

1 HIBRID: Histology and ct-DNA based Risk- 2 stratification with Deep Learning

3
4 Chiara M.L. Loeffler* (1, 2, 3), Hideaki Bando* (4,5,6), Srividhya Sainath (1),
5 Hannah Sophie Muti (1, 2, 7), Xiaofeng Jiang (1), Marko van Treeck (1), Nic Gabriel Reitsam
6 (1,8,9), Zunamys I. Carrero (1), Tomomi Nishikawa (4), Toshihiro Misumi (4), Saori Mishima
7 (5), Daisuke Kotani (5), Hiroya Taniguchi (10), Ichiro Takemasa (11), Takeshi Kato (12), Eiji
8 Oki (13), Tanwei Yuan (14), Durgesh Wankhede (14), Sebastian Foersch (15), Hermann
9 Brenner (14, 16), Michael Hoffmeister (14), Yoshiaki Nakamura (5,6), Takayuki Yoshino*
10 (4,5,6), Jakob Nikolas Kather* (1, 2, 3, 16)

11

12 * Equal contribution

13 + Correspondence to jakob-nikolas.kather@alumni.dkfz.de and tyoshino@east.ncc.go.jp

- 14 1. Else Kroener Fresenius Center for Digital Health, Technical University Dresden,
15 Dresden, Germany
- 16 2. Medical Department 1, University Hospital and Faculty of Medicine Carl Gustav
17 Carus, Technische Universität Dresden, Dresden, Germany
- 18 3. National Center for Tumor Diseases Dresden (NCT/UCC), a partnership between
19 DKFZ, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden
20 University of Technology, and Helmholtz-Zentrum Dresden - Rossendorf (HZDR),
21 Dresden, Germany
- 22 4. Department of Data Science, National Cancer Center Hospital East, Kashiwa, Japan
- 23 5. Department of Gastroenterology and Gastrointestinal Oncology, National Cancer
24 Center Hospital East, Kashiwa, Japan
- 25 6. Translational Research Support Office, National Cancer Center Hospital East,
26 Kashiwa, Japan
- 27 7. Department for Visceral, Thoracic and Vascular Surgery, University Hospital and
28 Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden,
29 Germany
- 30 8. Pathology, Faculty of Medicine, University of Augsburg, Augsburg, Germany
- 31 9. Bavarian Cancer Research Center (BZKF), Augsburg, Germany
- 32 10. Department of Clinical Oncology, Aichi Cancer Center Hospital, Nagoya, Japan
- 33 11. Department of Surgery, Surgical Oncology and Science, Sapporo Medical University,
34 Sapporo, Japan
- 35 12. Department of Surgery, NHO Osaka National Hospital, Osaka, Japan

36 13. Department of Surgery and Science, Graduate School of Medical Sciences, Kyushu
37 University, Fukuoka, Japan

38 14. Division of Clinical Epidemiology and Aging Research, German Cancer Research
39 Center (DKFZ), Heidelberg, Germany

40 15. Institute of Pathology, University Medical Center Mainz, Mainz, Germany.

41 16. Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital
42 Heidelberg, Heidelberg, Germany

43 Highlights

- 44 - This study combines MRD status measured by ctDNA with a DL-based risk
45 assessment trained on histological image data to enhance recurrence prediction.
- 46 - DL-based spatial assessment of tumor histopathology slides significantly improves
47 the risk stratification provided by MRD alone.
- 48 - MRD-negative patients with high DL-based risk had a significantly longer DFS if
49 treated with ACT, compared to MRD-negative and DL low risk patients
- 50 - The DL model is fully open-source and publicly available.

51 Keywords

52 Deep Learning, molecular residual disease, circulating tumor DNA, colorectal cancer, vision
53 transformers

54 Abstract

55 **Background:** Although surgical resection is the standard therapy for stage II/III colorectal
56 cancer (CRC), recurrence rates exceed 30%. Circulating tumor DNA (ctDNA) emerged as a
57 promising recurrence predictor, detecting molecular residual disease (MRD). However,
58 spatial information about the tumor and its microenvironment is not directly measured by
59 ctDNA. Deep Learning (DL) can predict prognosis directly from routine histopathology slides.

60 **Methods:** We developed a DL pipeline utilizing vision transformers to predict disease-free
61 survival (DFS) based on histological hematoxylin & eosin (H&E) stained whole slide images
62 (WSIs) from patients with resectable stage II-IV CRC. This model was trained on the DACHS
63 cohort (n=1766) and independently validated on the GALAXY cohort (n=1555). Patients
64 were categorized into high- or low-risk groups based on the DL-prediction scores. In the
65 GALAXY cohort, the DL-scores were combined with the four-weeks post-surgery MRD
66 status measured by ctDNA for prognostic stratification.

67 **Results:** In GALAXY, the DL-model categorized 307 patients as DL high-risk and 1248
68 patients as DL low-risk ($p < 0.001$; HR 2.60, CI 95% 2.11-3.21). Combining the DL scores
69 with the MRD status significantly stratified both the MRD-positive group into DL high-risk
70 ($n=81$) and DL low-risk ($n=160$) (HR 1.58 (CI 95% 1.17-2.11; $p=0.002$) and the MRD-
71 negative group into DL high-risk ($n=226$) and DL low-risk ($n=1088$) (HR 2.37 CI 95% 1.73-
72 3.23; $p < 0.001$). Moreover, MRD-negative patients had significantly longer DFS when
73 predicted as DL high-risk and treated with ACT (HR 0.48, CI 95% 0.27-0.86; $p=0.01$),
74 compared to the MRD-negative patients predicted as DL low-risk (HR=1.14, CI 95% 0.8-
75 1.63; $p=0.48$).

76 **Conclusion:** DL-based spatial assessment of tumor histopathology slides significantly
77 improves the risk stratification provided by MRD alone. Combining histologic information with
78 ctDNA yields the most powerful predictor for disease recurrence to date, with the potential to
79 improve follow-up, withhold adjuvant chemotherapy in low-risk patients and escalate
80 adjuvant chemotherapy in high-risk patients.

81

82 Introduction

83 Colorectal cancer (CRC) is one of the leading causes of cancer-related deaths worldwide¹.
84 Surgical resection remains the standard curative therapy in patients with Stage II-III CRC
85 and resectable metastases. Despite advancements in surgical and adjuvant therapies,
86 recurrence rates exceed 30% and 60%^{2,3}, respectively. Patients who relapse have an
87 increased mortality risk, hence identifying these patients at an early stage is crucial for
88 optimising follow-up treatment decisions. Current prognostication systems for risk
89 assessment, including imaging techniques, clinicopathological features and molecular data,
90 are moderate predictors for recurrence risk. Similarly, follow-up strategies, such as tumor
91 marker monitoring with carcinoembryonic antigen (CEA), lack sensitivity and specificity in
92 identifying recurrence⁴⁻⁶. In particular for stage II CRC, the decision on adjuvant
93 chemotherapy (ACT) is based on diverging risk assessment recommendations provided
94 through international oncological associations^{7,8}. Thus, a more fine-grained system for
95 estimating the risk of relapse is required, as no stage-specific survival benefit for adjuvant
96 chemotherapy has been proven. Therefore, new biomarkers for better and more precise
97 prognostication are needed. Circulating tumor DNA (ctDNA) has emerged as a promising
98 minimally invasive biomarker that measures a small fraction of ctDNA in the blood, allowing
99 for the detection of molecular residual disease (MRD) status⁹. Additionally, ctDNA can be
100 used for monitoring treatment response and early prediction of recurrence, as ctDNA
101 positivity after surgery is associated with a higher risk of disease recurrence^{10,11}. Previous

102 studies have shown that this correlation had already been found as early as four weeks after
103 primary tumor resection¹². However, ctDNA analysis alone does not capture the
104 morphological characteristic of the tumor. For instance, information such as histopathological
105 subtype, grading, vascular and lymphatic invasion, as well as the abundance of tumor-
106 infiltrating lymphocytes¹³⁻¹⁶, among many other morphological properties of the tumor
107 microenvironment (TME), have been shown to be prognostically relevant and are reflected in
108 current clinical guidelines^{17,18}. Deep Learning (DL) is an artificial intelligence technology
109 which is useful to extract quantitative biomarkers from routinely available clinical data in
110 oncology^{19,20}. DL models, trained on histopathological routine hematoxylin and eosin (H&E)
111 tumor slides have been shown to act as survival prediction models outperforming current
112 risk-stratifications systems²¹⁻²³. DL can extract highly relevant information from routine
113 pathology slides of CRC, including presence of microsatellite instability (MSI)^{24,25}, gene
114 mutations^{25,26}, response to neoadjuvant therapy²⁷, and overall survival (OS)²². Given the
115 ability of DL to extract meaningful biological information from pathology slides that ctDNA
116 cannot capture, we hypothesise that the combination of MRD assessment with a
117 transformer-based DL risk score from morphology could significantly improve prognosis
118 prediction. In this study, we aim to enhance patient stratification and recurrence prediction in
119 patients with CRC by integrating MRD status derived from ctDNA with a DL-based risk score
120 trained on routine histological images.

121 **Methods & Materials**

122 **Patient Data Acquisition**

123 In this study, we analysed histological whole slide images (WSIs) of hematoxylin & eosin
124 (H&E) stained tumour tissue of surgically curable CRC from two large cohorts in Germany
125 and Japan (Figure 1A-B, Supplementary Figure 1). The first cohort was the Darmkrebs:
126 Chancen der Verhütung durch Screening Study (DACHS), which includes 1774 WSI's
127 belonging to 1766 patients and was used as a training cohort (Supplementary Figure 1A).
128 The second cohort was the GALAXY trial from the CIRCULATE-Japan study
129 (UMIN000039205), which includes 1556 WSIs from 1555 patients and was used as an
130 independent external validation cohort (Supplementary Figure 1B). The GALAXY trial
131 comprised ctDNA data measuring the MRD status at the four weeks post-surgery interval:
132 MRD positivity was defined as at least 2 out of 16 tumour-specific ctDNA variants detected
133 above a predefined threshold based on Natera's method^{12,28}. Out of the 1555 patients
134 included in the trial, 241 were MRD-positive and 1314 patients were MRD-negative at the
135 respective 4 weeks interval¹² (Figure 1B). For both cohorts, disease free survival in months

136 (DFS) was available. DFS marked the time from primary surgery to last follow-up date or last
137 surgery to last follow-up date for patients for which primary surgery date was unavailable.

138 **Image Processing and Deep Learning Techniques**

139 **Data Preprocessing**

140 All whole-slide images (WSIs) were segmented into image patches with dimensions of
141 224×224 pixels and an edge length of $256 \mu\text{m}$, resulting in an effective magnification of
142 $1.14 \mu\text{m}$ per pixel. During this segmentation process, patches that primarily contained
143 background or blur (identified by having an average number of Canny edges below a
144 threshold of 2) were removed from the dataset. The retained image tiles were colour
145 normalised using the Macenko method in order to avoid stain-associated bias²⁹. For WSI
146 pre-processing, we employed our end-to-end publicly available pipeline, which can be found
147 here: <https://github.com/KatherLab/end2end-WSI-preprocessing>.

148

149 **Model Development**

150 To train and validate our prediction DL-models we used our open-source pipeline, marugoto
151 (<https://github.com/KatherLab/marugoto>). In the initial step, a self-supervised learning (SSL)
152 model called UNI, pretrained on over 100 M histology-specific images and 100k WSIs, was
153 employed to extract a 1024-dimensional feature vector from each image tile (Figure 1A-B)³⁰.
154 The obtained features were then preprocessed using a multi-headed self-attention
155 mechanism by the transformer network (Figure 1C). Here, the network views patch
156 embeddings as a sequence, with elements interacting through self-attention. For a WSI with
157 n patches of dimension d , self-attention calculates a query-key product. Multi-headed self-
158 attention repeats this in h heads, then concatenating and transforming the outputs. The
159 transformer architecture is designed with two layers, each featuring eight heads (total $h=8$),
160 a latent dimension of 512, and equal dimensions (each 64) for queries, keys, and values.
161 Post self-attention, the embeddings of each patch are combined into a sequence of
162 dimension $n \times 1024$ processed through a linear projection and ReLU activation to reduce
163 dimensionality to 512. A learnable class token is added to this sequence, resulting in an
164 input dimension of $(n+1) \times 512$ that is fed into the transform layer. Each transformer layer
165 consists of a layer normalisation block followed by multi-headed self-attention, a block of
166 layer normalisation and finally a multi-layer perceptron (MLP), with skip connections
167 integrated across each block to facilitate training²⁴. After processing through two transformer
168 layers, the class token is inputted into an MLP head designed to produce a continuous risk
169 score for each patient, serving as the output of the model. For model training, we used Cox
170 partial likelihood, as the loss function^{22,31}. We randomly split the DACHS cohort at the patient

171 level into training, validation, and test sets in a 4:4:2 ratio. The model was trained using the
172 training set, and the best checkpoint, determined by the highest C-index on the validation
173 set, was saved. This checkpoint was then validated on DACHS test set and CIRCULATE
174 data cohort

175

176 **Visualisation**

177 To interpret our model's output, we generated whole-slide patient heatmaps showing the DL
178 prediction scores. We used our trained Vision Transformer (ViT) model to process tile-level
179 features extracted from the WSI. The features are passed through the trained model to
180 obtain tile-level scores, which are then combined with Grad-CAM (Gradient-weighted Class
181 Activation Mapping) values to generate weighted scores, which were normalised to a range
182 of -1 to 1 facilitating the identification of the most significant tiles. Heatmaps were then
183 created using the weighted scores, with red indicating high-risk, and blue indicating low-risk.
184 To maintain interpretability, we blended these heatmaps with the original image features,
185 providing clear insights into the tumor morphology and the model's predictions.

186 **Experimental Design**

187 In our study we first trained a transformer-based DL model on the DACHS cohort, utilizing
188 clinical data on disease-free survival (DFS) events and DFS time in months to generate
189 patient level DL-based risk scores (Figure 1C). Next, we externally validated the trained DL
190 model on the GALAXY cohort. The continuous DL-risk score was binarized into DL high-risk
191 and DL low-risk categories based on a fixed threshold, defined as the median risk score in
192 the training cohort (0.9357855). Subsequently, we combined the four-week post-surgery
193 MRD status from the GALAXY trial with the DL-risk scores to analyze survival differences
194 between these subgroups (Figure 1D). We also looked at the effects of adjuvant
195 chemotherapy in the various subgroups. Survival Analysis was performed using Kapan-
196 Meier analysis and log-rank test to compare DFS time between the groups. Additionally,
197 multivariate analysis was conducted using Cox proportional hazard models, including the
198 covariates: age, gender, pathological T-Stage (pT) and pathological N-Stage (pN)²². Lastly,
199 we performed a morphological analysis to identify histopathological correlations between the
200 DL high-risk and low-risk subgroups, using classification heatmaps (Figure 1D).

201 **Data and Code availability**

202 Our whole slide image preprocessing pipeline is available here:
203 <https://github.com/KatherLab/end2end-WSI-preprocessing>. The code for the pretrained
204 vision encoder UNI can be found under: <https://github.com/mahmoodlab/uni>. Our DL model

205 codes are publicly available at [https://github.com/KatherLab/marugoto/tree/survival-](https://github.com/KatherLab/marugoto/tree/survival-transformer/marugoto/survival)
206 [transformer/marugoto/survival](https://github.com/KatherLab/marugoto/tree/survival-transformer/marugoto/survival). The respective study Principal Investigators provided the
207 remaining data. For detailed data sharing policies, please refer to the original publications.

208 **Results**

209 **DL stratifies patients by recurrence risk**

210 We trained a DL model to generate risk scores based on DFS and validated its performance
211 on the GALAXY cohort. Based on the DL-risk scores, we divided our cohort into DL high-
212 and DL low-risk groups, followed by a survival analysis using Kaplan Meier estimator and
213 Cox proportional hazard analysis (Figure 1D). These results were then compared with the
214 stratification outcomes of MRD status four weeks post-surgery in the GALAXY cohort.
215 Among the 1,555 patients 19.8% (n=307) were categorized as DL high-risk and 80.2%
216 (n=1248) as DL low-risk. Patients classified as DL high-risk exhibited a significantly elevated
217 risk of disease recurrence compared to DL low-risk patients (HR=2.6, CI 95% 2.11-3.21; $p <$
218 0.005), with a 20-month DFS of 59.3% vs. 82.1%, respectively (Figure 1D). The ctDNA
219 analysis alone stratified 15.5 % (n=241) patients as MRD-positive and 84.5% (n=1,314) as
220 MRD-negative, with an HR of 11.4 (CI 95% 9.28-14, $p < 0.001$, Supplementary Figure 2A). In
221 the multivariate analysis, including the covariates age, sex, pT, and pN, we found the most
222 prognostic indicator for recurrence risk to be MRD positivity (HR=10.57, CI 95% 8.26-13.53;
223 $p < 0.001$), followed by pT3-pT4-Stage (HR=2.00, CI 95% 1.20-3.35; $p < 0.05$, Figure 1F). The
224 DL-risk score was significant with an HR of 1.46 (CI 95% 1.11-1.90, $p < 0.05$). When
225 correlating the DL risk categories with patient characteristics, we found significant
226 differences in sex, pT-Stage, pN-Stage, pathological Stage, and MRD status (Table 1).
227 Together, these data demonstrate that the DL model can significantly stratify patients
228 according to their risk of recurrence.

229 **DL stratifies recurrence risk within MRD subgroups**

230 We hypothesised that by integrating the MRD status with our DL risk score we can further
231 stratify the patients according to risk of recurrence, particularly the MRD-negative patients.
232 To test this, we combined the binarized DL-derived risk score with the MRD status four-
233 weeks after curative surgery (Figure 2). In the MRD-positive group, 33.6% (81 out of 241
234 patients) were categorized as DL high-risk and 66.4% (160 out of 241 patients) as DL low-
235 risk, with an HR of 1.57 (CI 95% 1.18-2.12; $p = 0.002$, Figure 2A). The DFS-time interval was
236 longer in the DL low-risk group, with a 20-months DFS of 30.9% compared to 9.9% in the DL
237 high-risk group (Figure 2A).

238 In the MRD-negative group, 17.2% (226 out of 1,314 patients) were classified as high-risk by
239 the DL model and 82.8% (1,088 out of 1,314 patients) as DL low-risk with an HR of 2.36 (CI
240 95% 1.73-3.23; $p < 0.001$, Figure 2B). Additionally, the 20-month DFS was longer in the DL
241 low-risk group at 89.7%, compared to 77% in the DL high-risk group (Figure 2B). In a
242 multivariate Cox analysis with age, sex, pT and pN as covariates, the DL-score was the only
243 independent prognostic predictor in the MRD-positive group with an HR of 1.51 (CI 95%
244 1.07-2.14; $p = 0.018$, Supplementary Figure 2B). In the MRD-negative group, pT1-T2 was the
245 strongest prognostic indicator with an HR of 2.50 (CI 95% 1.73-3.64; $p < 0.001$,
246 Supplementary Figure 2C). The DL risk score, with an HR of 1.28 (CI 95% 0.83-1.99;
247 $p = 0.27$), was not an independent prognostic predictor (Supplementary Figure 2C). In
248 summary, these data show that the combination of MRD status with the DL risk score
249 enables a better stratification of patients with CRC.

250 **DL-based recurrence risk predicts benefit from adjuvant chemotherapy** 251 **in MRD-negative patients**

252 We hypothesised that our DL-risk score could identify patients with stage II-IV CRC who
253 might benefit from ACT, despite being MRD-negative. To test this hypothesis, we explored
254 the association of ACT with DFS by performing Kaplan-Meier analysis within the DL high-risk
255 and low-risk subgroups among both MRD-positive and MRD-negative patients (Figure 2 C-
256 F). For the MRD-positive group, patients receiving ACT had significantly longer DFS in
257 both the DL low-risk group (HR=0.20, CI 95% 0.14-0.30; $p < 0.001$, Figure 2C) and in the DL
258 high-risk group (HR=0.25, CI 95% 0.16-0.43; $p < 0.001$, Figure 2E). Without receiving ACT, all
259 MRD-positive and DL high-risk patients experienced recurrence within 20-months, whereas
260 18.6% of the MRD-positive and DL high-risk patients who received ACT remained disease-
261 free after 20 months (Figure 2E). In the MRD-negative group, patients in the low-risk DL
262 group did not have longer DFS when treated with ACT (HR=1.14, CI 95% 0.8-1.63; $p = 0.48$).
263 The 20-month DFS was 89.4% for patients treated with ACT vs 89.9% for patients not
264 receiving ACT (Figure 2D). Interestingly, patients in the MRD-negative and DL high-risk
265 group showed significantly longer DFS when treated with ACT (HR 0.48, CI 95% 0.27-0.86;
266 $p = 0.01$, Figure 2D). The 20-month DFS rate was 86.2% in patients who received ACT and
267 thus significantly higher than in patients who did not receive ACT (70.5%). This disease-free
268 survival advantage continued to be seen in the 40-month DFS rate at 83% (with ACT) vs
269 68.9% (without receiving ACT, Figure 2F).

270 Together, these data show that the DL prognostication model can successfully further stratify
271 MRD-negative patients. This indicated that even within the low-risk subgroup (according to

272 MRD), there are high-risk individuals for whom the omission of ACT may carry a higher risk
273 of recurrence.

274 **DL can identify histopathological features linked to prognosis**

275 Measurements of ctDNA provide information about viable and disseminated tumor cells,
276 serving as surrogate markers for their presence in the body and enabling a non-invasive
277 assessment of MRD after surgery. However, they do not provide any information regarding
278 tumor morphology as well as the TME, which is reflected in histopathology slides and is
279 known to be related to clinical outcomes. We investigated whether our model trained on
280 histopathology images without any manual annotation, learned to consider morphological
281 features of the tumor and the TME, which would be synergistic to MRD status. We used a
282 model trained on DACHS and deployed on GALAXY, visualising highly predictive regions at
283 both high and low magnification, as shown in Figure 3.

284 In the DL low-risk classified patients, the morphological analysis revealed a variety of benign
285 histopathological tissue features (Supplementary Figure 3A). As the DL score increased, the
286 histological image tiles still below the risk threshold displayed moderately differentiated
287 tumor components. These samples still displayed a balanced tumor-stroma ratio and tumor
288 glands with tubular to cribriform architecture, indicating an intermediate phenotype between
289 DL low and DL high-risk morphological characteristics (Figure 3A-B). The images, above the
290 risk threshold, displayed high-grade tumor cells with a significant desmoplastic stroma
291 reaction. There was a high intratumoral stroma fraction, and the presence of tumor
292 buds/poorly differentiated clusters, which are known to be associated with a higher
293 recurrence risk (Figure 3A-B).³²⁻³⁶ Taken together, we observed a clear morphological
294 continuum mirroring the progression from DL low to DL high-risk tumors. Moreover, we
295 analysed the distribution of the DL risk score with clinically relevant molecular information
296 namely MSI status, *BRAF* and *RAS* mutational status (Figure 3 C-F). We found that the
297 distribution was very similar for all these factors, suggesting that our DL model
298 independently detects and accounts for additional prognostically relevant morphological
299 features.

300 In summary, although our histopathology DL model was trained without human annotations,
301 and solely on non-processed WSIs, we found that the model learned to pay attention to
302 regions linked to tumor biological features plausibly associated with prognosis, thereby
303 synergizing with ctDNA. Moreover, our findings are consistent with previous DL-based end-
304 to-end prognostication approaches in CRC based on H&E histopathology alone^{22,37}.

305 **Discussion**

306 CRC can often be cured through surgery, but a subset of patients experience relapse, which
307 is associated with high mortality. To mitigate this risk, ACT is administered to locally
308 advanced CRC patients post-surgery. However, the majority of these patients do not benefit
309 from such treatment, which is associated with substantial side effects³⁸. Decades of research
310 have focused on identifying potential biomarkers to administer ACT selectively to high-risk
311 individuals who would benefit the most, while withholding it from low-risk individuals. To
312 date, one of the most promising biomarkers for this purpose is ctDNA. Measurement of MRD
313 through ctDNA is non-invasive, robust, and highly prognostic. However, ctDNA does not
314 capture the tumor's interaction with its microenvironment—the complex spatial ecology of
315 tumors³⁹-nor the tumor morphology itself. This is a limitation of ctDNA as a biomarker given
316 that, in addition to conventional histopathology tumor features, the interplay between tumors
317 and their microenvironment has been demonstrated to be highly prognostic and predictive
318 over the years. In our study, we demonstrate that combining DL-based risk assessment with
319 MRD measurement further enhances prognostic capabilities: MRD-negative patients who
320 were predicted to be at high-risk for relapse by our DL model had a significantly longer DFS
321 if treated with ACT, whereas in MRD-negative patients with a DL-based low-risk status no
322 DFS benefit was seen for those receiving ACT (Figure 2F). These observations suggest that
323 healthcare providers may identify a subset of patients who are at risk for relapse but are not
324 detected through current diagnostic tools, including a diagnostic as innovative as ctDNA.
325 Previous studies developing DL-based prognostication systems failed to provide evidence
326 for potentially different chemotherapy efficacy across DL categories, by which all potential
327 therapeutic implications of these models remain speculative^{23,40}

328 To our knowledge, our study provides the first evidence suggesting that a DL risk
329 assessment algorithm may indicate therapy efficacy in a real world setting in CRC. This
330 combined approach may improve patient selection, suggesting a way how ACT could be
331 restricted to those patients who are most likely to benefit from it. Furthermore, our DL
332 method is using the latest state-of-the-art models, is fully open source and can be reused
333 and adapted by anyone.

334 **Limitations**

335 A limitation of our study is that integrating our insights into clinical routine requires further
336 evaluation in additional cohorts, ideally in a prospective manner. Despite this, our study,
337 encompassing thousands of patients across different ethnicities, represents one of the
338 largest studies in this field. Moreover, we utilized a state-of-the-art foundation model for

339 digital pathology analysis, UNI³⁰. This is particularly relevant for clinical translatability, as the
340 capabilities of foundation models are rapidly advancing, suggesting that further performance
341 gains are conceivable with improved DL models. Nevertheless, medical device approval in
342 Japan, the US, and the European Union requires a static piece of software that cannot be
343 easily updated. Therefore, like any other DL-based biomarker, our method may be outdated
344 by the time of clinical approval. We urge regulators and policymakers to work towards
345 enabling the update of DL-based biomarkers with the latest technologies.

346 **Conclusion**

347 Despite these limitations, our data show that the excellent prognostic performance of ctDNA
348 in CRC can be further improved by DL-based end-to-end assessment of routine pathology
349 slides. After prospective validation, this approach provides a plausible and comprehensive
350 strategy for relapse risk assessment with potential therapeutic implications.

351 **Funding**

352 JNK is supported by the German Cancer Aid (DECADE, 70115166), the German Federal
353 Ministry of Education and Research (PEARL, 01KD2104C; CAMINO, 01EO2101; SWAG,
354 01KD2215A; TRANSFORM LIVER, 031L0312A; TANGERINE, 01KT2302 through ERA-NET
355 Transcan), the German Academic Exchange Service (SECAI, 57616814), the German
356 Federal Joint Committee (TransplantKI, 01VSF21048) the European Union's Horizon Europe
357 and innovation programme (ODELIA, 101057091; GENIAL, 101096312), the European
358 Research Council (ERC; NADIR, 101114631) and the National Institute for Health and Care
359 Research (NIHR, NIHR203331) Leeds Biomedical Research Centre. The views expressed
360 are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department
361 of Health and Social Care. This work was funded by the European Union. Views and
362 opinions expressed are however those of the author(s) only and do not necessarily reflect
363 those of the European Union. Neither the European Union nor the granting authority can be
364 held responsible for them. CIRCULATE-Japan receives financial support from the Japan
365 Agency for Medical Research and Development (grant 19ck0106447h0002-TY). SF is
366 supported by the German Federal Ministry of Education and Research (SWAG,
367 01KD2215A), the German Cancer Aid (DECADE, 70115166 and TargHet, 70115995) and
368 the German Research Foundation (504101714). The DACHS study (HB, TY, DW and MH)
369 was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4,
370 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
371 2354/3-2, RO 2270/8-1, RO 2270/8-2, BR 1704/17-1 and BR 1704/17-2), the
372 Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT;

373 Germany) and the German Federal Ministry of Education and Research (01KH0404,
374 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

375

376 **Disclosures**

377 **CMLL** reports honoraria from AstraZeneca. **HB** reports research funding from Ono
378 Pharmaceutical and honoraria from Ono Pharmaceutical, Eli Lilly Japan, and Taiho
379 Pharmaceutical. **TM** reports honoraria from Chugai, AstraZeneca, and Miyarisan. **SM** reports
380 honoraria from Taiho Pharmaceutical Co., Ltd., Chugai Pharmaceutical Co., Ltd., and Eli
381 Lilly CO, Ltd. **DK** reports honoraria from Takeda, Chugai, Lilly, MSD, Ono, Seagen,
382 Guardant Health, Eisai, Taiho, Bristol Myers Squibb, Daiichi-Sankyo, Pfizer,
383 Merckbiopharma, and Sysmex: research funding from Ono, MSD, Novartis, Servier,
384 Janssen, IQVIA, Syneoshealth, CIMIC, and Cimicshiftzero. **HT** reports speakers' bureau
385 from MSD K.K, Merck Biopharma, Takeda, Taiho, Lilly Japan, Bristol-Myers Squibb Japan,
386 Chugai Pharmaceutical, Ono Yakuhin, Amgen; research funding from Takeda, Daiichi
387 Sankyo. **IT** reports speakers' bureau from Medtronic, Johnson & Johnson, Intuitive,
388 Medcaroid, Eli Lilly and research funding from Medtronic, sysmex. **SF** has received
389 honoraria from MSD and BMS. **TK** reports nothing to declare. **EO** reports speakers' bureau
390 from Chugai Pharmaceutical Co., Ltd., Bristol Meyers, Ono Pharmaceutical Co., Ltd., Eli
391 Lilly, Takeda Pharmaceutical Co., Ltd.; research funding from Guardant Health, Inc.;
392 advisory role from Glaxosmithkline plc. **YN** reports advisory role from Guardant Health Pte
393 Ltd., Natera, Inc., Roche Ltd., Seagen, Inc., Premo Partners, Inc., Daiichi Sankyo Co., Ltd.,
394 Takeda Pharmaceutical Co., Ltd., Exact Sciences Corporation, and Gilead Sciences, Inc.;
395 speakers' bureau from Guardant Health Pte Ltd., MSD K.K., Eisai Co., Ltd., Zeria
396 Pharmaceutical Co., Ltd., Miyarisan Pharmaceutical Co., Ltd., Merck Biopharma Co., Ltd.,
397 CareNet, Inc., Hisamitsu Pharmaceutical Co., Inc., Taiho Pharmaceutical Co., Ltd., Daiichi
398 Sankyo Co., Ltd., Chugai Pharmaceutical Co., Ltd., and Becton, Dickinson and Company,
399 Guardant Health Japan Corp; research funding from Seagen ,Inc., Genomedica Inc.,
400 Guardant Health AMEA, Inc., Guardant Health, Inc., Tempus Labs, Inc., Roche Diagnostics
401 K.K., Daiichi Sankyo Co., Ltd., and Chugai Pharmaceutical Co., Ltd.. **TY** reports honoraria
402 from Taiho, Chugai, Eli Lilly, Merck, Bayer Yakuhin, Ono and MSD, and research funding
403 from Ono, Sanofi, Daiichi Sankyo, Parexel, Pfizer, Taiho, MSD, Amgen, Genomedica,
404 Sysmex, Chugai and Nippon Boehringer Ingelheim. S.S. The remaining authors declare no
405 competing interests. **JNK** declares consulting services for Owkin, France; DoMore
406 Diagnostics, Norway; Panakeia, UK; Scailyte, Switzerland; Mindpeak, Germany; and
407 MultiplexDx, Slovakia. Furthermore he holds shares in StratifAI GmbH, Germany, has

408 received a research grant by GSK, and has received honoraria by AstraZeneca, Bayer,
409 Eisai, Janssen, MSD, BMS, Roche, Pfizer and Fresenius. All the other authors report
410 nothing to declare.

411 **Ethics statement**

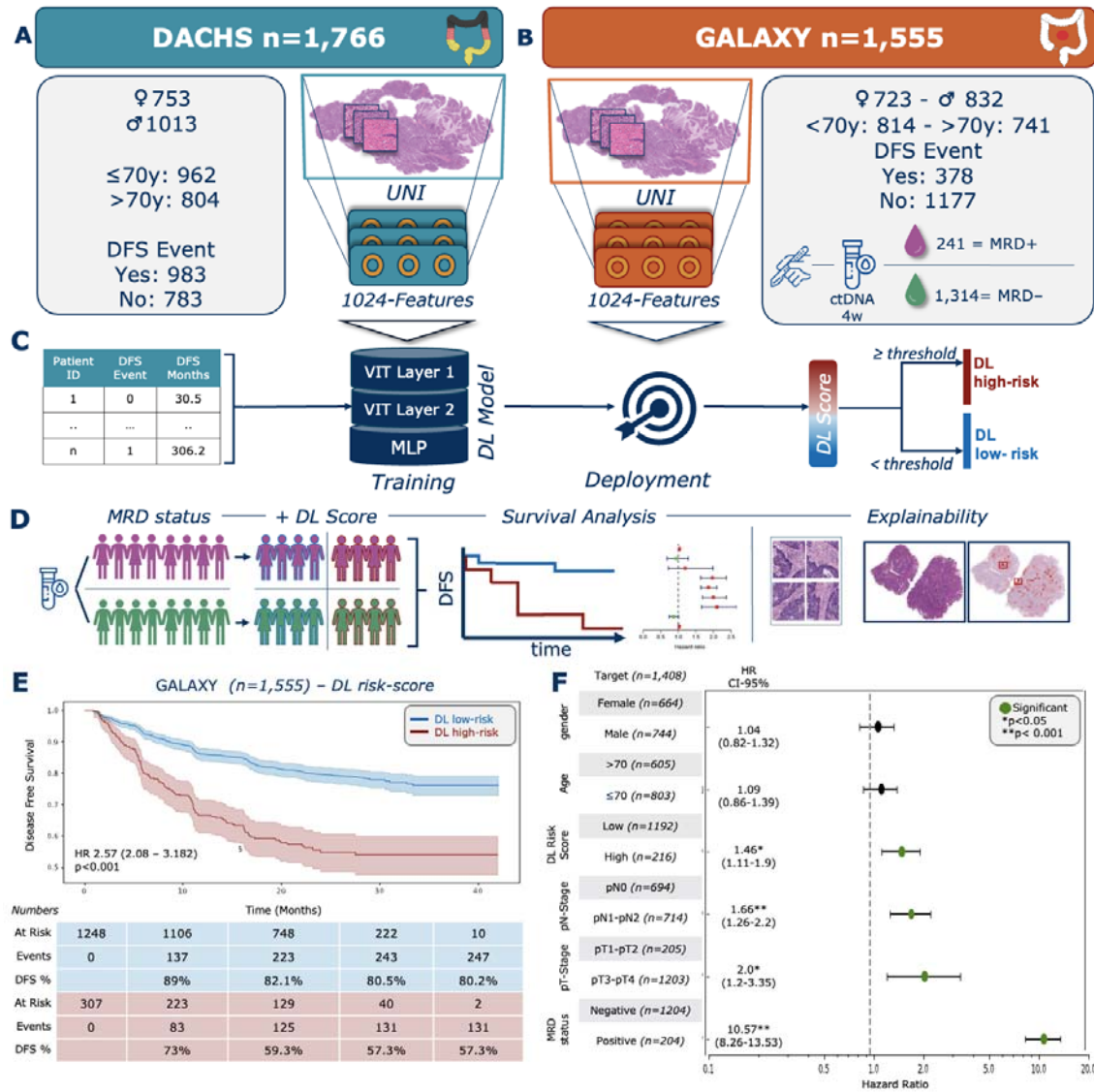
412 The experiments in this study were carried out according to the Declaration of Helsinki and
413 the International Ethical Guidelines for Biomedical Research Involving Human Subjects by
414 the Council for International Organizations of Medical Sciences (CIOMS). The present study
415 also adheres to the “Transparent reporting of a multivariable prediction model for individual
416 prognosis or diagnosis” (TRIPOD) statement.²⁰ The Ethics Board at the Medical Faculty of
417 Technical University Dresden (BO-EK-444102022) and Institutional Review Board of the
418 National Cancer Center Japan (2023-207) approved of the overall analysis in this study. The
419 patient sample collection in each cohort was separately approved by the respective
420 institutional ethics board.

421 **Author contributions**

422 CMLL, HB, JNK and TY conceptualised the study. HD, TN, TM, SM, DK, HT, IT, TK, EO
423 provided clinical and scanned whole slide image data for the GALAXY cohort. TY, DW, MH,
424 HB provided clinical and scanned whole slide image data for the DACHS cohort. CMLL
425 curated the source data. SS, XJ, MvT implemented the deep learning algorithm. SS
426 developed the code for data analysis and visualisation. CMLL and SS planned and
427 conducted the experiments. CMLL interpreted the data. HSM, ZIC, JNK assisted with the
428 interpretation of results. NR and SF did the pathological interpretation of the results. CMLL
429 wrote the first draft of the manuscript. All authors revised the manuscript draft, contributed to
430 the interpretation of the data and agreed to the submission of this paper.

431

432 **Figures & Tables**



433

434 **Figure 1: Study Design and DL risk stratification overall**

435 (A) DACHS cohort overview including patient characteristics and WSI preprocessing pipeline
 436 using UNI a pretrained vision encoder for feature extraction. (B) GALAXY cohort overview
 437 including patient characteristics and WSI preprocessing pipeline. (C) Flowchart of the study
 438 design: DFS data was analysed using a Cox-Regression model and fed into the DL-Model
 439 combined with the image features from the DACHS cohort for training. The DL-Model was
 440 then deployed onto the GALAXY features and a DL-Score was obtained. (D) Overview
 441 Experimental Setup: Patients were first categorised based on MRD status and then sub-
 442 categorized according to the DL score. Survival analysis with Kaplan-Meier estimator and
 443 Cox proportional hazard models were performed. Lastly, highly predictive Tiles and patient

444 whole slide heatmaps were generated. (E) Kaplan-Meier curves for DFS stratified by DL
 445 high-risk and DL low-risk patients. (F) Forest plot showing multivariate cox regression
 446 analysis including the covariates gender, age, DL risk score, pathological Nodal Stage (pN-
 447 Stage), pathological Tumor Stage (pT-Stage) and MRD-status and their association with
 448 DFS. HR and 95% CI were calculated by the Cox proportional hazard model. *P*-value was
 449 calculated using the two-sided log-rank test (**p*<0.05, ** *p*<0.001). Plot were generated
 450 using lifelines package in Python 3.11.5
 451 DACHS=Darmkrebs: Chancen der Verhütung durch Screening Study, WSI=whole-slide
 452 image, DFS=disease-free survival, DL=Deep Learning, MRD=molecular residual disease,
 453 HR=Hazard ratio, CI=Confidence interval

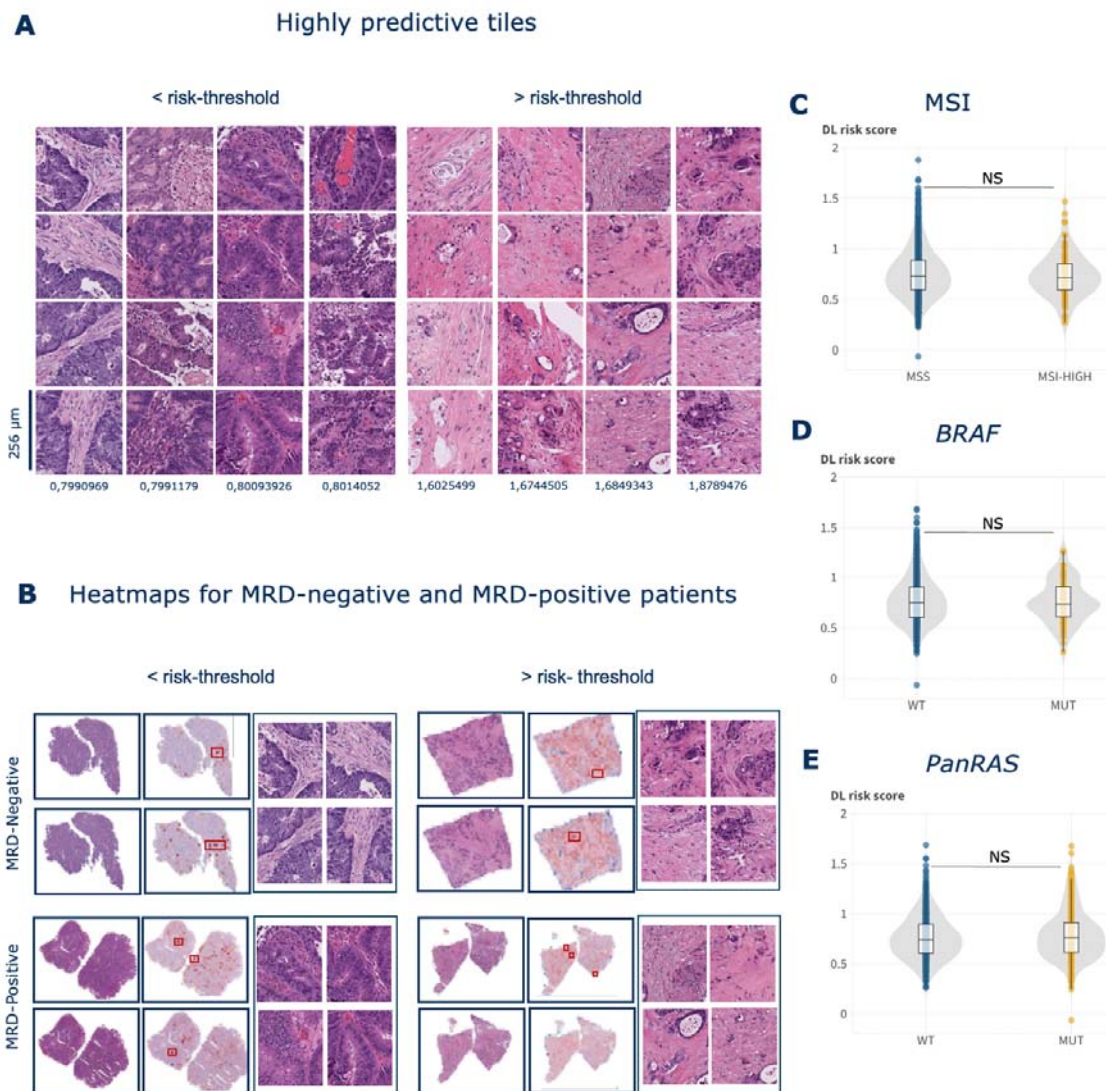


454

455

456 **Figure 2: DL stratifies recurrence risk within MRD subgroups**

457 Kaplan-Meier curves showing DFS stratification by DL high-risk and DL low-risk groups for
458 (A) MRD-positive and (B) MRD-negative groups, followed by Kaplan-Meier curves showing
459 DFS stratified by with or without ACT treatment in (C) MRD-positive and DL low-risk, (D)
460 MRD-negative and DL low-risk, (E) MRD-positive and DL high-risk and (F) MRD-negative
461 and DL low-risk subgroups. HR and 95% CI were calculated by the Cox proportional hazard
462 model. *P*-value was calculated using the two-sided log-rank test. Plots were generated using
463 the lifelines package in Python 3.11.5 DFS=disease-free survival, DL=Deep Learning,
464 ACT=adjuvant chemotherapy, MRD=molecular residual disease, HR=Hazard ratio,
465 CI=Confidence interval.
466



467

468 **Figure 3: Morphological and molecular features of the DL risk score**

469 (A) Highly predictive tiles for patients below the DL risk-threshold and above the DL risk-
 470 threshold exemplarily with DL score reported. (B) Whole slide patient heatmaps showing the
 471 DL prediction score, red indicating high-risk, and blue indicating low-risk. Box plot showing
 472 distribution of DL risk score among (C) MSI status (D) *BRAF-V600E* mutational status and
 473 (E) *PanRAS* mutational status. *P*-Value calculated using Kruskal-Wallis test. Figure was
 474 created using Flourish (<https://flourish.studio/>).

475 DL= Deep Learning, MRD= molecular residual disease, WT= wildtype, MUT=mutation,
 476 MSS=microsatellite stable, MSI= microsatellite instability, NS=not significant

477

478

Patient characteristics	Category	DL high-risk (n=307□), n (%)	DL low-risk (n=1248), n (%)	Chi squared and p values
Age	<=70	167 (54.4)	647 (51.8)	X ² = 0.54618 P-value = 0.4599
	>70	140 (45.6)	601 (48.2)	
Sex	Female	120 (39.1)	603 (48.3)	X ² = 8.0696 P-value = 0.004501
	Male	187 (60.9)	645 (51.7)	
ECOG performance Status	0	280 (91.2)	1124 (90)	X ² = 0.24735 P-value = 0.6189
	1	27 (8.8)	124 (10)	
pT-Stage	T1-T2	10 (3.3)	195 (15.6)	X ² = 19.26 P-value<0.001
	T3-T4	206 (67.1)	998 (80)	
	NA	91 (29.6)	55 (4.4)	
pN-Stage	N0	76 (35.2)	618 (51.8)	X ² = 19.646 P-value<0.001
	N1-2	140 (64.8)	574 (48.2)	
	NA	91 (29.6)	56 (4.5)	
pathological Stage	I	3 (1)	148 (11.9)	X ² = 201.46 P-value<0.001
	II	65 (21.1)	452 (36.2)	
	III	116 (37.8)	533 (42.7)	
	IV	123 (40.1)	115 (9.2)	
RAS status	RAS wild-type	102 (54.8)	394 ()	X ² = 0.32606 P-value=0.568
	RAS mutant	84 (45.2)	291 ()	

	NA	121 (39.4)	563 (45.1)	
BRAF status	<i>BRAF</i> wild-type	171 (55.7)	623 (49.9)	$X^2 = 1.2724e-28$ $P\text{-value}=1$
	<i>BRAF</i> mutant	13 (4.2)	48 (3.8)	
	NA	123 (40.1)	577 (46.2)	
MSI status	MSI	20 (6.5)	102 (8.2)	$X^2 = 0.46606$ $P\text{-value}=0.4948$
	MSS	262 (85.3)	1090 (87.3)	
	NA	25 (8.1)	56 (4.5)	
MRD Status	MRD-positive	81 (26.4)	160 (12.8)	$X^2 = 33.585$ $P\text{-value} < 0.001$
	MRD-negative	226 (73.6)	1088 (87.2)	

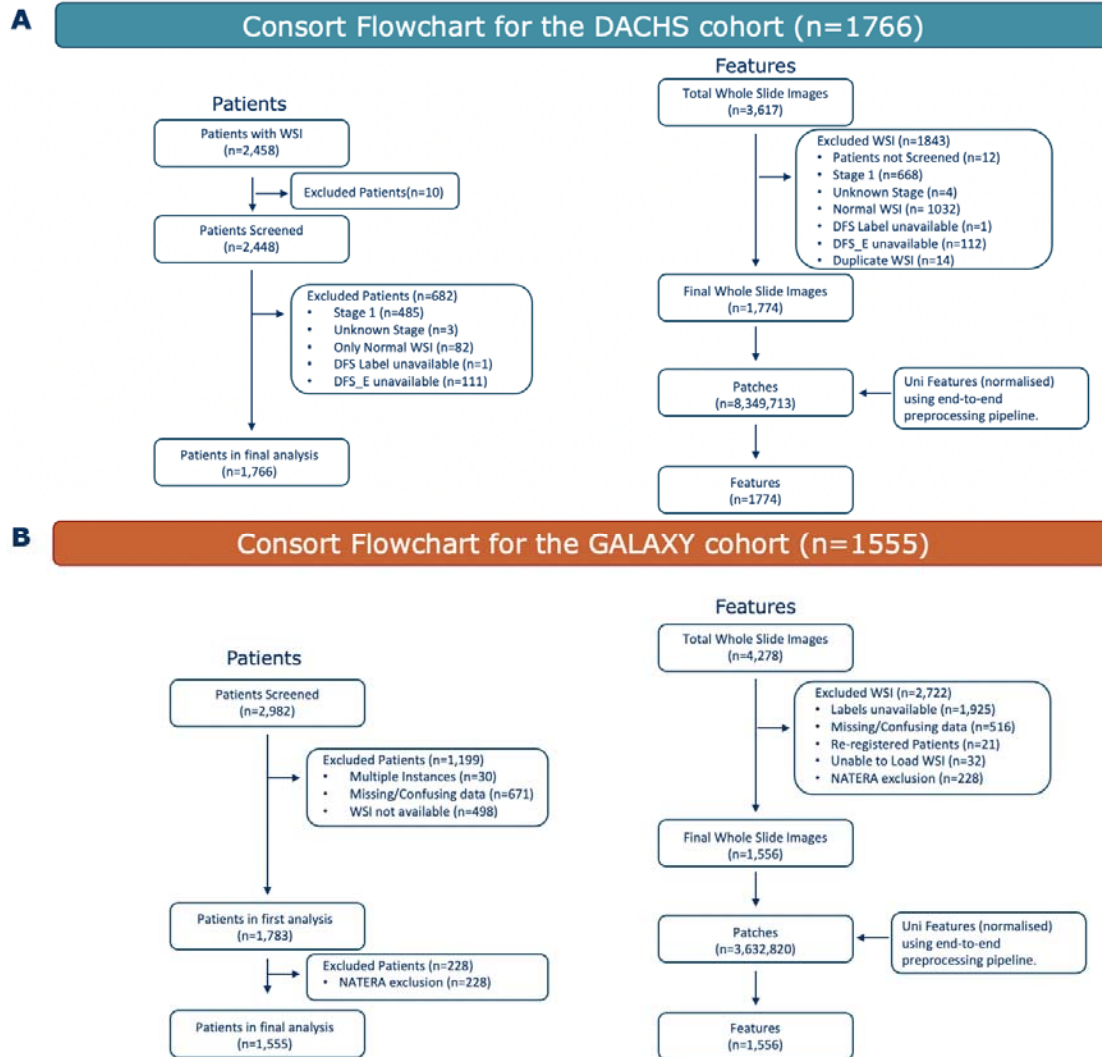
479

480 **Table 1 Patients characteristic for DL high-risk and low-risk patients**

481 P values were obtained by a pearsons chi-squared test with Yates' continuity correction
 482 comparing the distribution of the factors between the two columns (DL high-risk vs DL low-
 483 risk). Statistical analysis was performed on R 4.4.4. ECOG=Eastern Cooperativ Oncology
 484 Group, MSS=microsatellite stable, MSI=microsatellite instable, NA=Not available

485

486 **Supplementary Figures and Tables**



487

488

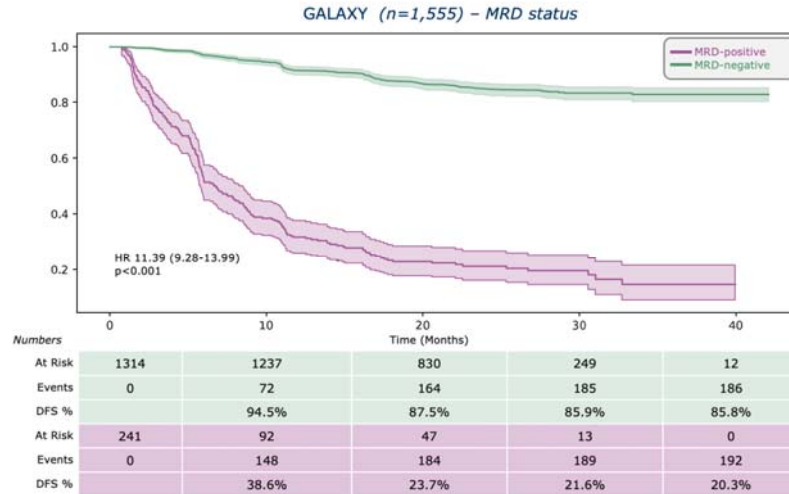
489 **Supplementary Figure 1: Consort diagram for both cohorts**

490 Flowchart showing initial screened patients and WSIs and exclusion criterias for (A) the
 491 DACHS cohort and (B) the GALAXY cohort.

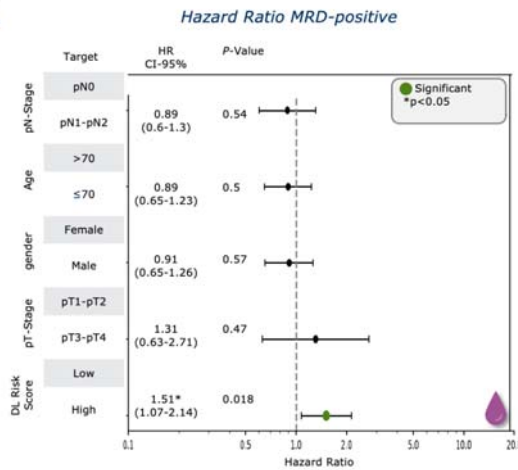
492 WSIs= whole slide images, DACHS=Darmkrebs: Chancen der Verhütung durch Screening

493 Study

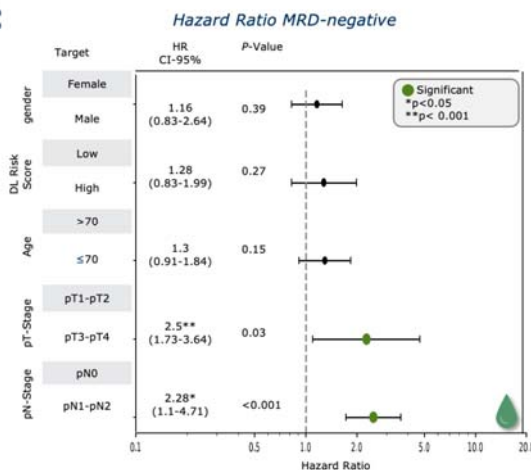
A



B



C



494

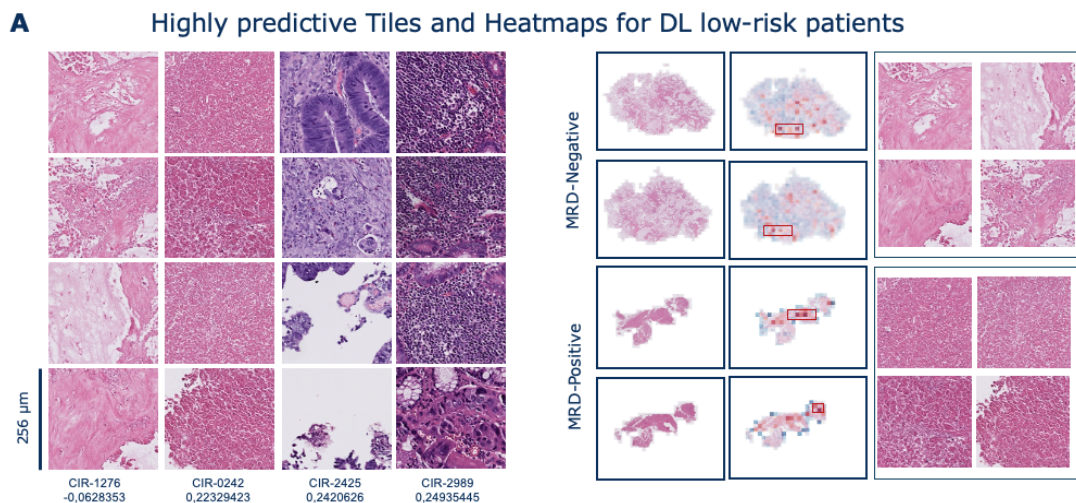
495 **Supplementary Figure 2: MRD status is predictive of survival outcomes and**
 496 **Multivariate analysis for MRD-subgroups**

497 (A) Kaplan-Meier curves for DFS stratified by MRD-positive and MRD-negative patients.
 498 Forest plot showing multivariate cox regression analysis for (B) MRD-positive and (C) MRD-
 499 negative subgroup including the covariates gender, age, DL risk score, pathological Nodal
 500 Stage (pN-Stage), pathological Tumor Stage (pT-Stage) and their association with DFS. HR
 501 and 95% CI were calculated by the Cox proportional hazard model. P-value was calculated
 502 using the two-sided log-rank test (*p<0.05, ** p<0.001). Plot were generated using lifelines
 503 package in Python 3.11.5

504 DFS=disease-free survival, DL=Deep Learning, MRD=molecular residual disease,
 505 HR=Hazard ratio, CI=Confidence interval.

506

507



508

509

Supplementary Figure 3: Morphological and molecular features for DL low-risk score

510

(A) Highly predictive tiles and Whole slide heatmaps for patients with lowest DL risk score.

511

Red indicating high-risk, and blue indicating low-risk. DL= Deep Learning, MRD= molecular

512

residual disease.

513

514

Declaration of generative AI and AI-assisted technologies in the writing

515

process

516

During the preparation of this study the author(s) used GPT-4 for grammar and spelling

517

checking. After using this tool/service, the author(s) reviewed and edited the content as

518

needed and take(s) full responsibility for the content of the publication.

519 References

- 520 1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*. January 17,
521 2024;74(1):12–49.
- 522 2. Boute TC, Swartjes H, Greuter MJE, Elferink MAG, van Eekelen R, Vink GR, et al. Cumulative
523 Incidence, Risk Factors, and Overall Survival of Disease Recurrence after Curative Resection of
524 Stage II-III Colorectal Cancer: A Population-based Study. *Cancer Res Commun*. February 29,
525 2024;4(2):607–616.
- 526 3. Nors J, Iversen LH, Erichsen R, Gotschalck KA, Andersen CL. Incidence of Recurrence and Time
527 to Recurrence in Stage I to III Colorectal Cancer: A Nationwide Danish Cohort Study. *JAMA*
528 *Oncol*. January 1, 2024;10(1):54–62.
- 529 4. Liemburg GB, Brandenburg D, Berger MY, Duijts SFA, Holtman GA, de Bock GH, et al.
530 Diagnostic accuracy of follow-up tests for detecting colorectal cancer recurrences in primary care:
531 A systematic review and meta-analysis. *Eur J Cancer Care*. September 2021;30(5):e13432.
- 532 5. Nicholson BD, Shinkins B, Pathiraja I, Roberts NW, James TJ, Mallett S, et al. Blood CEA levels
533 for detecting recurrent colorectal cancer. *Cochrane Database Syst Rev*. December 10,
534 2015;2015(12):CD011134.
- 535 6. Shinkins B, Nicholson BD, James T, Pathiraja I, Pugh S, Perera R, et al. What carcinoembryonic
536 antigen level should trigger further investigation during colorectal cancer follow-up? A systematic
537 review and secondary analysis of a randomised controlled trial. *Health Technol Assess*. April
538 2017;21(22):1–60.
- 539 7. O'Connor ES, Greenblatt DY, LoConte NK, Gangnon RE, Liou J-I, Heise CP, et al. Adjuvant
540 chemotherapy for stage II colon cancer with poor prognostic features. *J Clin Oncol*. September 1,
541 2011;29(25):3381–3388.
- 542 8. Baxter NN, Kennedy EB, Bergsland E, Berlin J, George TJ, Gill S, et al. Adjuvant Therapy for
543 Stage II Colon Cancer: ASCO Guideline Update. *J Clin Oncol*. March 10, 2022;40(8):892–910.
- 544 9. Moding EJ, Nabet BY, Alizadeh AA, Diehn M. Detecting Liquid Remnants of Solid Tumors:
545 Circulating Tumor DNA Minimal Residual Disease. *Cancer Discov*. December 1,
546 2021;11(12):2968–2986.
- 547 10. Kasi PM, Fehringer G, Taniguchi H, Starling N, Nakamura Y, Kotani D, et al. Impact of Circulating
548 Tumor DNA-Based Detection of Molecular Residual Disease on the Conduct and Design of
549 Clinical Trials for Solid Tumors. *JCO Precis Oncol*. March 2022;6:e2100181.
- 550 11. Tie J, Cohen JD, Wang Y, Li L, Christie M, Simons K, et al. Serial circulating tumour DNA
551 analysis during multimodality treatment of locally advanced rectal cancer: a prospective
552 biomarker study. *Gut*. April 2019;68(4):663–671.
- 553 12. Kotani D, Oki E, Nakamura Y, Yukami H, Mishima S, Bando H, et al. Molecular residual disease
554 and efficacy of adjuvant chemotherapy in patients with colorectal cancer. *Nat Med*. January
555 2023;29(1):127–134.
- 556 13. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, et al. Type,
557 density, and location of immune cells within human colorectal tumors predict clinical outcome.
558 *Science*. September 29, 2006;313(5795):1960–1964.
- 559 14. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou F-S, Bifulco C, et al. International validation of the
560 consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study.
561 *Lancet*. May 26, 2018;391(10135):2128–2139.
- 562 15. Jesinghaus M, Schmitt M, Lang C, Reiser M, Scheiter A, Konukiewitz B, et al. Morphology
563 Matters: A Critical Reappraisal of the Clinical Relevance of Morphologic Criteria From the 2019
564 WHO Classification in a Large Colorectal Cancer Cohort Comprising 1004 Cases. *Am J Surg*

- 565 *Pathol.* July 1, 2021;45(7):969–978.
- 566 16. Wankhede D, Yuan T, Kloor M, Halama N, Brenner H, Hoffmeister M. Clinical significance of
567 combined tumour-infiltrating lymphocytes and microsatellite instability status in colorectal cancer:
568 a systematic review and network meta-analysis. *Lancet Gastroenterol Hepatol.* July
569 2024;9(7):609–619.
- 570 17. Argilés G, Tabernero J, Labianca R, Hochhauser D, Salazar R, Iveson T, et al. Localised colon
571 cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol.*
572 October 1, 2020;31(10):1291–1305.
- 573 18. Roth AD, Delorenzi M, Tejpar S, Yan P, Klingbiel D, Fiocca R, et al. Integrated analysis of
574 molecular and clinical prognostic factors in stage II/III colon cancer. *J Natl Cancer Inst.* November
575 7, 2012;104(21):1635–1646.
- 576 19. Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology:
577 enhancing cancer research and clinical oncology. *Nat Cancer.* September 2022;3(9):1026–1038.
- 578 20. Perez-Lopez R, Ghaffari Laleh N, Mahmood F, Kather JN. A guide to artificial intelligence for
579 cancer researchers. *Nat Rev Cancer.* May 16, 2024; Available at:
580 <http://dx.doi.org/10.1038/s41568-024-00694-7>
- 581 21. Muti HS, Röcken C, Behrens H-M, Löffler CML, Reitsam NG, Grosser B, et al. Deep learning
582 trained on lymph node status predicts outcome from gastric cancer histopathology: a
583 retrospective multicentric study. *Eur J Cancer.* November 2023;194:113335.
- 584 22. Jiang X, Hoffmeister M, Brenner H, Muti HS, Yuan T, Foersch S, et al. End-to-end
585 prognostication in colorectal cancer by deep learning: a retrospective, multicentre study. *Lancet*
586 *Digit Health.* January 2024;6(1):e33–e43.
- 587 23. Kleppe A, Skrede O-J, De Raedt S, Hveem TS, Askautrud HA, Jacobsen JE, et al. A clinical
588 decision support system optimising adjuvant chemotherapy for colorectal cancers by integrating
589 deep learning and pathological staging markers: a development and validation study. *Lancet*
590 *Oncol.* September 2022;23(9):1221–1232.
- 591 24. Wagner SJ, Reisenbüchler D, West NP, Niehues JM, Zhu J, Foersch S, et al. Transformer-based
592 biomarker prediction from colorectal cancer histology: A large-scale multicentric study. *Cancer*
593 *Cell.* September 11, 2023;41(9):1650–1661.e4.
- 594 25. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can
595 predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med.* July
596 2019;25(7):1054–1056.
- 597 26. Gustav M, Reitsam NG, Carrero ZI, Loeffler CML, van Treeck M, Yuan T, et al. Deep learning for
598 dual detection of microsatellite instability and POLE mutations in colorectal cancer
599 histopathology. *NPJ Precis Oncol.* May 23, 2024;8(1):115.
- 600 27. Foersch S, Glasner C, Woerl A-C, Eckstein M, Wagner D-C, Schulz S, et al. Multistain deep
601 learning for prediction of prognosis and therapy response in colorectal cancer. *Nat Med.* January
602 9, 2023; Available at: <http://dx.doi.org/10.1038/s41591-022-02134-1>
- 603 28. Loupakis F, Sharma S, Derouazi M, Murgioni S, Biondi P, Rizzato MD, et al. Detection of
604 Molecular Residual Disease Using Personalized Circulating Tumor DNA Assay in Patients With
605 Colorectal Cancer Undergoing Resection of Metastases. *JCO Precis Oncol.* July 2021;5.
606 Available at: <http://dx.doi.org/10.1200/PO.21.00101>
- 607 29. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al. A method for
608 normalizing histology slides for quantitative analysis. In 2009 IEEE International Symposium on
609 Biomedical Imaging: From Nano to Macro 2009; 1107–1110.
- 610 30. Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, et al. Towards a general-purpose
611 foundation model for computational pathology. *Nat Med.* March 2024;30(3):850–862.

- 612 31. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized
613 treatment recommender system using a Cox proportional hazards deep neural network. *BMC*
614 *Med Res Methodol*. February 26, 2018;18(1):24.
- 615 32. Ueno H, Ishiguro M, Nakatani E, Ishikawa T, Uetake H, Murotani K, et al. Prognostic value of
616 desmoplastic reaction characterisation in stage II colon cancer: prospective validation in a Phase
617 3 study (SACURA Trial). *Br J Cancer*. March 2021;124(6):1088–1097.
- 618 33. Lugli A, Karamitopoulou E, Zlobec I. Tumour budding: a promising parameter in colorectal
619 cancer. *Br J Cancer*. May 22, 2012;106(11):1713–1717.
- 620 34. Platten M, Bunse L, Wick A, Bunse T, Le Cornet L, Harting I, et al. A vaccine targeting mutant
621 IDH1 in newly diagnosed glioma. *Nature*. April 2021;592(7854):463–468.
- 622 35. Huijbers A, Tollenaar RAEM, Pelt GW v, Zeestraten ECM, Dutton S, McConkey CC, et al. The
623 proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients:
624 validation in the VICTOR trial. *Ann Oncol*. January 2013;24(1):179–185.
- 625 36. Hu Q, Wang Y, Yao S, Mao Y, Liu L, Li Z, et al. Desmoplastic Reaction Associates with
626 Prognosis and Adjuvant Chemotherapy Response in Colorectal Cancer: A Multicenter
627 Retrospective Study. *Cancer Research Communications*. June 15, 2023;3(6):1057–1066.
- 628 37. Wulczyn E, Steiner DF, Moran M, Plass M, Reihls R, Tan F, et al. Interpretable survival prediction
629 for colorectal cancer using deep learning. *NPJ Digit Med*. April 19, 2021;4(1):71.
- 630 38. Grothey Axel, Sobrero Alberto F., Shields Anthony F., Yoshino Takayuki, Paul James, Taieb
631 Julien, et al. Duration of Adjuvant Chemotherapy for Stage III Colon Cancer. *N Engl J Med*.
632 March 29, 2018;378(13):1177–1188.
- 633 39. Zaborowski AM, Winter DC, Lynch L. The therapeutic and prognostic implications of
634 immunobiology in colorectal cancer: a review. *Br J Cancer*. November 2021;125(10):1341–1349.
- 635 40. Sun C, Li B, Wei G, Qiu W, Li D, Li X, et al. Deep learning with whole slide images can improve
636 the prognostic risk stratification with stage III colorectal cancer. *Comput Methods Programs*
637 *Biomed*. June 2022;221:106914.