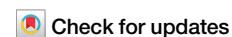


<https://doi.org/10.1038/s41698-024-00592-z>

Deep learning for dual detection of microsatellite instability and *POLE* mutations in colorectal cancer histopathology



Marco Gustav¹, Nic Gabriel Reitsam², Zunamys I. Carrero¹, Chiara M. L. Loeffler^{1,3}, Marko van Treeck¹, Tanwei Yuan⁴, Nicholas P. West⁵, Philip Quirke⁵, Titus J. Brinker⁶, Hermann Brenner^{4,7,8}, Loëtitia Favre^{9,10,11}, Bruno Märkl², Albrecht Stenzinger¹², Alexander Brobeil^{12,13}, Michael Hoffmeister⁴, Julien Calderaro^{9,10,11}, Anaïs Pujals^{9,10,11,15} & Jakob Nikolas Kather^{1,3,5,14,15} ✉

In the spectrum of colorectal tumors, microsatellite-stable (MSS) tumors with DNA polymerase ϵ (*POLE*) mutations exhibit a hypermutated profile, holding the potential to respond to immunotherapy similarly to their microsatellite-unstable (MSI) counterparts. Yet, due to their rarity and the associated testing costs, systematic screening for these mutations is not commonly pursued. Notably, the histopathological phenotype resulting from *POLE* mutations is theorized to resemble that of MSI. This resemblance not only could facilitate their detection by a transformer-based Deep Learning (DL) system trained on MSI pathology slides, but also indicates the possibility for MSS patients with *POLE* mutations to access enhanced treatment options, which might otherwise be overlooked. To harness this potential, we trained a Deep Learning classifier on a large dataset with the ground truth for microsatellite status and subsequently validated its capabilities for MSI and *POLE* detection across three external cohorts. Our model accurately identified MSI status in both the internal and external resection cohorts using pathology images alone. Notably, with a classification threshold of 0.5, over 75% of *POLE* driver mutant patients in the external resection cohorts were flagged as “positive” by a DL system trained on MSI status. In a clinical setting, deploying this DL model as a preliminary screening tool could facilitate the efficient identification of clinically relevant MSI and *POLE* mutations in colorectal tumors, in one go.

In colorectal cancer (CRC), microsatellite instability (MSI) is a key biomarker, indicating tumors that are hypermutated and highly immunogenic^{1–3}. As a result, CRC patients with MSI are considered as candidates for immunotherapy in both early and advanced stages⁴. Laboratory tests, including polymerase chain reaction (PCR) and immunohistochemistry (IHC), are used for the detection of MSI by identifying a lack in expression of mismatch repair deficiency proteins (dMMR). Since 2019, dozens of academic research studies have shown that Deep Learning (DL) can predict MSI status directly from routine hematoxylin and eosin (H&E) histology slides^{5–14}. These studies have led to the regulatory approval of at least one commercial DL-based MSI test in 2022, with several similar products being under development at other commercial entities^{14–16}. Although the performance of DL-based MSI tests may exhibit a somewhat

reduced specificity compared to gold standard methods, they offer a more rapid and cost-effective means of patient pre-screening at a high sensitivity, reducing the need for validating laboratory tests^{14,17}.

Unlike patients with MSI, those with microsatellite stable (MSS) colorectal tumors typically do not respond to immunotherapy. However, there is an exception: those harboring pathogenic mutations in the DNA synthesis proteins DNA Polymerase Epsilon (*POLE*) or DNA Polymerase Delta (*POLD1*). These mutations, specifically in the exonuclease domain of the DNA polymerase family B, disrupt the proof-reading function of the DNA polymerase enzyme^{18,19}. Consequently, these patients accumulate somatic mutations, resulting in a high immunogenicity, and are likely to respond to immunotherapy^{20–22}. Much like CRC with MSI, MSS CRC with *POLE* or *POLD1* mutations also

A full list of affiliations appears at the end of the paper.

✉ e-mail: jakob-nikolas.kather@alumni.dkfz.de

 THE HORMEL INSTITUTE
UNIVERSITY OF MINNESOTA

exhibit a response to cancer immunotherapy^{22,23}, with ongoing clinical trials for immuno-oncology therapy in such cases²⁴. However, these pathogenic mutations are found in only about 1% of CRC patients²⁵ and are not typically included in routine screening protocols. As a result, they often remain undiagnosed in clinical practice.

In this study, we aimed to investigate whether DL-based MSI detection methods, which were not explicitly trained to detect *POLE/POLD1* tumors, are capable of identifying such cases. Our hypothesis is grounded on research showing that tumors characterized by MSI, and those with pathogenic *POLE/POLD1* mutations (but MSS), exhibit shared clinical and morphological characteristics^{20,26–29}. Although these tumors are not biologically identical³⁰, they share biological features that arise from deficiencies in DNA repair mechanisms, leading to subsequent genomic instability and response to immunotherapy^{31–33}. Resulting from these similarities, we hypothesized that the morphological appearance of these tumors in histological H&E images should be comparable enough for DL methods to detect them, even if trained solely on detecting MSI. This would be clinically valuable, as there is currently a lack of rapid, cost-effective and widely accessible methods to diagnose CRC cases with pathogenic *POLE/POLD1* mutations.

Results

Prediction of MSI status from histological data via Deep Learning across three external cohorts

Digital whole slide images of H&E stained histopathology tissue of CRC cases were collected from four patient cohorts (Fig. 1a and Supplementary Table 1): DACHS (Darmkrebs: Chancen der Verhütung durch Screening, $N = 2039$)^{34,35}, TCGA (The Cancer Genome Atlas, $N = 429$, Fig. 1c), APHP (Assistance Publique–Hôpitaux de Paris/Public Assistance Hospitals of Paris) surgical resection ($N = 27$, Fig. 1d) and APHP biopsy ($N = 38$, Fig. 1d). The cases from the DACHS cohort were then used to train a transformer-based Deep Learning model (Fig. 1a, b) for the prediction of MSI status. We assessed the model's performance through patient-level cross-validation. Our findings revealed a state-of-the-art Area under the Receiver Operating Curve (AUROC) of 0.94 ± 0.03 ($p = 0.00$, Supplementary Fig. 2a) for MSI prediction, reiterating that MSI status can be reliably predicted from histopathology (Supplementary Fig. 2b). We used a pre-defined threshold of 0.5 to binarize outcomes based on the prediction scores obtained from the output neurons of the classifier network, with cases above and equal to 0.5 classified as positive, and those below as negative. This resulted in a sensitivity of 0.87 ± 0.07 (mean $\pm$ standard deviation), a specificity of 0.88 ± 0.04 , a negative predictive value (NPV) of 0.98 ± 0.01 and a positive predictive value (PPV) of 0.47 ± 0.08 . The distribution of prediction scores within the DACHS cohort indicates accurate classification for most MSI and MSS cases, including those patients under 50 years of age (Supplementary Fig. 2). These observations highlight the model's robust performance on internal data, underscoring the need for external validation. To address this, we tested the classifier's generalizability by utilizing three external cohorts (Fig. 2a–c) with previously unseen data. For the TCGA cohort, consisting of $N = 429$ patients, an AUROC of 0.87 ± 0.02 ($p = 0.00$, Supplementary Fig. 2a) was achieved. Furthermore, our model reached a sensitivity of 0.93 ± 0.02 , a specificity of 0.51 ± 0.10 , a NPV of 0.98 ± 0.00 and a PPV of 0.25 ± 0.03 in the TCGA cohort. The mean prediction scores ($\pm$ standard deviation) in this cohort were 0.83 ± 0.06 for MSI patients, 0.48 ± 0.14 for MSS patients, and 0.79 ± 0.12 for those patients harboring pathogenic *POLE* or *POLD1* mutations (Fig. 2a, d).

The APHP resection and biopsy cohorts were designed to include a nearly equal number of patients with MSI and MSS, ensuring a balanced representation between the two groups. Within the APHP resection cohort, 5.60 ± 0.49 (mean $\pm$ standard deviation) of 6 MSI patients were accurately predicted to be MSI. Conversely, 4.60 ± 0.80 out of 10 MSS patients were correctly categorized, with the remaining ones falsely predicted to be MSI. This resulted in a sensitivity of 0.93 ± 0.08 , specificity of 0.46 ± 0.08 , PPV of 0.51 ± 0.03 , and NPV of 0.93 ± 0.08 (Fig. 2b),

which is similar to commercially available methods for detection of MSI from H&E slides¹⁴. The mean MSI prediction score was 0.90 ± 0.07 for MSI patients, 0.55 ± 0.17 for MSS patients, and 0.70 ± 0.11 for patients with pathogenic *POLE* mutations (Fig. 2d). This implies that our model predicts MSI status on par with a commercially available assay and is a viable pre-screening tool.

For the APHP biopsy cohort, 9.20 ± 2.40 of the 13 MSI patients were accurately predicted to be MSI, while 5.80 ± 1.60 out of 9 MSS patients were incorrectly predicted to be MSI. This led to a sensitivity of 0.71 ± 0.18 , specificity of 0.36 ± 0.18 , PPV of 0.61 ± 0.05 , and NPV of 0.47 ± 0.12 (Fig. 2c). Of note, we found that two cases from the last group which were falsely predicted as MSI, originated from liver metastases (Fig. 2c, middle) as opposed to primary tumors like all cases in the training set, indicating that the system's performance on liver biopsy tissue could be limited. The mean MSI prediction scores for this cohort were: 0.68 ± 0.13 for MSI patients, 0.62 ± 0.13 for MSS patients, and 0.66 ± 0.14 for patients bearing pathogenic *POLE* mutations (Fig. 2d).

To conclude, our transformer-based Deep Learning approach demonstrated robust capabilities in predicting MSI status from pathology slides of surgical resection samples. Next, we assessed the predictions of this classifier in patients harboring *POLE* driver mutations.

Deep Learning for MSI classification also identifies *POLE* driver mutants in three external cohorts

We hypothesized that MSS tumors harboring *POLE* driver mutations share a common phenotype with MSI tumors. To investigate this, we obtained three cohorts with *POLE* sequencing data: TCGA, the APHP surgical resection and the APHP biopsy cohort. In TCGA, $N = 10$ tumors had *POLE* driver mutations (Fig. 1c and Supplementary Table 2), reflecting the natural prevalence of these rare mutations in CRC. The APHP resection cohort was composed of $N = 8$ *POLE* driver mutants, $N = 6$ MSI and $N = 12$ MSS patients, whereas the APHP biopsy cohort was composed of $N = 9$ *POLE* driver mutants, $N = 13$ MSI and $N = 8$ MSS patients (Fig. 1d, Supplementary Table 3). None of the three external cohorts included *POLD1* driver mutation cases.

We observed that 9.00 ± 0.63 out of 10 *POLE* driver mutant patients in the TCGA cohort were predicted as MSI. The two *POLE* driver mutant cases which were not classified as MSI by the median model (Fig. 2a, bottom) carried the *POLE* mutations S459F and V411L, which are considered pathogenic^{29,36–39}. Specifically, TCGA-AG-A002 had mutations S459F (d) and R150* and a median prediction score for MSI of 0.12, while TCGA-AA-A00N carried V411L (d) and L1255V and had a median MSI prediction score of 0.31 (Supplementary Table 2). Overall, there was a discernible trend towards high MSI prediction scores for the majority of *POLE/POLD1* mutant samples. For scores in the range of 0.8 and above, there was reduced variation among the model's predictions, as opposed to the transitional range (Fig. 2a bottom). Furthermore, it can be noted that for the two misclassified cases in TCGA based on the model with the median AUROC, the majority of the 5 deployed models predicted them to be as MSI when the 0.5 threshold was applied.

In the APHP resection cohort, 6.00 ± 0.63 out of 8 *POLE* driver mutant patients were predicted to be MSI (Fig. 2b, bottom). We further evaluated the performance of our model on biopsy samples, which are known to be inherently difficult to assess by DL-models⁴⁰. For the APHP biopsy cohort, 6.40 ± 0.80 out of 9 *POLE* mutant patients were predicted to be MSI (Fig. 2c, bottom). Just as in the MSI cases, we observed that there is less uncertainty regarding *POLE* mutations in the APHP resection cohort compared to the APHP biopsy cohort (Fig. 2b, c, bottom). These findings confirm that the analysis of biopsies using DL is more challenging compared to that of surgical resection slides due to inconsistent sample size, quality, and tumor content.

Understanding the pathogenicity and biological relevance of *POLE* mutations is still an evolving field^{41,42}. Hotspot mutations show a restriction depending on the tumor type⁴³, with *POLE* p.P286R mutations being of

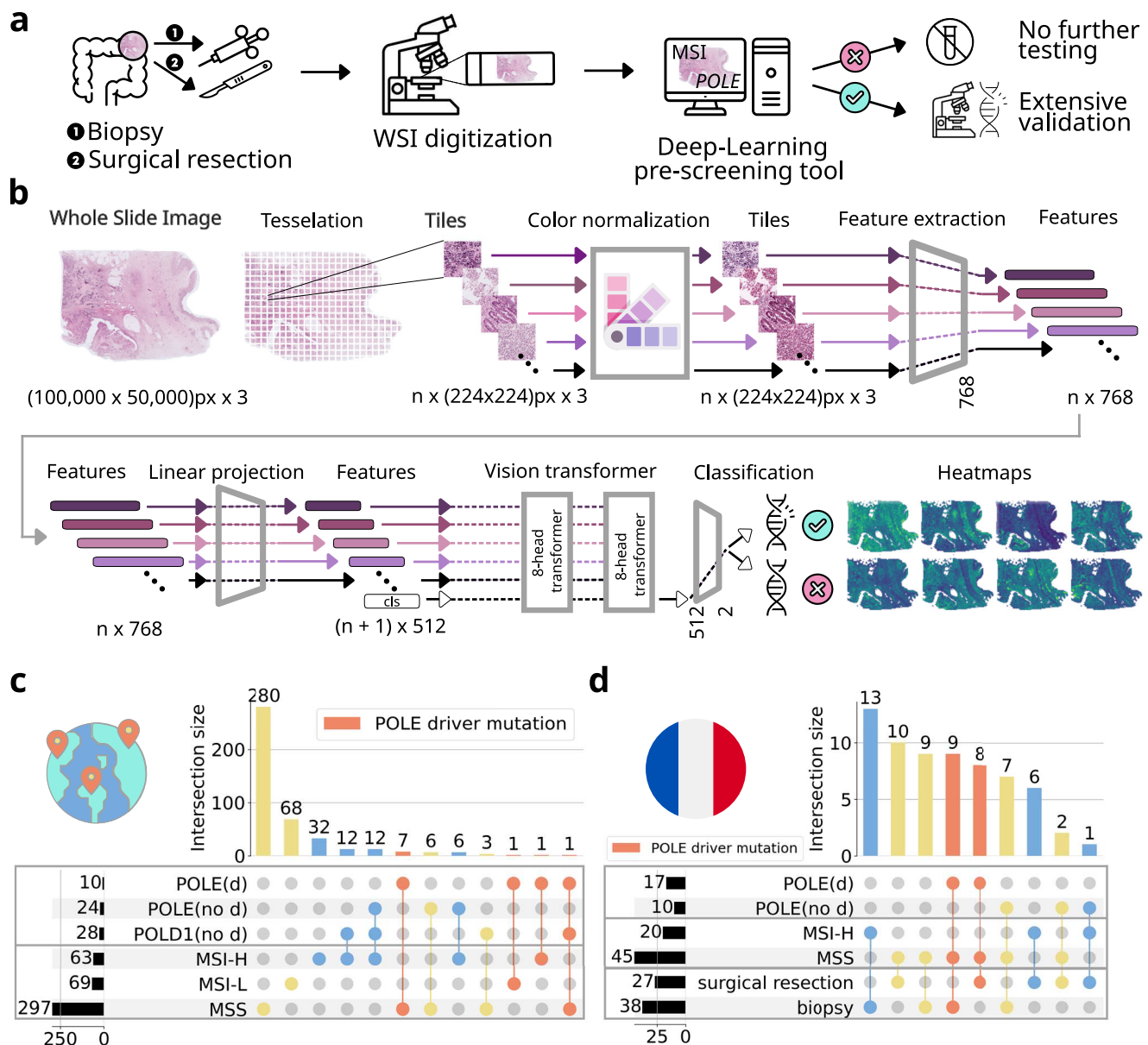


Fig. 1 | Experimental design, study overview and cohort characterization. **a** A suspected case of colorectal cancer (CRC) prompts a biopsy or surgical resection to obtain a tissue sample. This sample is then digitized into a Whole Slide Image (WSI) for analysis by a clinical Deep Learning system, which has the potential to pre-screen for MSI and *POLE* cases, pending external validation and regulatory approval as a medical device. **b** Our Deep Learning pipeline starts with tessellation of Whole Slide Images (WSIs) into smaller, relevant tiles while discarding non-informative background areas. We then extract n feature vectors from n color-normalized tiles, which range in size from 100,000 to 50,000 pixels across three color channels. These vectors are compressed into a more compact feature space and processed using a two-layer, eight-head Vision Transformer (ViT) architecture. Within this system, a 'class

token' is simultaneously trained to generate the final MSI prediction. To aid pathological evaluation, we create heatmaps that visualize the areas of focus determined by the ViT's attention mechanisms. **c** Molecular characterization of the TCGA (The Cancer Genome Atlas) cohort with respect to MSI and MSS (MSI-L/MSS). Combinations of microsatellite status and *POLE*/*POLD1* ("d": driver) mutations are shown (orange: *POLE* driver mutation, blue: MSI-H, yellow: MSI-L/MSS). **d** Molecular characteristics of the APHP (Assistance Publique-Hôpitaux de Paris, resection and biopsy) cohorts with respect to MSI and MSS cases. Combinations of microsatellite status and *POLE*/*POLD1* ("d": driver) mutations are shown (orange: *POLE* driver mutation, blue: MSI-H, yellow: MSS). The icons on all panels are obtained from www.flaticon.com.

particular relevance in CRC^{43,44}. For predicting *POLE* p.P286R mutations, our model reached a comparable or even better prediction scores (mean $\pm$ standard deviation) than calculated for all driver mutations in TCGA (0.96 ± 0.04), APHP resection (0.70 ± 0.29) and APHP biopsy (0.73 ± 0.16), indicating that our algorithm is able to identify morphological features associated with this particular CRC-specific *POLE* mutation. Taken together, our findings demonstrate that DL-based MSI detection methods possess the capacity to identify *POLE* mutants. This provides compelling evidence for a common morphological phenotype between MSI and *POLE* mutations in CRC, thereby confirming our initial hypothesis.

Explainability of the morphological patterns associated with MSI and *POLE* mutations

Having established that a morphology-based classifier trained on MSI tumors can also identify *POLE* mutant tumors, we sought to determine the distinct morphological patterns associated with these tumors. In order to visualize the distribution of attention at a slide-level, we conducted an analysis using attention heatmaps for samples in TCGA (Fig. 3) and DACHS (Supplementary Fig. 3a). Specifically, we examined samples with the highest and lowest prediction scores across all ground truth classes, including MSI, MSS, and *POLE* driver mutations (*POLE* only for TCGA,

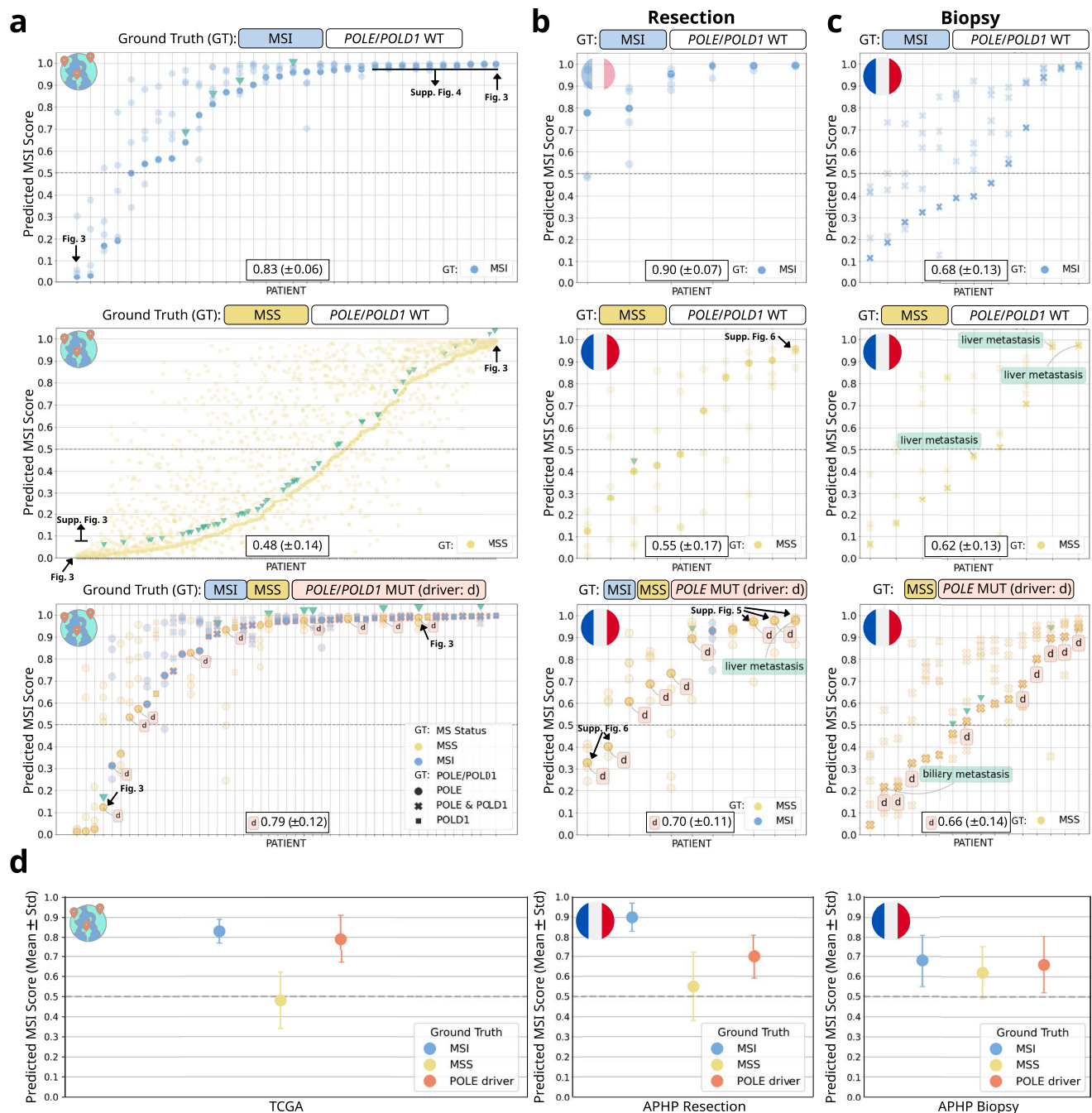


Fig. 2 | Results of MSI and *POLE* prediction experiments with Vision Transformer based pipeline. Results are shown for external validation on TCGA (The Cancer Genome Atlas, International) and AHPH Resection and Biopsy (Assistance Publique–Hôpitaux de Paris, France). Each patient is shown with 5 dots representing their prediction scores from 5-fold cross-validation. Highlighted dots correspond to the median AUROC model based on the TCGA cohort. The classification threshold is set to 0.5 for all cohorts. Mean and standard deviation of the predicted MSI score are shown, with values computed for driver mutations only in case of *POLE/POLD1*. Microsatellite status is indicated by color, wild type by WT, mutated by MUT, driver mutations by “d”, liver metastasis samples by green boxes and early-onset colorectal cancer patients (age at diagnosis <50 years)⁵⁸ by a green triangle. Arrows point to the corresponding heatmaps of selected samples. **a** In the TCGA testing cohort,

prediction scores were calculated for patients with MSI (excluding *POLE/POLD1* mutations, top chart), MSS (excluding *POLE/POLD1* mutations, middle chart) and *POLE/POLD1* mutations (bottom chart). MSS group includes MSI-L and MSS.

b, c In AHPH resection (**b**) and biopsy (**c**) testing cohort, prediction scores were calculated for patients with MSI (excluding *POLE/POLD1* mutations, top chart), MSS (excluding *POLE* mutations, middle chart) and *POLE* (bottom chart). **d** The mean and standard deviation for Predicted MSI Scores in panel a–c based split are calculated for the cohort and ground truth regarding microsatellite and *POLE/POLD1* status. For *POLE/POLD1* mutated cases, only driver mutations are considered. The icons indicating the origin of the cohorts are sourced from www.flaticon.com.

regardless of MS status). Here, we observed that high-attention areas included almost exclusively tumor tissue, whereas other tissue types, such as regular tumor-free intestinal mucosa or uninvolved pericolonic/mesorectal adipose tissue, were not highlighted. The heatmaps also reveal that the

model seems to be quite robust against pen markings, since prevalent pen markings are not highlighted in the majority of cases (Supplementary Fig. 3). The results depicted in the figures also indicate that the eight attention maps of the first transformer layer show more diverse attention

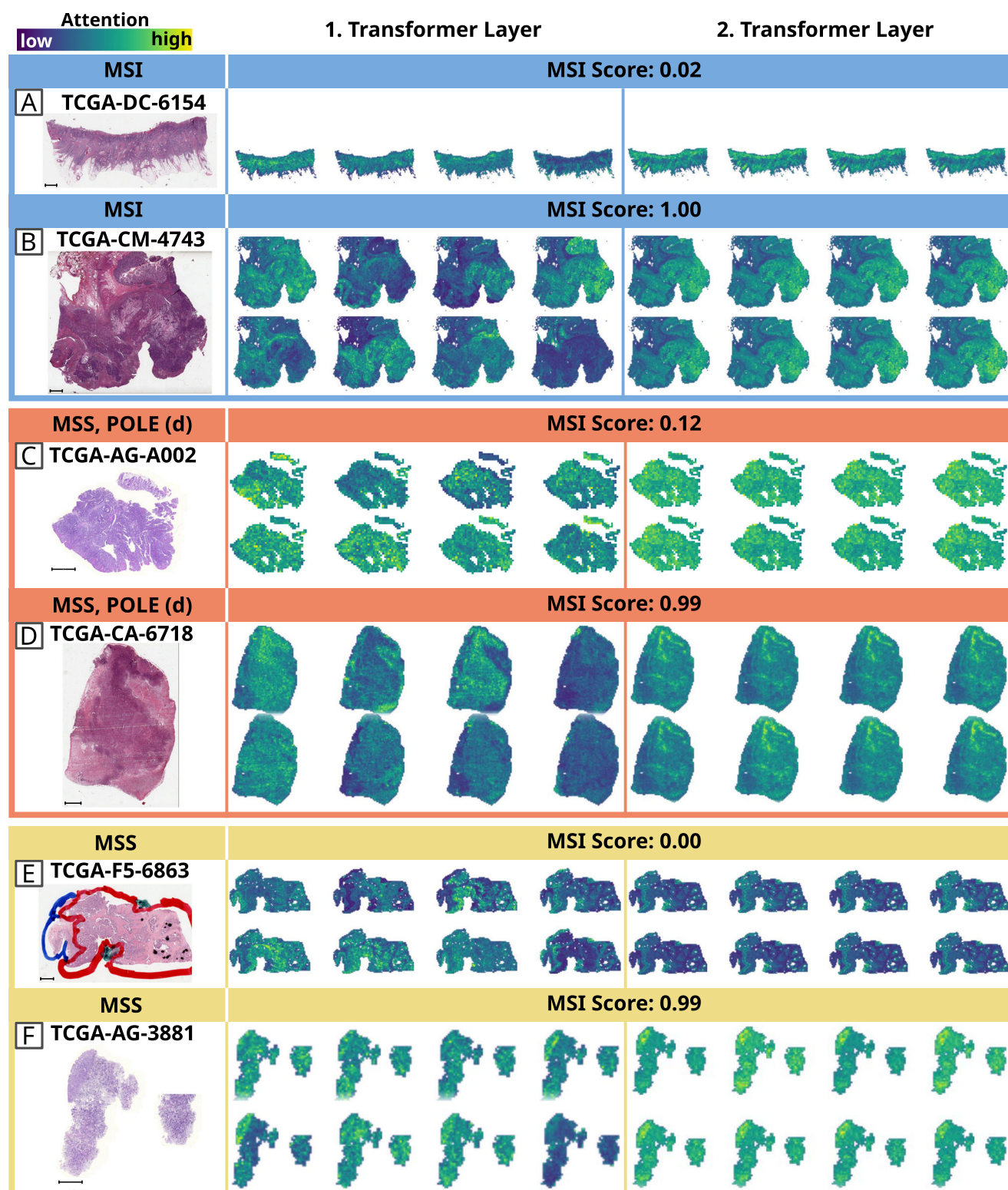


Fig. 3 | Attention heatmaps of selected samples from the TCGA (The Cancer Genome Atlas) cohort. For each group with ground truth MSI (blue, A, B), MSS & *POLE* mutated (orange, C, D), and MSS (yellow, E, F) we chose a sample with a high and a low MSI prediction score, respectively. The associated whole slide images can be viewed under the GDC Data Portal: <https://portal.gdc.cancer.gov/>. Heatmaps were derived from a model that showcases the median AUROC among 5 trained models for this cohort. Due to the transformer's architecture, which consists of 2 layers and 8 heads, each slide yields 16 distinct heatmaps. Blue regions signify areas

with low attention, while yellow designates high attention areas critical for predictions. The scale bars indicate a length of 2 mm. A, C, E illustrate low MSI-prediction scores, characterized by adenocarcinoma NOS (not otherwise specified) histology, the presence of dirty necrosis, and a reduced number of tumor-infiltrating lymphocytes. B, D, F demonstrate high MSI prediction scores, associated with morphological hallmarks such as abundant extracellular mucin (seen in B and F), a medullary growth pattern (visible in D), and an enhanced presence of tumor-infiltrating lymphocytes.

patterns than in the second transformer layer. For two samples (Supplementary Fig. 3b), attention within the transformer varies significantly among different attention heads: some focus on larger regions, while others concentrate on smaller areas. This variance becomes more uniform in the second transformer layer (Fig. 3, Supplementary Fig. 3). Furthermore, we revised the histomorphology of the cases with lowest ($N = 10$, heatmaps in Supplementary Fig. 4) and highest ($N = 10$, heatmaps in Supplementary Fig. 5) MSI prediction scores within TCGA with regards to histology, tumor-infiltrating lymphocytes (TILs) and the presence of dirty necrosis (Supplementary Table 5). High MSI prediction scores were significantly associated with the presence of extracellular mucin or a medullary growth pattern ($p < 0.001$), higher numbers of TILs ($p = 0.002$) and the absence of dirty necrosis ($p = 0.002$) (Supplementary Table 6), which is in line with the characteristic phenotype Shia et al. have already described in their comprehensive histomorphologic characterization of *POLE* mutant CRCs²⁷. In the APHP resection cohort, prediction and attention heatmaps for selected samples have been delineated in Supplementary Figs. 6–7. Within these figures, a specific MSS patient sample, taken from a liver metastasis, was classified as MSI (refer to Supplementary Fig. 6A). Additionally, two MSS patients with pathogenic mutations in *POLE*, both of whom were classified as MSI, are presented in Supplementary Fig. 6B–C. These findings lend considerable weight to the hypothesis of this study. In contrast, the sole two MSS patients with pathogenic mutations in *POLE*, who were classified as MSS, are shown in Supplementary Fig. 7A–B, accompanied by another MSS case which was incorrectly predicted as MSI (Supplementary Fig. 7C). To gain further insights into morphological features associated with our predictions, we qualitatively reviewed another ten misclassified TCGA MSS cases which were *POLE* wild type (WT) but were assigned high MSI prediction scores by our model. Three of these cases were consistent with the diagnosis of mucinous adenocarcinoma ($\geq 50\%$ extracellular mucin content), and one case showed a partly mucinous differentiation ($\sim 40\%$ extracellular mucin content). The other cases showed gland-forming and cribriform architecture (adenocarcinoma, non otherwise specified, low-grade), but all cases had a relatively high tumor-to-stroma ratio, resembling a solid/medullary growth pattern, and also partially showing high number of TILs (Supplementary Table 7). All of these morphological features are known to be associated with MSI, hence they are plausible confounding factors^{26,27,45}. Thus, the misclassification of the model is reasonable as it reflects the categorization a pathologist would likely make. Combining these results, it is clear that the morphological patterns detected through our method provide strong validation of our study's underlying hypothesis.

Discussion

The advent of cancer immunotherapy has fundamentally changed the treatment landscape in oncology, notably in tumor types such as lung cancer and melanoma where immune checkpoint inhibitors elicit prolonged responses even in metastatic cases^{46,47}. Given the current advance in medical treatments, immunotherapy in neoadjuvant and adjuvant settings is increasingly utilized, promising potential for improved response rates across different cancer entities. Despite being one of the most common types of cancer, CRC is notoriously unresponsive to immunotherapy. Only a small fraction ($<10\%$ in metastatic stages⁴⁸) of CRC cases, specifically MSI tumors, benefit from immunotherapy⁴⁹, leading to its use even in the neoadjuvant setting, and necessitating upfront MSI testing in all CRC patients. Several techniques for MSI detection include IHC, PCR, next-generation sequencing (NGS)-based testing (via computational methods such as MSIsensor or MANTIS⁵⁰) and a DL-based assay that predicts MSI status via the analysis of digital images of H&E stained histopathology slides (MSIntuitTM, Owkin, France¹⁴). Among the majority of non-MSI CRC patients, a rare, immunotherapy-responsive subgroup exists with *POLE* or *POLD1* mutations, affecting fewer than 1% of CRC cases but highly significant for impacted individuals. Identifying these mutations, however, involves expensive genetic testing, not routinely conducted in CRC cases nor widely available.

Our study demonstrates the effectiveness of a DL method in MSI screening, capable of identifying not only MSI but also *POLE* mutant

tumors, classifying them as MSI. This model, initially trained on CRC data to differentiate MSI from MSS cases, utilizes specific characteristics in H&E images for this purpose. The ability of our DL model to detect *POLE* mutations — typically identified via extensive genetic data acquisition and processing — represents a major breakthrough, suggesting our model's broader utility beyond MSI detection alone. Despite the challenge in conclusively determining these features within our current technological framework, tools like heatmaps assist in correlating visible characteristics with established pathological patterns, indicating that our model captures features common to both MSI and mutations recognizable by trained pathologists. Therefore, describing our model solely as an MSI detector would be an oversimplification as it effectively identifies various mutation patterns, highlighting its diagnostic capabilities beyond initial expectations. This evidence points to a common histopathological profile shared by MSI and *POLE* mutations, allowing a DL system trained on MSI alone to reliably detect *POLE* mutations. The overlapping morphological features between MSI and many *POLE* tumors, such as high levels of tumor-infiltrating lymphocytes, a medullary growth pattern, substantial mucin production, and the absence of dirty necrosis^{20,26–29}, are not novel findings. Interestingly, some of these features seem to be site-agnostic and also already known for *POLE* mutant endometrial carcinomas⁵¹. The assessment of attention heatmaps in our study reinforced the foundational morphological attributes critical to our DL approach. Our analysis not only corroborated our initial observations with respect to the MSI classifier but also underscored the model's resilience to artifacts like pen markings.

Our model's categorization of *POLE* cases as MSI, while not technically predicting *POLE*, is still clinically significant. Both MSI and *POLE* mutations, associated with hypermutation⁵², increased immune response, and better outcomes with immunotherapy^{20,22,52}, present valuable targets for treatment. Hence, deploying DL screenings could help in identifying potential *POLE* mutation carriers clinically. In practice, after thorough external validation and the necessary regulatory approvals, the model could help in ruling out *POLE* and MSI mutations in samples classified as MSS. This serves as a similar approach to existing DL-based MSI testing tools used for preliminary screening. Following an MSI-positive result from our model, a cost-effective IHC test could be conducted to detect dMMR, correlating with recommendations for existing DL products¹⁴. The detection of dMMR suggests MSI presence. Conversely, an MSI prediction with proficient MMR necessitates further analysis. According to our findings, in CRC cases of proficient MMR, *POLE* mutations might be the underlying cause for an MSI-like prediction. In such scenarios, it is advised to use cost-effective Sanger sequencing for *POLE* mutations or small NGS panels covering *POLE/POLD1* mutations. This is particularly recommended for young male patients with right-sided tumors, as these clinical-pathological features are often associated with *POLE* mutations²⁰. Adopting this tiered testing strategy could significantly reduce the need for broad, expensive panel sequencing for *POLE* mutations, thereby streamlining and providing more cost-effective clinical workflows. This becomes increasingly pertinent with the anticipated incorporation of Immuno-oncology (IO) therapies for *POLE* mutant CRCs into routine clinical management.

Technically, our study enhances a robust, previously established DL pipeline⁵³ with some innovative elements. For final predictions, we employed an ensemble of 5 models, each trained on distinct 80% subsets of the DACHS training cohort. This ensemble approach generates a consensus indication among the models, which can be interpreted as a measure of confidence in the predictions. Our findings indicate that the models are more confident in their predictions when the scores are markedly high or low, as opposed to those around the mid-range (0.5 ± 0.3). This trend is especially evident in MSI or *POLE* mutation cases derived from surgical resections in the TCGA and APHP datasets. Conversely, there is more uncertainty in predictions for biopsies and MSS cases. Further investigation into the learning mechanics of the model and understanding the elements influencing these extreme prediction values and the origins of uncertainty could lead to further refinement of the model. Our future research should focus on leveraging this methodology to quantify the clinical relevance of

ensemble-based prediction uncertainty, ideally incorporating additional cohorts with *POLE* mutant samples. We suggest that data on therapeutic outcomes should be collected in future patient cohorts.

Our research is subject to some constraints, predominantly the limited number of *POLE* mutants and significant class imbalance impacting statistical analysis. Notably, many DL studies in CRC use the TCGA cohort, but it includes only ten relevant *POLE* driver mutants and lacks *POLD1* driver mutants. Despite the inaccessibility of *POLD1* mutant samples, our study sought to solidify our results' robustness and general applicability by incorporating a considerable number of *POLE* mutation cases from a major French reference center (APHP). It is pertinent to note that the prevalence of *POLE* testing in CRC remains limited, impacting the sample size of *POLE* mutations in our study, although their significance is only just being recognized. Our dataset included 28 *POLD1* mutants and 61 *POLE* mutants, 27 of which were *POLE* driver mutations (Fig. 1c, d). The APHP cohort, which comprised two sub-cohorts of surgical resection specimens and biopsies, respectively, confirmed the generalizability of our findings. Consistent with prior research⁸, our results revealed lower performance on biopsy samples, suggesting current DL approaches are more effective with surgical specimen analysis. However, as biopsy-based testing is becoming more common, it is important to develop algorithms that can effectively process these samples in the future⁴. Additionally, the total sample number in the APHP cohort complicates the interpretation of MSI and MSS, with the liver metastasis cases being the most challenging. Results from DL models, trained on primary tumors but applied to metastatic tissues, often demonstrate unsatisfactory discriminative performance⁵⁴. This becomes apparent in the distinct scatter of the model predictions for MSS cases. To address these challenges, further studies could explore how the selection criteria for defining final predictions impact outcomes. For example, instead of basing the final prediction on the median fold guided by the AUROC, it might be more effective to consider the majority of predictions from the 5 folds. Investigating this approach could lead to a more refined method of setting an optimized threshold for binarization. However, adjusting the threshold for specific cohorts would limit generalizability and would ideally require more data. The goal would be to encompass a greater number of MSI and *POLE* cases while more accurately classifying MSS, thereby minimizing the incidence of falsely predicted cases lacking *POLE/POLD1* driver mutations.

In conclusion, our study demonstrates the capability of a DL screening tool, initially trained for MSI classification, to extend its utility beyond identifying MSI status in patients. Notably, it can also discern MSS CRCs harboring *POLE* mutations. This finding implies a shared histopathologic phenotype between MSI and *POLE* mutations, which a DL system, focused solely on MSI, can effectively detect. As we move forward, such technology might play a pivotal role in the preliminary screening for *POLE* mutations within MSS CRCs, potentially identifying a small yet significant patient group for targeted treatment strategies. Our study underscores the importance of collecting these rare samples, as the significance of these mutations is becoming increasingly evident. Nevertheless, it is paramount to further validate and couple these DL-based screenings with subsequent genetic confirmation to bolster diagnoses. The ability to accurately and efficiently identify CRC patients with pathogenic *POLE* gene mutations could greatly enhance both diagnostic processes and therapeutic outcomes in this field.

Methods

Patient samples

In this study we used the following independent patient cohorts (Supplementary Table 1): DACHS (Darmkrebs: Chancen der Verhütung durch Screening, Southwest Germany, $N_{\text{all}} = 2039$, $N_{\text{MSI}} = 210$ (10.3%))^{34,35}, TCGA (The Cancer Genome Atlas, $N_{\text{all}} = 429$, $N_{\text{MSI}} = 63$ (14.7%)), APHP (Assistance Publique–Hôpitaux de Paris/Public Assistance Hospitals of Paris) resection ($N_{\text{all}} = 27$, $N_{\text{MSI}} = 7$ (25.9%), Fig. 1d) and APHP biopsy ($N_{\text{all}} = 38$, $N_{\text{MSI}} = 13$ (34.2%), Fig. 1d). DACHS is a population-based case-control and patient cohort study on CRC including samples from patients of all tumor stages (I–IV) collected from different laboratories in the south-west of

Germany coordinated by the German Cancer Research Center (Heidelberg, Germany). The APHP cohorts are consecutive case series of a total of $N = 27$ *POLE* mutant, with $N = 17$ *POLE* driver mutant, colorectal cancers collected between 2015 and 2023, as well as a matched cohort of $N = 19$ MSI and $N = 20$ MSS tumors from the same pathology archive. In this cohort the control tumors (MSI and MSS tumors) underwent *POLE* testing and are proven to be *POLE* WT. TCGA is the public repository “The Cancer Genome Atlas”, available at <https://portal.gdc.cancer.gov/>, USA^{55,56}, which includes colorectal cancer of any stage. The microsatellite status for the TCGA samples was determined by Polymerase Chain Reaction (PCR) testing⁵⁷. Molecular features of DACHS, TCGA, APHP surgical resection and biopsy cohorts are shown in Fig. 1c, d (Supplementary Table 1). Sociodemographic and relevant clinical features are shown in Supplementary Table 1. Additional molecular information is presented in Fig. 4 for TCGA and in Supplementary Fig. 1 for DACHS and APHP. A comprehensive list with molecular characteristics of patients with *POLE/POLD1* mutations can be found in Supplementary Table 2 for TCGA and Supplementary Table 3 for APHP. As shown in the tables, this study does not include any *POLD1* driver mutations. Therefore, only *POLE* may be referenced in parts of the following sections. The distribution of MSI status was available for all cohorts (Supplementary Table 1). Details on the methods used for genetic testing of patients in these cohorts are given in Supplementary File 1: Supplementary Methods. Cases of early-onset colorectal cancer (EO-CRC) with patient age under 50 at point of diagnosis are highlighted in the results⁵⁸. This retrospective analysis of scanned images of anonymized tissue samples from various cohorts of cancer patients was conducted in accordance with the Declaration of Helsinki. The data were collected, anonymized, and ethical approval was obtained. The use of tissue samples from DACHS was approved by the ethics committees of the Medical Faculty at Heidelberg University (310/2001) and the state medical boards of Baden-Wuerttemberg and Rhineland-Palatinate⁵⁹. All participants the DACHS study provided written informed consent for the scientific analysis of their data and samples. The use of tissue samples from the APHP cohort was approved by the local ethics committee of Henri Mondor University Hospital (IRB N° 00011558 ; 2021-123). For APHP and TCGA, there was no informed consent required by local regulations for a retrospective analysis of anonymized data. The overall analysis was approved by the ethics board of the Medical Faculty of Technical University Dresden under the ID BO-EK-444102022.

Experimental setup

We used the image data from the DACHS study to train a Deep Learning network to detect MSI in CRC. In this study, only information on MSI status was available, but no information on either *POLE* or *POLD1* status. The methodology we used here corresponds to that of Wagner et al.⁵³. Briefly, we tessellated all digitized whole slide images (WSIs) into tiles and stain normalized them. We further processed the normalized tiles by applying a pre-trained network to extract features from each tile. Based on these features we trained a transformer-based architecture with a classifying multilayer perceptron to classify each slide as MSI or MSS. The ground truth for all cohorts is taken from the clinical data. For TCGA the microsatellite classes contain MSI-H (high grade MSI), MSI-L (low-grade MSI) and MSS. For this study, MSI-H cases are assumed to be MSI while MSI-L and MSS cases both are assigned to MSS as performed by Wagner et al.⁵³. All experiments were performed using 5-fold cross-validation with internal validation on DACHS and external validation on two datasets. For testing the trained models from the 5-fold cross-validation, they were deployed on the external cohorts, TCGA and APHP, for which the *POLE/POLD1* mutation status was available. Our hypothesis applies to *POLE/POLD1* mutations that are meant to include driver mutations only while other genetic alterations like splice or passenger mutations were considered as wild type. For each slide we calculated the classification probabilities of being MSI predicted by the model. We assessed the model performance using the median AUROC of the 5 deployed models on the TCGA external cohort. To investigate the scattering of the models' prediction results, we visualized the results of all folds for each sample respectively and highlighted the prediction score obtained from the

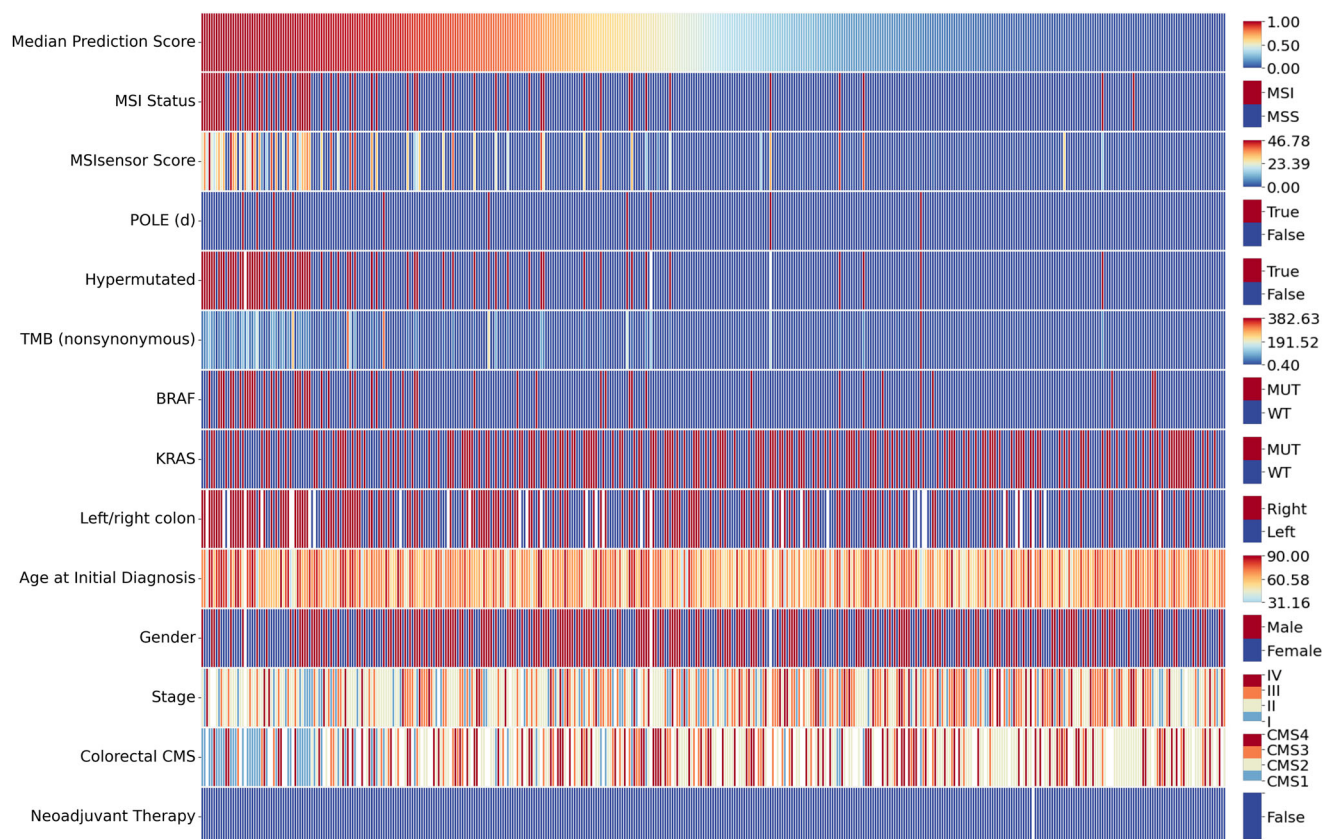


Fig. 4 | Selected molecular characteristics for the TCGA (The Cancer Genome Atlas) cohort. The columns correspond to individual patients and are sorted by “Median Prediction Score”, representing the median model of 5-fold cross-validation evaluated using AUROC. “MSI Status” is the ground truth of microsatellite status. *POLE* (d) denotes patients with a *POLE* driver mutation. TMB represents tumor mutational burden. For BRAF and KRAS “MUT” indicates the presence of the

respective mutation and “WT” the absence of the mutation (wild type). “Stage” ranges from one to four and indicates the cancer stage. The same range is used by “Colorectal CMS” which stands for the consensus molecular subtypes (CMS) of colorectal cancer. Empty cells represent cases where the information regarding specific characteristics is missing.

median model. For explainability, attention heatmaps as well as prediction heatmaps of selected slides were generated. Training and deployment were performed on a NVIDIA RTX A6000 with 48 GB GPU memory. Detailed information on the Deep Learning methods (Fig. 1b) is provided in Supplementary File 1: Supplementary Methods.

Statistics

The primary statistical endpoint for assessment of model performance during training was the AUROC including 95% confidence intervals (CIs). For internal validation (DACHS) as well as for external validation on TCGA, the CI was calculated based on all 5-fold-wise AUROCs. For calculating the statistical power of the AUROCs, we used a two-sided t-test comparing the prediction scores. For individual folds, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated for each subgroup consisting of MSI, MSS and *POLE* cases respectively. For the TCGA and APHP cohorts, we calculated the mean and standard deviation of the prediction scores for each subgroup (MSI, MSS, *POLE*/*POLD1*) for the 5 folds. To do this, we first calculated the mean and standard deviation of the prediction scores for each sample over the 5 folds and then determined the mean of these values over all samples within each subgroup. In the *POLE*/*POLD1* subgroup, only driver mutations were taken into account. For hypothesis testing of differences between relative frequencies of morphologic features, a two-sided Fisher’s exact test was used.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The whole slide images (WSIs), molecular and clinical data for the TCGA cohort are publicly accessible at <https://portal.gdc.cancer.gov/> and <https://www.cbiportal.org/> (accessed 12 October 2022). For this study we used the MSI status for the TCGA cohort based on the findings of Liu et al.⁵⁷ at <https://github.com/KatherLab/cancer-metadata/blob/main/tcga/liu.xlsx> (accessed 06 November 2022). The datasets from DACHS and APHP are available from the corresponding author on reasonable request. All data generated or analyzed during this study are included in this published article and its supplementary information files.

Code availability

All source code for image processing is publicly available at <https://github.com/KatherLab> and given with the exact commit ID for reasons of reproducibility. The tessellation script is available at <https://github.com/KatherLab/preprocessing-ng> and followed by normalization with the script and the reference image available at <https://github.com/KatherLab/preProcessing>. Extraction of CTransPath features was carried out with scripts from <https://github.com/KatherLab/marugoto> as well as the transformer-based Deep Learning model and the heatmaps for explainability used for the study. Links to the exact version to the repositories including the scripts used can be found in Supplementary Table 4.

Received: 7 November 2023; Accepted: 14 April 2024;

Published online: 23 May 2024

References

- Motta, R. et al. Immunotherapy in microsatellite instability metastatic colorectal cancer: current status and future perspectives. *Transl. Res.* **7**, 511–522 (2021).
- Westdorp, H. et al. Opportunities for immunotherapy in microsatellite unstable colorectal cancer. *Cancer Immunol. Immunother.* **65**, 1249–1259 (2016).
- Mulet-Margalef, N. et al. Challenges and therapeutic opportunities in the dMMR/MSI-H colorectal cancer landscape. *Cancers* **15**, 1022 (2023).
- Chalabi, M. et al. LBA7 - Neoadjuvant immune checkpoint inhibition in locally advanced MMR-deficient colon cancer: The NICHE-2 study. *Ann. Oncol.* **33**: S808–S869 (2022).
- Echle, A. et al. Deep learning for the detection of microsatellite instability from histology images in colorectal cancer: a systematic literature review. *Immunoinformatics* **3–4**, 100008 (2021).
- Kather, J. N. et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **25**, 1054–1056 (2019).
- Saldanha, O. L. et al. Swarm learning for decentralized artificial intelligence in cancer histopathology. *Nat. Med.* <https://doi.org/10.1038/s41591-022-01768-5> (2022).
- Echle, A. et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. *Gastroenterology* **159**, 1406–1416.e11 (2020).
- Yamashita, R. et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *Lancet Oncol.* **22**, 132–141 (2021).
- Schirris, Y., Gavves, E., Nederlof, I., Horlings, H. M. & 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. *Med. Image Anal.* <https://doi.org/10.1016/j.media.2022.102464> (2022).
- Schrammen, P. L. et al. Weakly supervised annotation-free cancer detection and prediction of genotype in routine histopathology. *J. Pathol.* **256**, 50–60 (2022).
- Bilal, M. et al. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. *Lancet Digit. Health* **3**, e763–e772 (2021).
- Echle, A. et al. Artificial intelligence for detection of microsatellite instability in colorectal cancer—a multicentric analysis of a pre-screening tool for clinical application. *ESMO Open* **7**, 100400 (2022).
- Saillard, C. et al. Validation of MSIntuit as an AI-based pre-screening tool for MSI detection from colorectal cancer histology slides. *Nat. Commun.* **14**, 1–11 (2023).
- Kleppe, A. et al. A clinical decision support system optimising adjuvant chemotherapy for colorectal cancers by integrating deep learning and pathological staging markers: a development and validation study. *Lancet Oncol.* **23**, 1221–1232 (2022).
- Arsilan, S. et al. Evaluation of a predictive method for the H&E-based molecular profiling of breast cancer with deep learning. *bioRxiv* <https://doi.org/10.1101/2022.01.04.474882> (2022).
- Kacaw, A. J. et al. Artificial intelligence can cut costs while maintaining accuracy in colorectal cancer genotyping. *Front. Oncol.* **11**, 630953 (2021).
- Ma, X., Dong, L., Liu, X., Ou, K. & Yang, L. POLE/POLD1 mutation and tumor immunotherapy. *J. Exp. Clin. Cancer Res.* **41**, 216 (2022).
- Strauss, J. D. & Pursell, Z. F. Replication DNA polymerases, genome instability and cancer therapies. *NAR Cancer* **5**, zcad033 (2023).
- Domingo, E. et al. Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. *Lancet Gastroenterol. Hepatol.* **1**, 207–216 (2016).
- Garmezy, B. et al. Clinical and molecular characterization of POLE mutations as predictive biomarkers of response to immune checkpoint inhibitors in advanced cancers. *JCO Precis. Oncol.* **6**, e2100267 (2022).
- Wang, F. et al. Evaluation of POLE and POLD1 mutations as biomarkers for immunotherapy outcomes across multiple cancer types. *JAMA Oncol.* **5**, 1504–1506 (2019).
- Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
- Sun Yat-sen University. *Toripalimab as Monotherapy in Participants With POLE or POLD-1 Mutated and Non-MSI-H Advanced Solid Tumors*. CTG Labs - NCBI. <https://clinicaltrials.gov/study/NCT03810339> (2019).
- Kawai, T. et al. Clinical and epigenetic features of colorectal cancer patients with somatic POLE proofreading mutations. *Clin. Epigenet.* **13**, 117 (2021).
- Greenon, J. K. et al. Pathologic predictors of microsatellite instability in colorectal cancer. *Am. J. Surg. Pathol.* **33**, 126–133 (2009).
- Shia, J. et al. Morphological characterization of colorectal cancers in The Cancer Genome Atlas reveals distinct morphology-molecular associations: clinical and biological implications. *Mod. Pathol.* **30**, 599–609 (2017).
- Reitsam, N. G. et al. Concurrent loss of MLH1, PMS2 and MSH6 immunoexpression in digestive system cancers indicating a widespread dysregulation in DNA repair processes. *Front. Oncol.* **12**, 1019798 (2022).
- Hwang, H. S., Kim, D. & Choi, J. Distinct mutational profile and immune microenvironment in microsatellite-unstable and POLE-mutated tumors. *J. Immunother. Cancer* **9**, e002797 (2021).
- Morales-Juarez, D. A. & Jackson, S. P. Clinical prospects of WRN inhibition as a treatment for MSI tumours. *NPJ Precis. Oncol.* **6**, 85 (2022).
- Roberts, S. A. & Gordenin, D. A. Hypermutation in human cancer genomes: footprints and mechanisms. *Nat. Rev. Cancer* **14**, 786–800 (2014).
- Park, V. S. & Pursell, Z. F. POLE proofreading defects: contributions to mutagenesis and cancer. *DNA Repair* **76**, 50–59 (2019).
- Caracciolo, D. et al. Error-prone DNA repair pathways as determinants of immunotherapy activity: an emerging scenario for cancer treatment. *Int. J. Cancer* **147**, 2658–2668 (2020).
- Brenner, H., Chang-Claude, J., Seiler, C. M. & Hoffmeister, M. Long-term risk of colorectal cancer after negative colonoscopy. *J. Clin. Oncol.* **29**, 3761–3767 (2011).
- Hoffmeister, M. et al. Statin use and survival after colorectal cancer: the importance of comprehensive confounder adjustment. *J. Natl Cancer Inst.* **107**, djv045 (2015).
- Hühns, M. et al. High mutational burden in colorectal carcinomas with monoallelic POLE mutations: absence of allelic loss and gene promoter methylation. *Mod. Pathol.* **33**, 1220–1231 (2020).
- Hu, H. et al. Ultra-mutated colorectal cancer patients with POLE driver mutations exhibit distinct clinical patterns. *Cancer Med.* **10**, 135–142 (2021).
- Silberman, R. et al. Complete and prolonged response to immune checkpoint blockade in POLE-mutated colorectal cancer. *JCO Precis. Oncol.* **3**, 1–5 (2019).
- Wang, C., Gong, J., Tu, T. Y., Lee, P. P. & Fakhri, M. Immune profiling of microsatellite instability-high and polymerase ϵ (POLE)-mutated metastatic colorectal tumors identifies predictors of response to anti-PD-1 therapy. *J. Gastrointest. Oncol.* **9**, 404–415 (2018).
- Kanavati, F. et al. A deep learning model for the classification of indeterminate lung carcinoma in biopsy whole slide images. *Sci. Rep.* **11**, 8110 (2021).
- Golmard, L., Golmard, J.-L. & Stoppa-Lyonnet, D. Evaluation of POLE/POLD1 Variants as Potential Biomarkers for Immune Checkpoint Inhibitor Treatment Outcomes. *JAMA Oncol.* **6**, 588–589 (2020).
- Maddalena, G. et al. 631P Using the unique somatic mutation profile of POLE loss of proof-reading mutation helps in selection of patients

- who may benefit from immunotherapy. *Ann. Oncol.* **34**, S448–S449 (2023).
43. Campbell, B. B. et al. Comprehensive analysis of hypermutation in human cancer. *Cell* **171**, 1042–1056.e10 (2017).
 44. Ahn, S.-M. et al. The somatic POLE P286R mutation defines a unique subclass of colorectal cancer featuring hypermutation, representing a potential genomic biomarker for immunotherapy. *Oncotarget* **7**, 68638–68649 (2016).
 45. Malik, A., Bhatia, J. K., Sahai, K., Boruah, D. & Sharma, A. Evaluating morphological features for predicting microsatellite instability status in colorectal cancer. *Armed Forces Med. J. India* **78**, S96–S104 (2022).
 46. Hodi, F. S. et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010).
 47. Garon, E. B. et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.* **372**, 2018–2028 (2015).
 48. Zhang, C. et al. Incidence and detection of high microsatellite instability in colorectal cancer in a Chinese population: a meta-analysis. *J. Gastrointest. Oncol.* **11**, 1155–1163 (2020).
 49. André, T. et al. Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. *N. Engl. J. Med.* **383**, 2207–2218 (2020).
 50. Bonneville, R. et al. Detection of microsatellite instability biomarkers via next-generation sequencing. *Methods Mol. Biol.* **2055**, 119–132 (2020).
 51. Van Gool, I. C. et al. Blinded histopathological characterisation of POLE exonuclease domain-mutant endometrial cancers: sheep in wolf's clothing. *Histopathology* **72**, 248–258 (2018).
 52. McGrail, D. J. et al. High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann. Oncol.* **32**, 661–672 (2021).
 53. Wagner, S. J. et al. Transformer-based biomarker prediction from colorectal cancer histology: a large-scale multicentric study. *Cancer Cell* **41**, 1650–1661.e4 (2023).
 54. Lee, S. H., Song, I. H. & Jang, H.-J. Feasibility of deep learning-based fully automated classification of microsatellite instability in tissue slides of colorectal cancer. *Int. J. Cancer* **149**, 728–740 (2021).
 55. Cancer Genome Atlas Network Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
 56. Isella, C., Cantini, L., Bellomo, S. E. & Medico, E. *TCGA CRC 450 dataset* (2014).
 57. Liu, Y. et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell* **33**, 721–735.e8 (2018).
 58. Brockway-Lunardi, L. et al. Early-onset colorectal cancer research: gaps and opportunities. *Colorectal Cancer* **9**, CRC34 (2020).
 59. Brenner, H., Chang-Claude, J., Seiler, C. M., Stürmer, T. & Hoffmeister, M. Does a negative screening colonoscopy ever need to be repeated? *Gut* **55**, 1145–1150 (2006).

Acknowledgements

J.N.K. is supported by the German Federal Ministry of Health (DEEP LIVER, ZMW1-2520DAT111) and the Max-Eder-Programme of the German Cancer Aid (grant #70113864), the German Federal Ministry of Education and Research (PEARL, 01KD2104C; CAMINO, 01EO2101; SWAG, 01KD2215A; TRANSFORM LIVER, 031L0312A; TANGERINE, 01KT2302 through ERA-NET Transcan), the German Academic Exchange Service (SECAI, 57616814), the German Federal Joint Committee (Transplant.KI, 01VSF21048) the European Union's Horizon Europe and innovation programme (ODELIA, 101057091; GENIAL, 101096312) and the National Institute for Health and Care Research (NIHR, NIHR213331) Leeds Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. N.P.W. and P.Q. are funded by Yorkshire Cancer Research L386 and L394 and NIHR as above. The DACHS study (H.B. and M.H.) was

supported by the German Research Foundation (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HO 5117/2-2, HE 5998/2-1, HE 5998/2-2, KL 2354/3-1, KL 2354/3-2, RO 2270/8-1, RO 2270/8-2, BR 1704/17-1, and BR 1704/17-2), the Interdisciplinary Research Program of the NCT (Germany), and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A, and 01ER1505B). The study was further supported by project funding for the PEARL consortium from the German Federal Ministry of Education and Research (01KD2104A).

Author contributions

M.G., J.C., C.M.L.L., A.P., and J.N.K. conceptualized the study. J.C., T.Y., N.P.W., P.Q., T.J.B., L.F., A.B., A.S., H.B., M.H., and A.P. provided clinical and histopathological data. M.G. curated the source data. M.G. and M.v.T. implemented the source code for the analysis. M.v.T. developed the code for the heatmaps. M.G. conducted the experiments and data analysis. M.G., N.G.R., Z.I.C., and C.M.L.L. interpreted the data. N.G.R. and B.M. did the pathological examination of the samples and analyzed the heatmaps. M.G. wrote the first draft of the manuscript. All authors revised the manuscript draft, jointly interpreted the data and agreed to the submission of this article.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

J.N.K. declares consulting services for Owkin, France; DoMore Diagnostics, Norway; AstraZeneca, UK and Panakeia, UK and has received honoraria for lectures by Eisai, Roche, MSD, and Fresenius. J.N.K. is Deputy Editor at “npj Precision Oncology”. M.G. has received honoraria from Techniker Krankenkasse. P.Q. declares research funding from Roche and honoraria for lectures by Roche, Bayer, Amgen. B.M. has received honoraria from Merck, Roche, MSD, BMS, and Bayer. A.P. declares consulting services for AstraZeneca, Biocartis and declares research funding from Bayer. No other potential conflicts of interest are reported by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41698-024-00592-z>.

Correspondence and requests for materials should be addressed to Jakob Nikolas Kather.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

¹Else Kroener Fresenius Center for Digital Health, Medical Faculty Carl Gustav Carus, Technical University Dresden, Dresden, Germany. ²Pathology, Faculty of Medicine, University of Augsburg, Augsburg, Germany. ³Department of Medicine I, University Hospital and Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. ⁴Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁵Pathology & Data Analytics, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, United Kingdom. ⁶Digital Biomarkers for Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁷Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. ⁸German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁹Université Paris Est Créteil, INSERM, IMRB, Créteil, France. ¹⁰Assistance Publique-Hôpitaux de Paris, Henri Mondor-Albert Chenevier University Hospital, Department of Pathology, Créteil, France. ¹¹INSERM, U955, Team Oncogenèse des lymphomes et tumeurs de la Neurofibromatose 1, Créteil, France. ¹²Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany. ¹³Tissue Bank of the National Center for Tumor Diseases (NCT) Heidelberg, Heidelberg, Germany. ¹⁴Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany. ¹⁵These authors jointly supervised this work: Anaïs Pujals, Jakob Nikolas Kather.

✉ e-mail: jakob-nikolas.kather@alumni.dkfz.de