)	258 (71·07%)	246 (91·79%)	12 (12·63%)
High grade (G3)	105 (28·93%)	22 (8·21%)	83 (87·37%)
FIGO stage			
I-II	291 (80·17%)	239 (89·18%)	52 (54·74%)
III-IV	71 (19·56%)	28 (10·45%)	43 (45·26%)
Unknown	1 (0·28%)	1 (0·37%)	0

Table 2: Clinicopathologic features of the validation set.

Variable	Total	NSMP	p53abn
Total	288	193 (67·01%)	95 (32·99%)
Age at diagnosis			
<60 yrs	81 (28·13%)	72 (37·70%)	9 (9·47%)
≥60 yrs	205 (71·18%)	119 (62·30%)	86 (90·53%)
Unknown	2 (0·69%)	2 (1·04%)	0
Histotype			
Endometrioid	195 (67·71%)	172 (89·12%)	23 (24·21%)
Non-endometrioid	91 (31·60%)	19 (9·84%)	72 (75·79%)
Unknown	2 (0·69%)	2 (1·04%)	0
Tumor grade			
Low grade (G1–2)	151 (52·43%)	146 (75·65%)	5 (5·26%)
High grade (G3)	137 (47·57%)	47 (24·35%)	90 (94·74%)
FIGO stage			
I-II	216 (75·00%)	166 (86·01%)	50 (52·63%)
III-IV	69 (23·96%)	24 (12·44%)	45 (47·37%)
Unknown	3 (1·04%)	3 (1·55%)	0