Machine Learning of Bone Marrow Histopathology Identifies Genetic and Clinical Determinants in Patients with MDS

Brück, Oscar E.

2021-05-01

Brück , O E , Lallukka-Brück , S E , Hohtari , H R , Ianevski , A , Ebeling , F T , Kovanen , P E , Kytölä , S I , Aittokallio , T A , Ramos , P M , Porkka , K V & Mustjoki , S M 2021 , ' Machine Learning of Bone Marrow Histopathology Identifies Genetic and Clinical Determinants in Patients with MDS ' , Cancer Discovery , vol. 2 , no. 3 , pp. 238-249 . https://doi.org/10.1158/2643-3

http://hdl.handle.net/10138/351824
https://doi.org/10.1158/2643-3230.BCD-20-0162

unspecified
acceptedVersion

Machine Learning of Bone Marrow Histopathology Identifies Genetic and Clinical Determinants in MDS Patients

Oscar E. Brück*[1,2,3,4], Susanna E. Lallukka-Brück†[1], Helena R. Hohtari †[1,2], Aleksandr Ianevski†[5], Freja T. Ebeling[4], Panu E. Kovanen[6], Soili Kytölä[7], Tero A. Aittokallio[3,5,8], Pedro M. Ramos[9], Kimmo V. Porkka[1,2,3,4], Satu M. Mustjoki*[1,2,3,10]

1 *Hematology Research Unit Helsinki, University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center, Helsinki, Finland*
2 Translational Immunology Research Program, University of Helsinki, Helsinki, Finland
3 iCAN Digital Precision Cancer Medicine Flagship, Helsinki, Finland
4 Helsinki University Hospital Comprehensive Cancer Center, Department of Hematology, Helsinki, Finland
5 Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland
6 Department of Pathology, HUSLAB, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
7 HUS Diagnostic Center, HUSLAB, Helsinki University Hospital, Helsinki, Finland
8 Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, and Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway
9 Novartis Pharmaceuticals, Basel, Switzerland
10 Department of Clinical Chemistry and Hematology, University of Helsinki, Helsinki, Finland

* Corresponding authors
† These authors contributed equally to this work.

Address for correspondence:
Drs S Mustjoki and O Brück, Hematology Research Unit Helsinki, University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center, Haartmaninkatu 8, FIN-00290 Helsinki, Finland. E-mail: satu.mustjoki@helsinki.fi, oscar.bruck@hus.fi

Running title: Histopathological Landscape of MDS
Word count: Abstract 150, Manuscript 4490
Number of Figures and tables: 4 figures; 1 table, 10 supplemental figures, 3 supplemental tables

Disclosure of Conflicts of Interest
K.P. received honoraria and research funding from Celgene, Incyte, Novartis and Bristol-Myers Squibb. S.M. received honoraria and research funding from Pfizer,

Novartis and Bristol-Myers Squibb. H.H. received research funding from Incyte. P.R.M. is employed by Novartis Pharmaceuticals.

Abstract

In myelodysplastic syndrome (MDS) and myeloproliferative neoplasm (MPN), bone marrow (BM) histopathology is assessed to identify dysplastic cellular morphology, cellularity, and blast excess. Yet, other morphological findings may elude the human eye. We used convolutional neural networks to extract morphological features from 236 MDS, 87 MDS/MPN, and 11 control BM biopsies. These features predicted genetic and cytogenetic aberrations, prognosis, age, and gender in multivariate regression models. Highest prediction accuracy was found for *TET2* (area under the receiver-operating curve [AUROC] 0.94) and spliceosome mutations (0.89) and chromosome 7 monosomy (0.89). Mutation prediction probability correlated with variant allele frequency and number of affected genes per pathway, demonstrating the algorithms' ability to identify relevant morphological patterns. By converting regression models to texture and cellular composition, we reproduced the classical del(5q) MDS morphology consisting of hypolobulated megakaryocytes. In summary, this study highlights the potential of linking deep BM histopathology with genetics and clinical variables.

Statement of significance

Histopathology is elementary in the diagnostics of MDS patients, but its high-dimensional data are underused. By elucidating the association of morphological features with clinical variables and molecular genetics, this study highlights the vast potential of convolutional neural networks in understanding MDS pathology and how genetics is reflected in BM morphology.

Introduction

Current diagnosis of myelodysplastic syndrome (MDS) is based on identifying cellular dysplasia by visual inspection of bone marrow (BM) aspirate or biopsy (1). Karyotype, blast proportion, and peripheral blood (PB) cell count are assessed for disease subclassification according to WHO guidelines and for risk stratification by the Revised International Prognostic Scoring System (IPSS-R) criteria (1,2).

Deep learning enables accurate visual pattern recognition with convolutional neural networks (CNNs), where multiple processing layers detect and decode image data into numerical features (3). CNNs have recently led to significant breakthroughs in the analysis of biomedical images, facilitating diagnosis of skin tumors, retinal disease, intracranial hemorrhage, and breast cancer (3–6). In the context of routine hematoxylin and eosin (H&E) tissue stains, similar algorithms have improved Gleason scoring in prostate cancer, outcome prediction in colorectal cancer, and even discrimination of solid cancer patients by driver mutation status (4,7–9).

Here, we investigate the potential of CNN-based morphological analysis in hematology. To improve our understanding of MDS histopathology and its association with clinical factors, we predict diagnosis, prognosis, IPSS-R risk score, mutated genes, cytogenetics, and patient age and gender by utilizing solely BM morphological features (study flow presented in Supplementary Fig. 1). We demonstrate the highest detection accuracy for point mutations, such as *TET2* and *ASXL1*. The mutation prediction probability correlates with variant allele frequency (VAF) within the sample, confirming the identification of mutation-specific features. To study interactions between disease determinants and BM histology, we introduce a novel multidimensional image analysis approach that combines information at tile,

segmented nucleated hematopoietic cell (NHC) and pixel levels, ultimately facilitating the interpretation of complex BM histopathological patterns.

Results

Unsupervised modeling of BM texture recognized morphologic lineages and distinct myelodysplastic clusters

To dissect BM morphology, we extracted ImageNet-configured visual features of H&E-stained BM biopsies from patients with MDS (236 samples from 143 subjects), MDS-MPN (87 samples from 51 subjects), and healthy controls (11 samples from 11 subjects) using VGG16 and Xception CNNs (Fig. 1A, Supplementary Fig. 1). Images of tissue microarray (TMA) cores were grayscaled and split into 500 tiles enabling the examination of detailed tissue patterns. To investigate the spectrum of CNN texture patterns, we mapped image tiles from diagnostic BM samples of MDS, MDS/MPN, and control subjects with two-dimensional UMAP representation (Fig. 1B). Unsupervised image segregation was principally driven by stromal and cellular texture. Images with cellular content were observed to subcluster according to NHC and RBC abundance and lipid droplet density. As up to 60% of MDS patients may present with a hypercellular BM phenotype at diagnosis, complete discrimination of MDS and MDS/MPN patients with an unsupervised approach is unlikely (10). Here, MDS/MPN patients harbored an increased number of hypercellular tiles, while MDS patients demonstrated heterogeneous histopathological phenotypes (Fig. 1B). Notably, tiles from healthy subjects represented a balanced cellular and lipid droplet composition, with scarce proportion of stroma.

By clustering image tiles with PhenoGraph, we observed distinct subgroups of tissue texture. In turn, we hypothesized that these patterns could be interconnected into morphological entities (Fig. 1B). Slingshot lineage analysis is commonly-used in single-cell RNA sequencing data analysis to identify cellular development trajectories (11). The method aims to identify evolutionary branches from a common starting point using minimum spanning tree-based clustering (11). Here, we repurposed slingshot analysis for image data and demonstrate that tiles with high proportion of either lipid droplets, hypercellularity, or RBCs were connected to tiles of hypocellular tissue (Fig. 1B). These connections suggest that adipose, hypercellular, and RBC-rich tissue areas rarely transform from one to another but instead are more likely to arise from areas of hypocellular texture.

Tile features from MDS, MDS/MPN, and healthy subjects were averaged by adding the features of all tiles from one TMA spot and dividing by the number of tiles. Aggregated features were 2D-projected with UMAP (Fig. 1C and http://hruh-20.it.helsinki.fi/mds_visualization). Unsupervised k-means clustering was used for subgrouping aggregated features, and each cluster was analyzed for associations with clinical variables. A distinct cluster representing healthy subjects (Cluster 1) was clearly distinguishable and homogenous in contrast to the four myelodysplastic subgroups (Fig. 1C, Supplementary Fig. 2A). It also confirmed that the selected transfer learning approach identifies relevant biological features, even though these have been designed with the non-histological ImageNet image dataset (Fig. 1D). Clustering was principally driven by tissue texture content, which included the proportion of NHC, stroma, and lipid droplets and RBC and nuclear metrics of NHCs (Fig. 1D-G).

Cluster 3, defined histologically by high bone stroma content, was enriched with the World Health Organization (WHO) MDS subtypes elevated blast type 1 (EB-1, p=0.03, Chi^2 test) and 2 (EB-2, p=0.01, Chi^2 test), dimmer hematoxylin staining, and leukopenic PB blood count (Fig. 1F-G). Cluster 5 was characterized by high NHC proportion and hypercellular BM typical for chronic myelomonocytic leukemia (CMML, p=0.005, Chi^2 test) and unclassifiable MDS/MPN-U patients (p=0.13, Chi^2 test). Moreover, these were also associated with darker hematoxylin staining. Cluster 4 harbored hypoplastic MDS (p=0.002, Chi^2 test) as well as low erythroid cell frequency linked with del(5q) MDS (12). Of note, the unsupervised clustering structure of MDS and MDS/MPN samples only partly overlapped with the WHO disease classification, likely as it was driven by tissue texture features whereas the WHO classes are defined by BM cytomorphology, cytogenetics, blast proportion and cytopenias. However, CNN-guided tissue analysis could potentially assist classification of challenging cases according to WHO criteria.

The MDS BM morphology is linked to mutation, karyotype, gender, and prognostic status

MDS is characterized by recurrent oncogenic somatic variants in driver genes and chromosomal aberrations, and the mutation profile detected in our patient cohort corresponded well to the known mutational landscape in MDS (Fig. 2A-B, Supplementary Fig. 2B-C) (13–15). We adapted a transfer learning approach, where we used VGG16 and Xception network features extracted from tile-level H&E images to develop elastic net-regularized regression models (Supplementary Fig. 1). Predictions at the TMA spot level outweighed tile-level detection values due to uncertainties related to the limited image data in individual tiles. After averaging tile-level predictions at the TMA sample level, we could accurately detect notably *TET2,*

ASXL1, and *STAG2* mutations, chromosome 7 monosomy, and 7q deletion from morphological features (Fig. 3A-D). Moreover, morphological features were associated with point mutations in genes regulating splicing, cell differentiation, and cell cycle, which are commonly affected in MDS.

Secondary MDS and complex karyotype were often observed in the same patients ($p=0.008$, Chi^2 test), and their morphological features were most challenging to learn possibly due to higher variability in tissue texture. Features extracted with the Xception CNN model were more generalizable than those extracted with VGG16, concomitant with higher feature heterogeneity observed in the Xception correlation matrix (Supplementary Fig. 3 and 4). In addition, lasso and elastic net penalization provided more generalizable models than ridge regression (Supplementary Fig. 4).

Next, we investigated the validity of models predicting mutated genes and dysregulated pathways. The detection probability of a distinct mutation correlated significantly with its variant allele frequency within the sample for *ASXL1, KRAS/NRAS, IDH1/IDH2,* and *RUNX1* genes, and showed a trend for *TET2* and *TP53* genes (Fig. 3E-F and Supplementary Fig. 5). Moreover, the predicted likelihood of *RAS*, cell differentiation, and chromatin structure regulating gene pathway dysregulation correlated with the number of genes mutated in the respective pathways (Fig. 3G). This signifies that the prediction probability is higher when the sample contains more mutated cells, represented as higher VAF, or the sample has higher proportion of mutated genes in the same pathway. This implies that prominent genetic changes translate to distinct morphological image features. Of note, the initial models were developed using data from all patients with available myeloid mutation panel analysis, while the correlation analysis was restricted to samples with known

mutation, indicating BM tissue morphology to be impacted by molecular genetics and emphasizing the algorithms' ability to identify variant-related histopathological patterns in an unprecedented fashion.

Contrary to the histopathological evaluation of solid tumors, tissue morphology is not factored in risk stratification of MDS patients. Instead, the IPSS-R score accounting for PB cell count, BM blast burden, and cytogenetics is the most established stratification tool, and were associated in this dataset with risk for AML and OS (Supplementary Fig. 6A-B) (2). We could predict IPSS-R score, OS and progression to AML by solely employing H&E-stained slides (Fig. 3A-B, H). When evaluating progression to AML, BM morphology was associated with inferior prediction capability compared to the IPSS-R score but performed slightly better than IPSS-R when predicting OS within 2 years (Supplementary Fig. 6C-D). The best model was achieved by combining both deep histopathology and IPSS-R scores (Supplementary Fig. 6D). These results might be impacted by differences in patient selection, as our cohort represented an unbiased real-world cohort, while patients treated with disease-modifying therapies were excluded from the landmark prognostic study by Greenberg *et al.* (2). However, patient prognostication and treatment stratification could be significantly improved by including morphological features.

Interpretation of supervised prediction models with multilevel image analysis

MDS is differentially diagnosed from MDS/MPN by cytomorphological and histopathological examination, evaluation of PB counts, flow cytometry, karyotype, and increasingly, molecular genetics. In our study, tissue samples were partitioned into smaller TMA cores possibly limiting effective diagnostic segregation. Yet, we could discern MDS patients from MDS/MPN patients with an AUROC validation

accuracy of 0.81 (Fig. 3A). To help in understanding which texture pattern are associated with various clinical variable, we interpreted regression models by correlating tile-level tissue texture predictions with pixel and NHC-level metrics (Fig. 4A). In addition, we present both tile and TMA spot level images with the highest and lowest association with the predicted endpoint as well as TMA-level activation maps (Supplementary Fig. 7). Pixel classification was trained to identify NHCs, RBCs, stroma including fibrotic stroma and bone trabeculae, and lipid droplets, constituting the major tissue elements in standard H&E staining. Additionally, NHCs were segmented to extract nuclear measurements such as size, circularity, and H&E dye variations. As expected, MDS likelihood increased if the sample represented hypoplastic texture (Fig. 4B-C). MDS morphology was also associated with dimmer H&E staining and larger nuclear size likely due to a technically larger cell segmentation area in hypocellular BM (Fig. 4A, Supplementary Fig. 7A).

Chromosome 5q deletion is associated with a decrease in erythroid precursor cells and an increase in hypolobular megakaryocytes (12). When inspecting tiles associated with 5q deletion, we discovered enrichment of megakaryocytes with abnormally circular nuclei in line with previous findings (Fig. 4D). While hMDS was not associated with 5q deletion syndrome in our cohort (7/51 vs. 14/183 samples had 5q deletion syndrome in hMDS vs. non-hMDS, $p=0.29$, Chi^2 test), del(5q) was identified more often in hypocellular samples and samples with higher stroma content (Fig. 4E). These findings could be due to prior treatment with lenalidomide indicated for MDS with del(5q). Moreover, del(5q) MDS samples displayed lower nucleus circularity. This could be due to different white blood cellular content rather than dysplasia, as this pattern was absent in samples with complex karyotype (Fig. 4F).

Discussion

Here, we demonstrate how the intricate and heterogeneous BM morphological landscape can be dissected and associated with clinical data using multilevel computer vision.

While the presented analytical platform is not accurate and extensive enough to replace sequencing technologies in identifying genetic lesions, deep BM morphology predicted mutations and cytogenetic aberrations with accuracy exceeding that of similar applications in solid tumors (4,7,16,17). We suspect that the homogeneous BM tissue consistency and lower mutation burden in MDS may account for the results. To increase reliability, we employed transfer learning and separated the training and validation datasets at the TMA sample level. However, in the absence of an independent test cohort, the results should be interpreted with caution.

Based on our findings, transfer learning of H&E images can shed light on disease pathology and gene targets by linking molecular genetics with tissue dysplasia. It would be interesting to compare the texture of MDS BM samples collected in clinical trials or derived from animal models to images in our interactive platform (http://hruh-20.it.helsinki.fi/mds_visualization). The mechanism of action of any drug could, thus, be examined with respect to the signaling pathways or molecular genetics aberrations described in this study. The potential of this approach could further expand if explored by combining genetic variants with protein production quantitated with immunohistochemistry to comprehensively observe the effects of transcriptional, spliceosomal and translation regulation on eventual protein production.

The analysis was conducted for images digitized with 20x objective as we initially aimed to combine cellular-level morphological patterns with clinical variables. However, the resolution remains suboptimal for the distinction of subcellular organelles, and quantification of chromatin or nucleus-to-cytoplasm ratio. Addressing these questions would likely require analysis of MGG-stained BM aspirate samples with better morphological quality using high-magnification (100x) microscopes and oil immersion.

While presence of bone trabeculae could be related to sampling artefacts, the BM microenvironment and erythropoietin are known to regulate the osteoblast-osteoclast balance, and ultimately bone remodeling (18,19). Moreover, both cellular and acellular components contribute to the BM niche (20). Therefore, bone trabeculae were not excluded from image analysis, and further examination of the BM microenvironment is needed to validate associations with molecular genetics.

Digital pathology and adaptation of CNNs have had substantial influence in the field of solid tumors (21). Standardized clinical application would require multicenter collaboration to both collect sufficient training data and to impact clinical routine. Our current results suggest that image analysis of H&E-stained BM samples is insufficient alone to reliably differentiate MDS and MDS/MPN subtypes. However, especially the trained eye of the pathologist may benefit by learning novel associations between morphology and molecular genetics. Moreover, CNN-based analytics could reduce intra- and interobserver variability of histopathological analysis, and by incorporating additional clinical information could assist in automating and objectively classifying MDS patients. A recent study took a slightly different approach and used hematopathological reports instead of image analysis to link BM morphology and

clinical variables (22). Cytomorphological findings such as dysplasia, monocytosis and elevated megakaryocytes from MGG-stained BM smears and presence of myelofibrosis from H&E-stained BM biopsies were shown to associate with cytopenia and mutations emphasizing that machine learning-based platforms could support MDS diagnosis.

The black box dilemma hinders clinical translation of deep learning algorithms (23). To increase model transparency, we decoded CNN-extracted morphological patterns associated with molecular and clinical determinants with a holistic methodology linking image analysis at the tile, pixel, and cellular levels. CNN-based feature extraction, pixel classification and NHC segmentation required little parameter optimization to increase reproducibility and scalability to any image analysis study employing deep learning. Image features associated with MDS vs. MDS/MPN diagnosis and del(5q) MDS were affirmed by a hematopathologist and were consistent with previous knowledge (1,12). Dissection of complex computer vision models could be achieved with similar approaches based on neural network-based semantic and instance segmentation methods.

Taken together, deep mining of the BM tissue texture on a larger scale could assist pathologists by revealing intricate morphological patterns defining disease subtypes and eventually improving clinical stratification of MDS patients.

Materials and methods

Patients. The study population comprised MDS (n=143) and myelodysplastic/myeloproliferative neoplasm (MDS/MPN) patients (n=51) and control subjects (n=11) treated in 2000-2018 at the Department of Hematology in the

Helsinki University Hospital (HUH), Finland (Supplementary Fig. 1 and Table 1). The formal MDS or MDS/MPN diagnosis and classification adhered to WHO guidelines and were assigned by the treating clinical hematologist on the basis of patient history, PB cell counts, BM cytomorphology evaluated by two laboratory hematology specialists, and cytogenetics according to WHO guidelines. If needed, flow cytometry results were included in the evaluation. The diagnostic conclusion, especially if challenging, was confirmed by a regular tumor board of clinical hematologists and a laboratory hematologist at our institution (24,25).

BM trephine biopsies are used in standard diagnostic procedures and occasionally for assessing treatment response and disease progression. As all available samples were obtained, no selection bias in sample collection occurred. Diagnostic (MDS n=143, MDS/MPN n=51) and follow-up specimens (MDS n=93, MDS/MPN n=36) were collected prior to and after initiating a disease-modifying treatment, respectively. According to ethics board guidelines, control BM trephine samples were collected from subjects without diagnosis of hematologic malignancy, chronic infection, or autoimmune disorder in the six-year follow-up (Supplementary Table 1). Of the 11 control subjects, six were males and their median age was 57.0 [range 40.0-82.0] years at the time of BM sampling. MDS and MDS/MPN patients were slightly older than control subjects ($p=0.02$ and $p=0.002$, respectively) but did not differ significantly by sex (Table 1).

The study complied with the Declaration of Helsinki and the HUS institutional ethics committee. All study patients gave written informed consent. All clinical data were collected from the HUS datalake, a GDPR-compliant database.

Sequencing. Genomic DNA was isolated from diagnostic BM aspirates (n=108) using the QIAsymphony DSP DNA kit. According to the manufacturer's protocols, 20 ng of DNA for each sample was used for library preparation with the Ion AmpliSeq™ Library Kit 2.0 (Thermo Fisher Scientific) and the AmpliSeq primers (Thermo Fisher Scientific) for the in-house myeloid cancer gene panel. This panel covers all exons in ASXL1, BCOR, CDKN2A, CEBPA, CREBBP, CUX1, DNMT3A, EP300, ETV6, EZH2, GATA2, KDM6A, NF1, PHF6, RAD21, SETD2, STAG2, TET2, TP53 and ZRSR2 genes and hot spot exons in BRAF, CALR, CBL, CSF3R, FLT3, GATA1, IDH1, IDH2, JAK2, KIT, KRAS, MPL, NPM1, NRAS, PDGFRA, PTPN11, RUNX1, SETBP1, SF3B1, SMC1A, SMC3, SRSF2, U2AF1 and WT1 genes. Libraries were barcoded with the Ion Xpress Barcode Adapters (Thermo Fisher Scientific) and quantified with the Ion Library TaqMan™ Quantitation Kit. Template preparation and chip loading were performed on the Ion Chef system (Thermo Fisher Scientific). Sequencing was carried out on the Ion Proton or Ion GeneStudio S5 system (Thermo Fisher Scientific). Data analysis was carried out with Torrent Suite Software v.5.8 (Life Technologies). The Ion Reporter software v4.6 (Thermo Fisher Scientific) was used to filter out non-coding and polymorphic variants. All variants listed after filtering were visualized in the Integrative Genomics Viewer (IGV) to manually discard alterations generated by incorrect calling.

Driver gene mutations were defined using a clinical-grade, myeloid amplicon sequencing panel capable of identifying mutations with variant allele frequency >2% (mean sequencing depth 6000x; Supplementary Fig. 2B-C). In addition, data from an additional 40 samples analyzed with the Illumina TruSight myeloid amplicon sequencing panel were included (mean sequencing depth 100x). Sequencing panels produced similar results for 305/315 genes (96.8%) by analyzing nine samples in

parallel with both assays.

Tissue microarrays (TMAs). Upon sampling, fresh BM biopsies were conformed to routine EDTA-based decalcification softening the tissue and formalin-fixation and paraffin-embedding (FFPE). TMAs were cast by a single 2 mm core (MDS and MDS/MPN) or double 1 mm cores (controls) per sample from representative BM biopsy areas (Fig. 1A). Duplicate control cores did not differ by their morphological patterns (Supplementary Fig. 2A). Tissue blocks were cut into 4 μm thick sections and stained with hematoxylin and eosin (H&E). Slides were digitized at 0.22 μm/pixel (20x objective magnification) with the whole-slide scanner Pannoramic 250 FLASH (3DHISTECH Ltd.).

Image analysis. Images were analyzed with three methods: (A) transfer learning, (B) pixel classification, and (C) NHC analysis to capture information from different representation levels (Supplementary Fig. 1).

A. Transfer learning.

Image preprocessing. As sample processing and slide digitization affect color distribution, H&E-stained RGB images were converted into grayscale with the OpenCV Python module to reduce possible technical noise. Moreover, we standardized background pixels that did not represent BM tissue by converting them into the pixel value 255 (white). H&E images were split into 500 equal-sized tiles (small subimages of the original H&E image), in average 470x470 pixels. Tile size was optimized to outsize the BM lipid droplets to avoid their classification as non-tissue background. Tiles with mean pixel intensity over 240 represented non-tissue background and were excluded from the analysis resulting in 73 531 tiles.

Feature extraction. In transfer learning, a prior algorithm or its configurations are reused to develop a new machine learning algorithm. Visual features that were previously defined in a larger training dataset can be repurposed for new image analysis tasks reducing the need for training models from scratch (26). We adapted a transfer learning approach where image tile features were extracted with pretrained Xception and VGG16 convolutional networks which have achieved high accuracy in classifying the ImageNet dataset commonly used to develop and benchmark computer vision algorithms (27,28). Individual tiles were resized into equal sizes (224x224 pixels for VGG16 and 299x299 pixels for Xception). Colors were rescaled between values [0,1]. Images were reformatted as 3-channel grayscale where each pixel was assigned an identical value for each channel to ensure compatibility with pretrained CNNs as ImageNet configurations have been developed for color images. For each tile image, a 2.048-bin feature vector was extracted at the second last Xception network layer. As VGG16 and Xception networks do not possess similar number of parameters, we exported features from the last layer of the VGG16 network (n=25.088) and retained only 2.048 features with the highest variance-to-mean ratio (Supplementary Fig. 3A-B).

Regression models. Visual features extracted from image tiles with CNNs contain numerical values. Therefore, these can be introduced as covariates to develop regression models. Image tiles were first split at the TMA sample level into training (2/3) and test (1/3) datasets. In average, 213 (range 70-336) tiles per sample were available. Models were trained with L1 (alpha = 1), L2 (alpha = 0) or elastic net-regularized (alpha = 0.5) regression models. Here, multivariate regression models are optimized for both regularizing a factor alpha and penalization rate lambda using 5-fold crossvalidation. Lambda values were optimized for each fixed parameter alpha to reach minimum cross-validation error (lambda.min) and the lambda at one

standard error of this minimum (lambda.1se). Training of the regression models occurred at the tile level to highlight individual tissue texture and cell morphology features. The prediction results were assessed both at tile level and TMA sample level. Prediction probabilities for TMA samples were calculated by summing probabilities of each tile of the same TMA sample and dividing the total by the number of tiles. In summary, each predicted variable was estimated based on 12 algorithms (feature matrices extracted with two CNN models and three alpha and two lambda elastic net regularization parameter values).

Follow-up for survival analyses was restricted to 2 years to identify morphological patterns associated with aggressive disease and to ensure concordance between AUROC model evaluation and the predicted variable. Patients with shorter follow-up time or receiving allogenic hematopoietic stem cell transplant before the occurrence of an event were censored. Azacytidine was administered 75 mg/m^2 per day with a median interval of 29 days [27-33 days 25-75 percentiles] and a median of six treatment cycles [3-11 cycles, 25-75%]. Positive treatment response was defined as continuation of treatment over 6 months or complete remission prior to 6 months. Disease progression was determined to represent treatment failure. Patients treated for less than 6 months due to short follow-up or therapy-related side effects were censored.

As longitudinal samples were not characterized with lower variance in morphological features compared to samples from other patients, we included samples from multiple timepoints of the same patient for model training (Supplementary Fig. 8A-B). Prediction models were trained with MDS BM images only except for differential diagnosis, where both MDS and MDS/MPN samples were used. Separate prediction models for IPSS-R score and age at diagnosis were developed with linear regression and using only diagnostic MDS samples. Age was transformed into categories of

458 <50.0, 50.0-59.9, 60.0-69.9, 70.0-79.9, and >80.0 years age categories, which
459 ameliorated the accuracy and interpretation of results. Gender, mutations,
460 cytogenetic aberrations, and MDS etiology were predicted with logistic regression
461 using both diagnostic and follow-up samples. Prognosis was measured by predicting
462 overall survival (OS) and acute myeloid leukemia (AML) progression using both
463 diagnostic and follow-up samples. Only mutations and chromosomal aberrancies
464 present in over 9% of the samples were selected. Disease etiology was assigned as
465 either “de novo” or “secondary MDS”. Azacytidine response was predicted with
466 logistic regression using samples taken 0-365 days before treatment start. Samples
467 taken during HMA treatment did not differ by their morphological profile from other
468 samples suggesting that HMA does not alter the BM morphology in a distinct fashion
469 (Supplementary Fig. 8C-D).
470
471 **B. Pixel classification.** Each RGB TMA spot image was analyzed with the Trainable
472 Weka Segmentation module of the Fiji software using default parameters (29). With
473 the assistance of hematopathologic expertise, we manually delineated areas
474 representing NHCs, red blood cells (RBCs), stroma or lipid droplet/background from
475 ten images and trained a tissue pixel classifier. NHCs represented nucleated cells
476 consisting primarily of leukocytes and occasional erythroblasts and megakaryocytes.
477 Stroma included fibrotic stroma and bone trabeculae as both are associated with
478 MDS pathology (30,31). We manually annotated additional ten test images to
479 evaluate the accuracy of the classifier. Each pixel was classified with global 97%
480 accuracy composed of NHC (99% accuracy), RBC (84% accuracy), stroma (80%
481 accuracy), or lipid droplet/background (100% accuracy; Supplementary Fig. 9A-B).
482 The relative area of individual classes was calculated as their proportion to a binary
483 tissue mask. The tissue mask was created by converting H&E images into binary

format where white areas represent tissue and black areas adipose tissue and empty background. Then, we performed mask dilatation, empty hole fill and mask erosion steps to differ adipose tissue from the background (Supplementary Fig. 9C). Lipid droplets were defined as filled image holes from the initial tissue mask.

C. NHC analysis. Each RGB TMA spot image was analyzed with the open-source software QuPath (v0.2.0) (32). NHCs were segmented with the watershed cell detection and background radius 30px, median filter radius 0px, sigma 6px, minimum area 10px^2, maximum area 10.000px^2, threshold 0.1, maximum background intensity 2, and cell expansion 5px (Supplementary Fig. 10). Nucleus hematoxylin and eosin staining intensity was defined as their optical density after color deconvolution. Nucleus size was calculated based on area and eccentricity based on the deviance from a perfect circle. These metrics were extracted for individual NHCs and averaged at the TMA spot level (Supplementary Table 2).

Statistical analysis. Continuous variables were compared with Wilcoxon Rank-Sum test (unpaired, two-tailed) and correlated with Spearman's rank correlation coefficient. Categorical variables were compared with Chi^2 test. P-values were adjusted with Benjamini-Hochberg's correction when necessary. The log-rank test for Kaplan-Meier analysis was used for survival analysis. Model fitness was assessed by calculating statistical significance of the area under the receiver operating characteristic curves (AUROC). AUROC values for predicting AML progression, OS and IPSS-R were further compared with DeLong's test.

For unsupervised analysis, we selected a uniform manifold approximation and projection (UMAP) method (33). PhenoGraph is a graph-based community detection

method designed for high-resolution single-cell data analogous to visual features (34). Single tiles were clustered with PhenoGraph with default settings to attain higher granularity to distinguish morphological features. K-means clustering was selected for sample grouping to simplify interpretation of TMA spot grouping. The *k* parameter was harmonized using the consensus of 30 indices based on Euclidean distance and Ward agglomeration metrics (35). Feature extraction, regression models, and statistical analysis were performed with R 3.5.1 (Supplementary Table 3).

Code and data availability. Codes used for data analysis are available at https://github.com/obruck/MDS_HE_IA/. Image data and activation maps are available at http://hruh-20.it.helsinki.fi/mds_visualization.

Acknowledgments

We are grateful to the members of the Hematology Research Unit Helsinki for discussions and technical help. We thank the Helsinki Biobank for sample collection and the Digital and Molecular Pathology Unit supported by the University of Helsinki and Biocenter Finland for digital microscopy services. We thank the University of Helsinki language center for editing and proofreading services.

This study was supported by the University of Helsinki, the Doctoral Programme in Biomedicine (DPBM), personal grants (O. Brück) from Biomedicum Helsinki Foundation, Finska Läkaresällskapet, research grants (S. Mustjoki) from the Cancer Foundation Finland, Sigrid Juselius Foundation, Signe and Ane Gyllenberg Foundation, Relander Foundation, State funding for university-level health research in Finland, a Helsinki Institute of Life Sciences Fellow grant, and an investigator-initiated research grant from Novartis.

Authorship Contributions

Conception and design: O.B., S.M. Collection and assembly of clinical data: H.H., K.P., O.B., S.L.-B. Collection and assembly of image data: O.B., P.K. Collection and assembly of sequencing data: F.E., O.B., S.K., S.M. Image analysis and data analysis: O.B. Visualization: A.I., O.B., T.A. Data interpretation: All authors. Manuscript writing: All authors. Final approval of manuscript: All authors.

References

1. Montalban-Bravo G, Garcia-Manero G. Myelodysplastic syndromes: 2018 update on diagnosis, risk-stratification and management. Am J Hematol. 2018;93:129–47.
2. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Solé F, et al. Revised International Prognostic Scoring System for Myelodysplastic Syndromes. Blood [Internet]. 2012;120:2454–65. Available from: https://doi.org/10.1182/blood-2012-03-420489
3. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015. page 436–44.
4. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat Med. 2018;24:1559–67.
5. Titano JJ, Badgeley M, Schefflein J, Pain M, Su A, Cai M, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat Med [Internet]. 2018;24:1337–41. Available from: https://doi.org/10.1038/s41591-018-0147-y
6. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature [Internet]. 2017;542:115–8. Available from: https://doi.org/10.1038/nature21056
7. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med [Internet]. Springer US; 2019;25:1054–6. Available from: http://dx.doi.org/10.1038/s41591-019-0462-y
8. Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a

population-based, diagnostic study. Lancet Oncol [Internet]. Elsevier; 2020;21:222–32. Available from: https://doi.org/10.1016/S1470-2045(19)30738-7

9. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci Rep [Internet]. Springer US; 2018;8:1–11. Available from: http://dx.doi.org/10.1038/s41598-018-21758-3

10. Schemenau J, Baldus S, Anlauf M, Reinecke P, Braunstein S, Blum S, et al. Cellularity, characteristics of hematopoietic parameters and prognosis in myelodysplastic syndromes. Eur J Haematol [Internet]. John Wiley & Sons, Ltd; 2015;95:181–9. Available from: https://doi.org/10.1111/ejh.12512

11. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics [Internet]. 2018;19:477. Available from: https://doi.org/10.1186/s12864-018-4772-0

12. VAN DEN BERGHE H, CASSIMAN J-J, DAVID G, FRYNS J-P, MICHAUX J-L, SOKAL G. Distinct haematological disorder with deletion of long arm of No. 5 chromosome. Nature [Internet]. 1974;251:437–8. Available from: https://doi.org/10.1038/251437a0

13. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia [Internet]. Nature Publishing Group; 2013;28:241–7. Available from: http://dx.doi.org/10.1038/leu.2013.336

14. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood [Internet]. 2013;122:3616 LP – 3627. Available from:

http://www.bloodjournal.org/content/122/22/3616.abstract

15. Mason CC, Khorashad JS, Tantravahi SK, Kelley TW, Zabriskie MS, Yan D, et al. Age-related mutations and chronic myelomonocytic leukemia. Leukemia [Internet]. 2016;30:906–13. Available from: https://doi.org/10.1038/leu.2015.337

16. Kather JN, Heij LR, Grabsch HI, Loeffler C, Echle A, Muti HS, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer [Internet]. 2020; Available from: https://doi.org/10.1038/s43018-020-0087-6

17. Fu Y, Jung A, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer [Internet]. 2020; Available from: https://doi.org/10.1038/s43018-020-0085-8

18. Mellibovsky L, Diez A, Serrano S, Aubia J, Pérez-Vila E, Mariñoso ML, et al. Bone remodeling alterations in myelodysplastic syndrome. Bone [Internet]. 1996;19:401–5. Available from: http://www.sciencedirect.com/science/article/pii/S8756328296002104

19. Balaian E, Wobus M, Weidner H, Baschant U, Stiehler M, Ehninger G, et al. Erythropoietin inhibits osteoblast function in myelodysplastic syndromes via the canonical Wnt pathway. Haematologica [Internet]. 2017/10/27. Ferrata Storti Foundation; 2018;103:61–8. Available from: https://pubmed.ncbi.nlm.nih.gov/29079596

20. Pronk E, Raaijmakers MHGP. The mesenchymal niche in MDS. Blood [Internet]. 2019;133:1031–8. Available from: https://doi.org/10.1182/blood-2018-10-844639

21. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly

supervised deep learning on whole slide images. Nat Med [Internet]. 2019;25:1301–9. Available from: https://doi.org/10.1038/s41591-019-0508-1

22. Nagata Y, Zhao R, Awada H, Kerr CM, Mirzaev I, Kongkiatkamon S, et al. Machine learning demonstrates that somatic mutations imprint invariant morphologic features in myelodysplastic syndromes. Blood [Internet]. 2020;136:2249–62. Available from: https://doi.org/10.1182/blood.2020005488

23. Medicine TLR. Opening the black box of machine learning. Lancet Respir Med [Internet]. Elsevier; 2018;6:801. Available from: https://doi.org/10.1016/S2213-2600(18)30425-9

24. Guidelines C. MDS and CMML Guidelines Guidelines for the diagnosis and treatment of Myelodysplastic Syndrome and Chronic Myelomonocytic Leukemia Nordic MDS Group. 2017;

25. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood [Internet]. 2016;127:2391–405. Available from: https://doi.org/10.1182/blood-2016-03-643544

26. Shin H, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans Med Imaging. 2016;35:1285–98.

27. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv Prepr arXiv14091556. 2014;

28. Chollet F. Xception: Deep learning with depthwise separable convolutions. Proc IEEE Conf Comput Vis pattern Recognit. 2017. page 1251–8.

29. Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, et al. Trainable Weka Segmentation: a machine learning tool for microscopy

pixel classification. Bioinformatics. Oxford University Press; 2017;33:2424–6.

30. Della Porta MG, Malcovati L. Myelodysplastic syndromes with bone marrow fibrosis. Haematologica [Internet]. Ferrata Storti Foundation; 2011;96:180–3. Available from: https://pubmed.ncbi.nlm.nih.gov/21282718

31. Weidner H, Rauner M, Trautmann F, Schmitt J, Balaian E, Mies A, et al. Myelodysplastic syndromes and bone loss in mice and men. Leukemia [Internet]. 2017;31:1003–7. Available from: https://doi.org/10.1038/leu.2017.7

32. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep [Internet]. 2017;7:16878. Available from: http://www.nature.com/articles/s41598-017-17204-5

33. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv Prepr arXiv180203426. 2018;

34. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir EAD, Tadmor MD, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell [Internet]. Elsevier Inc.; 2015;162:184–97. Available from: http://dx.doi.org/10.1016/j.cell.2015.05.047

35. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust Package: finding the relevant number of clusters in a dataset. J Stat Softw. 2012;

669 Table 1. Patient characteristics at diagnosis.

	MDS (n=143)	MDS-MPN (n=51)
Gender, male	73 (51.0%)	36 (70.6%)
Age, mean [range]	64.8 [15.5-88.6]	67.5 [36.4-87.0]
Etiology		
• De novo	118 (82.5%)	47 (92.2%)
• Secondary (treatment-related)	25 (17.5%)	4 (7.8%)
WHO MDS classification		
• MDS-SLD	20 (14.0%)	
• MDS-MLD	33 (23.1%)	
• MDS-RS	16 (11.2%)	
• MDS-EB-1	25 (17.5%)	
• MDS-EB-2	35 (24.5%)	
• MDS with isolated del(5q)	10 (7.0%)	
• MDS-U	4 (2.8%)	
WHO MDS/MPN classification		
• CMML		35 (68.8%)
• MDS/MPN-U		16 (31.2%)
Survival status 2 years after diagnosis		
• Alive	60 (42.0%)	32 (15.7%)
• Deceased	70 (49.0%)	19 (84.3%)
• Not defined or censored due to alloHSCT	13 (9.1%)	0 (0.0%)
Progression to AML within 2 years of diagnosis		
• No	51 (35.7%)	28 (54.9%)
• Yes	42 (29.4%)	11 (21.6%)
• Not defined or censored due to alloHSCT	50 (35.0%)	12 (23.5%)
Azacytidine-treated within 1 year of diagnosis		
• No	78 (60.1%)	31 (66.7%)
• Yes	65 (39.9%)	20 (33.3%)
IPSS risk class		
• Low	24 (16.8%)	
• Intermediate-1	57 (39.9%)	
• Intermediate-2	26 (18.2%)	
• High	24 (16.8%)	
• Not defined	12 (8.4%)	
IPSS-R risk class		
• Very low	17 (11.9%)	
• Low	33 (23.1%)	
• Intermediate	27 (18.9%)	
• High	26 (18.2%)	
• Very high	28 (19.6%)	
• Not defined	12 (8.4%)	
AlloHSCT		
• No	117 (81.8%)	43(84.3%)
• Yes	24 (16.8%)	8 (15.7%)
• Not defined	2 (1.4%)	0 (0.0%)

670 Abbreviations: World Health Organization (WHO); Myelodysplastic syndrome with
671 single lineage dysplasia (MDS-SLD); MDS with multilineage dysplasia (MDS-MLD);
672 MDS with ring sideroblasts (MDS-RS); MDS with excess blasts 1 (MDS-EB-1); MDS
673 with excess blasts 2 (MDS-EB-2); MDS, unclassifiable (MDS-U); Chronic
674 myelomonocytic leukemia (CMML); Myelodysplastic/Myeloproliferative neoplasm,

675 unclassifiable (MDS/MPN-U); Acute myeloid leukemia (AML); Allogeneic
676 hematopoietic stem cell transplantation (alloHSCT); International Prognostic Scoring
677 System (IPSS); Revised IPSS (IPSS-R)

678 **Figure Legends**

679 **Fig. 1. Study design.** (a) Tissue microarrays (TMAs) were constructed from
680 formalin-fixed paraffin-embedded bone marrow (BM) trephine biopsies and stained
681 with hematoxylin and eosin. Images were analyzed at tile, pixel, and nucleated
682 hematopoietic cells (NHC) levels. In the tile-level analysis, TMA spot images were
683 split into 500 small patches, and morphological features were extracted with
684 ImageNet-pretrained convolutional neural networks. We developed elastic net-
685 regularized algorithms to predict multiple clinical and molecular genetic variables
686 using only morphological features. Tile-level feature features for each prediction
687 model were joined into activation maps where the probability of each tile to associate
688 with an endpoint is visualized with a heatmap color panel. A Weka pixel classifier
689 was trained to identify NHC, red blood cell (RBC), stroma, and lipid droplet from
690 images. Moreover, the nuclei of NHCs were segmented, and their geometrical
691 measures were extracted. (b) Extracted visual features from image tiles of diagnostic
692 myelodysplastic syndrome (MDS), myelodysplastic/myeloproliferative neoplasm
693 syndrome (MDS/MPN), and healthy subjects are plotted using uniform manifold
694 approximation and projection (UMAP). We have selected 10 random tiles from each
695 TMA sample for the UMAP visualization. Colors represent different PhenoGraph
696 clusters (larger image) or diagnoses (upper right image). Dashed circles represent
697 morphological subcategories where PhenoGraph clusters have been regrouped with
698 k-means clustering. Superposed lines demonstrate morphological trajectories
699 generated with a slingshot analysis. Study flow is described in detail in
700 Supplementary Fig. 1. (c) UMAP projection of TMA samples. Visual features of tile
701 images are combined at the TMA sample level by their mean value. TMA samples
702 are then clustered with the k-means method (numbered large circles). Circle colors
703 represent the corresponding diagnosis, (d) tissue proportion of NHCs and (e) stroma,

and (f) WHO classification. (g) Each k-means cluster has been compared to remaining clusters (Wilcoxon Rank-Sum test and Benjamini-Hochberg p-value correction) for segmented NHC, pixel-level image analysis parameters and clinical information. Clinical information is not reported for healthy patients (Cluster 1). Variables are described in Supplementary Table 2.

Fig. 2. Distribution of genomic alterations. (a) Oncoprint visualization of mutation pattern and gene groups in myelodysplastic syndrome (MDS) patients. The top bars represent the frequency of mutations per patient. The right bars represent the frequency of patients with a given mutated gene. The bottom bars represent the survival status of each patient. (b) Oncoprint visualization of cytogenetics pattern in samples from MDS patients. The top bars represent the frequency of the described karyotype aberrations per patient. The right bars represent the frequency of patients with a given karyotype aberration.

Fig. 3. Supervised learning on bone marrow (BM) morphological features. (a) Heatmap displaying area under the receiver operating characteristic curve (AUROC) values of elastic net-regularized logistic regression models. The left-most column shows the number of samples included in the analysis and the cell color represent the distribution of the binary variables to be predicted. The following two columns inform the AUROC values of the models in the training (2/3) and test (1/3) dataset in the tile-level images and the last two columns at the tissue microarray (TMA) spot level. Abbreviations used for predicted variables; SPLICINGmut: spliceosome mutations. CELLDIFFERENTIATIONmut: mutation in genes regulating cell differentiation. CELLCYCLEmut: mutation in genes regulating cellcycle. DNACHROMATINmut: mutation in genes regulating DNA chromatin structure.

RASmut: mutation in *NRAS* or *KRAS*. ABNCHR: presence of any abnormal chromosome. IDHmut: mutation in *IDH1* or *IDH2*. OS2y: overall survival event within 2 years of follow-up. RASPATHWAYmut: mutation in genes regulating RAS pathway. AML2y: progression to AML within 2 years of follow-up. AZA: azacytidine treatment response. (b) Similar plot for elastic net-regularized linear regression models. The left-most column shows the Spearman correlation value in the training dataset and right-most in the test dataset. (c) Tile-level (left) and TMA spot-level (right) AUROC for the logistic regression of monosomy 7 and (d) ASXL1 mutation status. The analysis shows the consistency of the predicted and true occurrence of an aberration. (e) Scatter plot for the Spearman correlation (x-axis) between logistic regression predicting mutation status and the observed gene variant allele frequency for individual genes or number of genes mutated for functional pathways. (f) Linear regression (R represents Spearman correlation) between prediction probability of ASXL1 mutation and its detected variant allele frequency. (g) Wilcoxon comparison for predicted mutation probability and detected frequency of altered genes in the *RAS* pathway. (h) Linear regression (R represents Spearman correlation) between the predicted and observed IPSS-R score in the training (left) and test (right) datasets.

Fig. 4. Interpretation of supervised prediction models. (a) Correlation matrix for logistic and linear regression model predictions (rows) and pixel-level and segmented nucleated hematopoietic cell-level (NHC) image analysis metrics aggregated per sample. The color of individual matrix cells represents the Spearman correlation and the asterisks the Benjamini-Hochberg-adjusted significance values: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Variables are described in Supplementary Table 2. (b) H&E-stained TMA spots and corresponding activation maps for the prediction of myelodysplastic syndrome or myelodysplastic/myeloproliferative neoplasm diagnosis

(MDS or MDS/MPN). For the activation map, tile-level predictions have been color-scaled (blue: high probability for MDS/MPN, red: high probability for MDS). (c) Scatter plot and Wilcoxon Rank-Sum test to compare NHC proportion by MDS and MDS/MPN diagnoses. Boxplots define the interquartile ranges and median values by diagnoses. (d) Image tiles representing highest and lowest computed probability of chromosome 5q deletion. (e) Scatter plot and Wilcoxon Rank-Sum test to compare the proportion of stroma and (f) nuclei eccentricity (opposite of circularity) by del(5q) status. Boxplots define the interquartile ranges and median values by diagnoses.

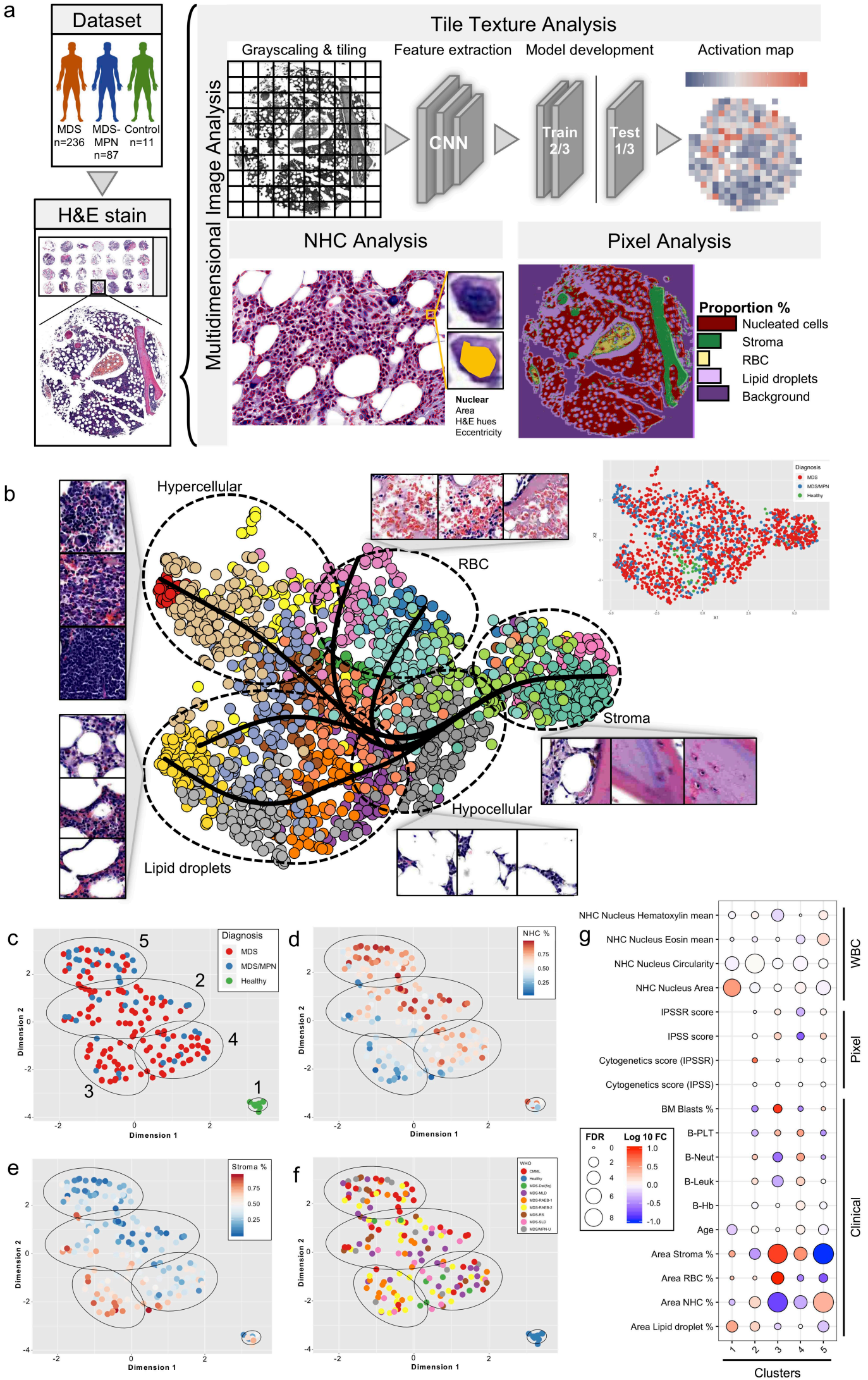
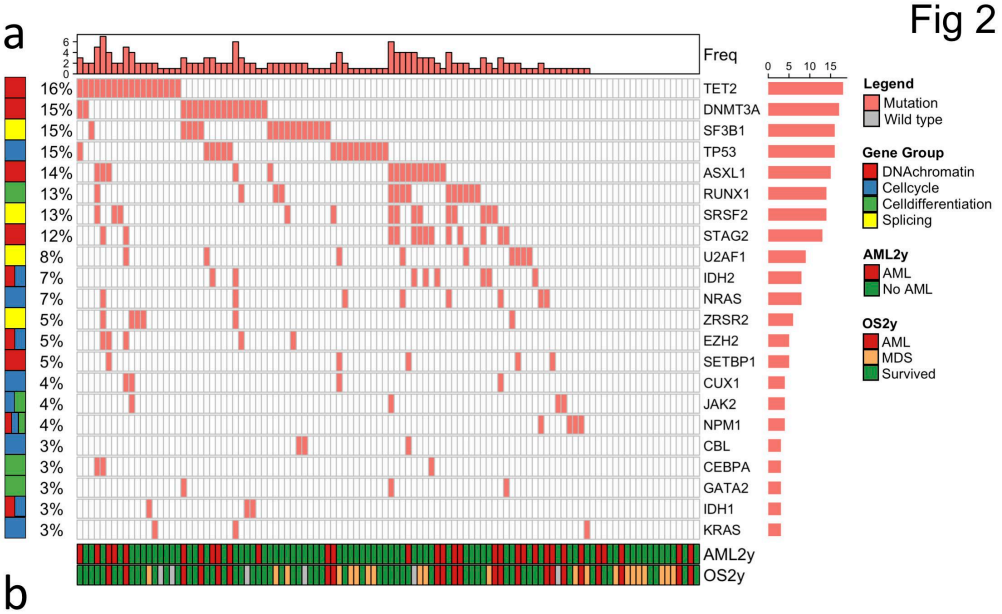

a
Dataset
MDS n=236
MDS-MPN n=87
Control n=11
H&E stain
Multidimensional Image Analysis
Tile Texture Analysis
Grayscaling & tiling
Feature extraction
Model development
Activation map
CNN
Train 2/3
Test 1/3
NHC Analysis
Pixel Analysis
Nuclear
Area
H&E hues
Eccentricity
Proportion %
Nucleated cells
Stroma
RBC
Lipid droplets
Background
b
Hypercellular
RBC
Stroma
Hypocellular
Lipid droplets
Diagnosis
MDS
MDS/MPN
Healthy
c
d
e
f
Dimension 1
Dimension 2
Diagnosis
NHC %
Stroma %
WHO
g
NHC Nucleus Hematoxylin mean
NHC Nucleus Eosin mean
NHC Nucleus Circularity
NHC Nucleus Area
IPSSR score
IPSS score
Cytogenetics score (IPSSR)
Cytogenetics score (IPSS)
BM Blasts %
B-PLT
B-Neut
B-Leuk
B-Hb
Age
Area Stroma %
Area RBC %
Area NHC %
Area Lipid droplet %
FDR
Log 10 FC
WBC
Pixel
Clinical
Clusters

Fig 2


a
Freq
TET2
DNMT3A
SF3B1
TP53
ASXL1
RUNX1
SRSF2
STAG2
U2AF1
IDH2
NRAS
ZRSR2
EZH2
SETBP1
CUX1
JAK2
NPM1
CBL
CEBPA
GATA2
IDH1
KRAS
AML2y
OS2y
Legend
Mutation
Wild type
Gene Group
DNAchromatin
Cellcycle
Celldifferentiation
Splicing
AML2y
AML
No AML
OS2y
AML
MDS
Survived

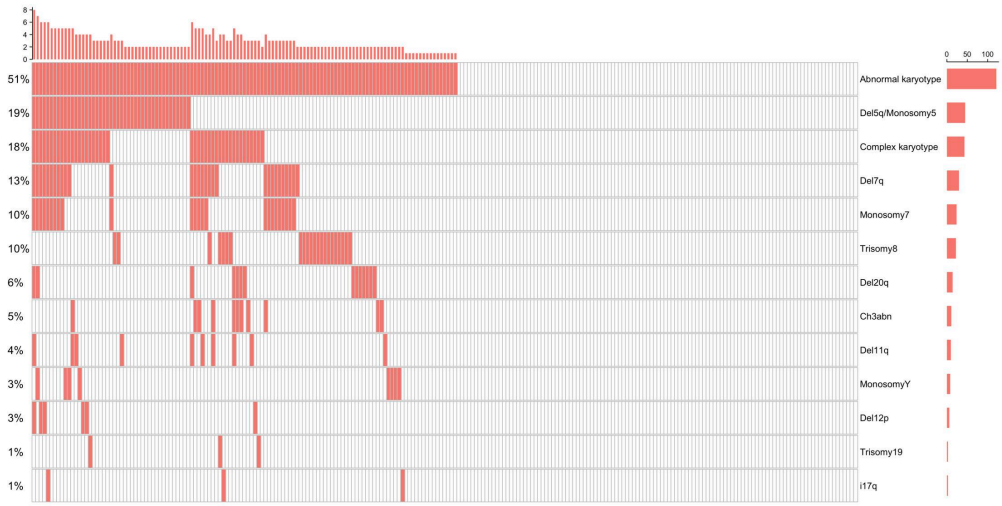

b
Abnormal karyotype
Del5q/Monosomy5
Complex karyotype
Del7q
Monosomy7
Trisomy8
Del20q
Ch3abn
Del11q
MonosomyY
Del12p
Trisomy19
i17q

Fig 3

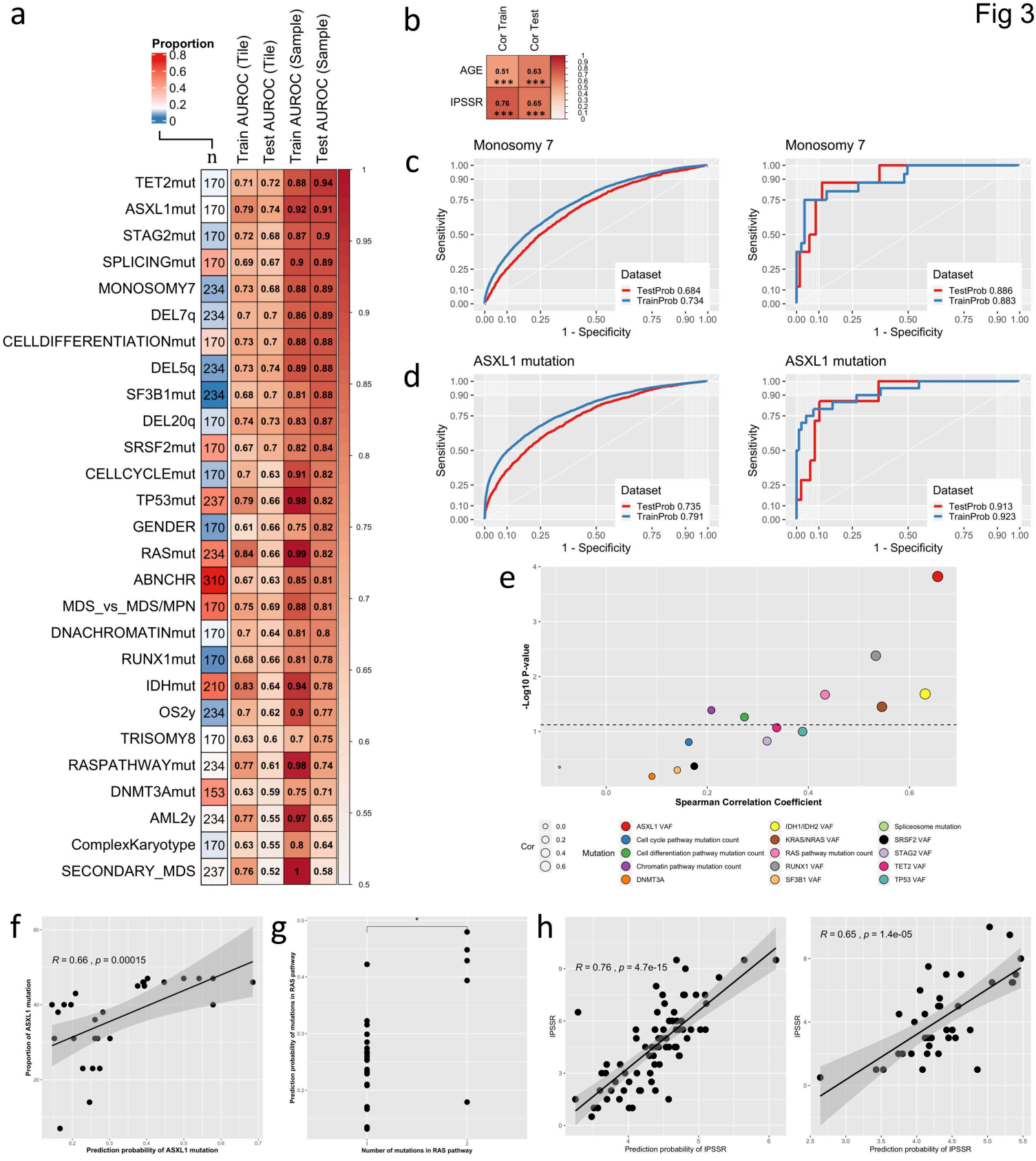
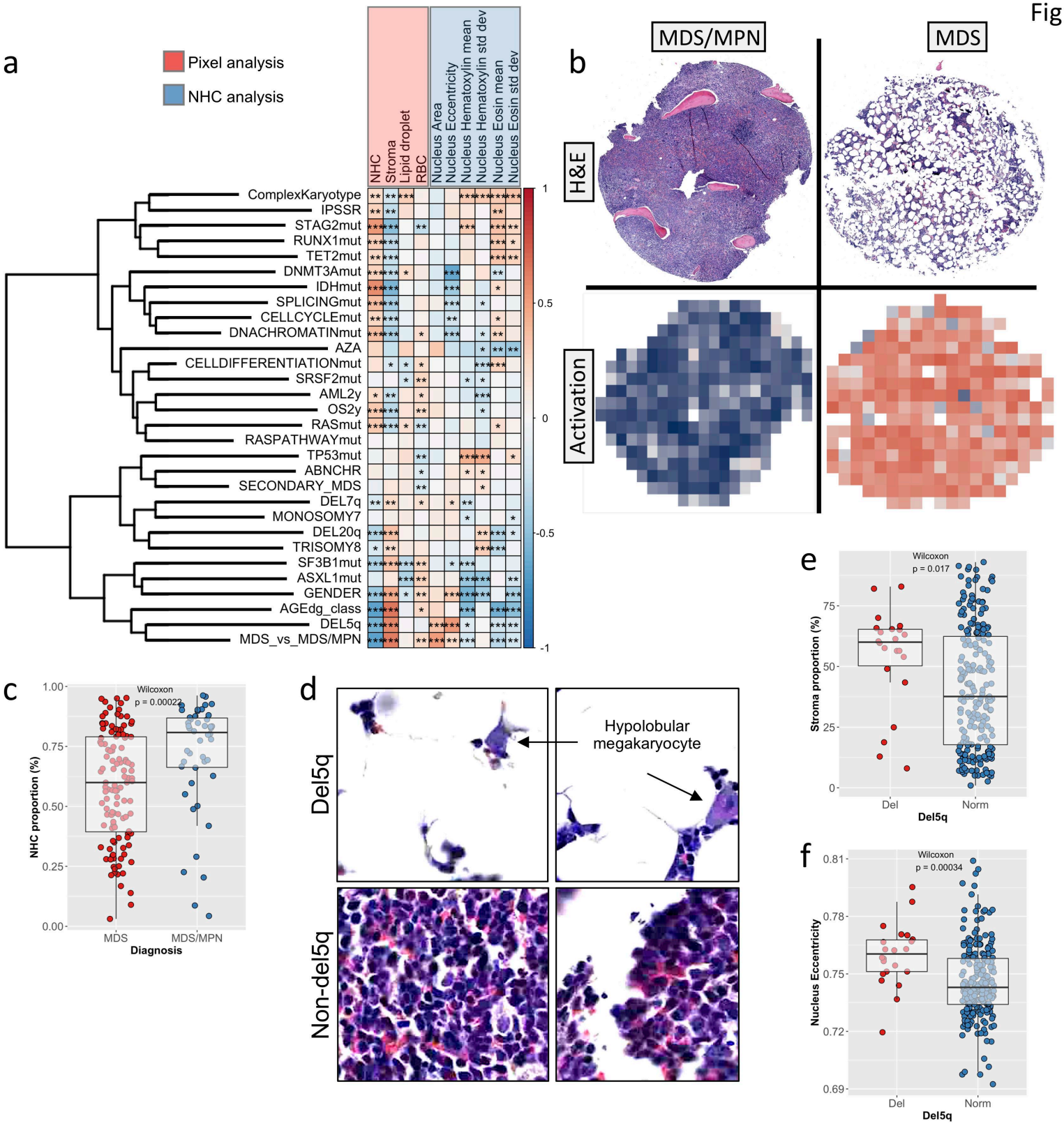

a
Proportion
0.8
0.6
0.4
0.2
0
n
Train AUROC (Tile)
Test AUROC (Tile)
Train AUROC (Sample)
Test AUROC (Sample)
TET2mut 170 0.71 0.72 0.88 0.94
ASXL1mut 170 0.79 0.74 0.92 0.91
STAG2mut 170 0.72 0.68 0.87 0.9
SPLICINGmut 170 0.69 0.67 0.9 0.89
MONOSOMY7 234 0.73 0.68 0.88 0.89
DEL7q 234 0.7 0.7 0.86 0.89
CELLDIFFERENTIATIONmut 170 0.73 0.7 0.88 0.88
DEL5q 234 0.73 0.74 0.89 0.88
SF3B1mut 234 0.68 0.7 0.81 0.88
DEL20q 170 0.74 0.73 0.83 0.87
SRSF2mut 170 0.67 0.7 0.82 0.84
CELLCYCLEmut 170 0.7 0.63 0.91 0.82
TP53mut 237 0.79 0.66 0.98 0.82
GENDER 170 0.61 0.66 0.75 0.82
RASmut 234 0.84 0.66 0.99 0.82
ABNCHR 310 0.67 0.63 0.85 0.81
MDS_vs_MDS/MPN 170 0.75 0.69 0.88 0.81
DNACHROMATINmut 170 0.7 0.64 0.81 0.8
RUNX1mut 170 0.68 0.66 0.81 0.78
IDHmut 210 0.83 0.64 0.94 0.78
OS2y 234 0.7 0.62 0.9 0.77
TRISOMY8 170 0.63 0.6 0.7 0.75
RASPATHWAYmut 234 0.77 0.61 0.98 0.74
DNMT3Amut 153 0.63 0.59 0.75 0.71
AML2y 234 0.77 0.55 0.97 0.65
ComplexKaryotype 170 0.63 0.55 0.8 0.64
SECONDARY_MDS 237 0.76 0.52 1 0.58
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
b
Cor Train
Cor Test
AGE 0.51 *** 0.63 ***
IPSSR 0.76 *** 0.65 ***
c
Monosomy 7
Sensitivity
1 - Specificity
Dataset
TestProb 0.684
TrainProb 0.734
Monosomy 7
Dataset
TestProb 0.886
TrainProb 0.883
d
ASXL1 mutation
Dataset
TestProb 0.735
TrainProb 0.791
ASXL1 mutation
Dataset
TestProb 0.913
TrainProb 0.923
e
-Log10 P-value
Spearman Correlation Coefficient
Cor
0.0
0.2
0.4
0.6
Mutation
ASXL1 VAF
Cell cycle pathway mutation count
Cell differentiation pathway mutation count
Chromatin pathway mutation count
DNMT3A
IDH1/IDH2 VAF
KRAS/NRAS VAF
RAS pathway mutation count
RUNX1 VAF
SF3B1 VAF
Spliceosome mutation
SRSF2 VAF
STAG2 VAF
TET2 VAF
TP53 VAF
f
R = 0.66 , p = 0.00015
Proportion of ASXL1 mutation
Prediction probability of ASXL1 mutation
g
Prediction probability of mutations in RAS pathway
Number of mutations in RAS pathway
h
R = 0.76 , p = 4.7e-15
IPSSR
Prediction probability of IPSSR
R = 0.65 , p = 1.4e-05
IPSSR
Prediction probability of IPSSR

Fig 4


a
Pixel analysis
NHC analysis
NHC
Stroma
Lipid droplet
RBC
Nucleus Area
Nucleus Eccentricity
Nucleus Hematoxylin mean
Nucleus Hematoxylin std dev
Nucleus Eosin mean
Nucleus Eosin std dev
ComplexKaryotype
IPSSR
STAG2mut
RUNX1mut
TET2mut
DNMT3Amut
IDHmut
SPLICINGmut
CELLCYCLEmut
DNACHROMATINmut
AZA
CELLDIFFERENTIATIONmut
SRSF2mut
AML2y
OS2y
RASmut
RASPATHWAYmut
TP53mut
ABNCHR
SECONDARY_MDS
DEL7q
MONOSOMY7
DEL20q
TRISOMY8
SF3B1mut
ASXL1mut
GENDER
AGEdg_class
DEL5q
MDS_vs_MDS/MPN
1
0.5
0
-0.5
-1
b
MDS/MPN
MDS
H&E
Activation
c
Wilcoxon
p = 0.00022
NHC proportion (%)
1.00
0.75
0.50
0.25
0.00
MDS
MDS/MPN
Diagnosis
d
Del5q
Non-del5q
Hypolobular
megakaryocyte
e
Wilcoxon
p = 0.017
Stroma proportion (%)
75
50
25
0
Del
Norm
Del5q
f
Wilcoxon
p = 0.00034
Nucleus Eccentricity
0.81
0.78
0.75
0.72
0.69
Del
Norm
Del5q

Supplementary Table 1. Control subject characteristics

Sex	Age at sampling	Bone marrow biopsy indication
male	69	Persistent erythrocytosis, leukocytosis and thrombocytosis
female	57	Persistent thrombocytopenia
male	60	Persistent thrombocytosis and leukocytosis
male	61	Persistent thrombocytopenia
male	44	Persistent eosinofilia
male	65	Persistent anemia
female	40	Persistent thrombocytosis and leukocytosis
female	44	Persistent thrombocytosis and erythrocytosis
female	82	Persistent thrombocytosis
male	54	Persistent erythrocytosis
female	43	Persistent anemia and leukopenia

*Control subjects have not been diagnosed with any hematologic malignancy, chronic infection nor autoimmune disorder and are no longer under active surveillance at the Department of Hematology.

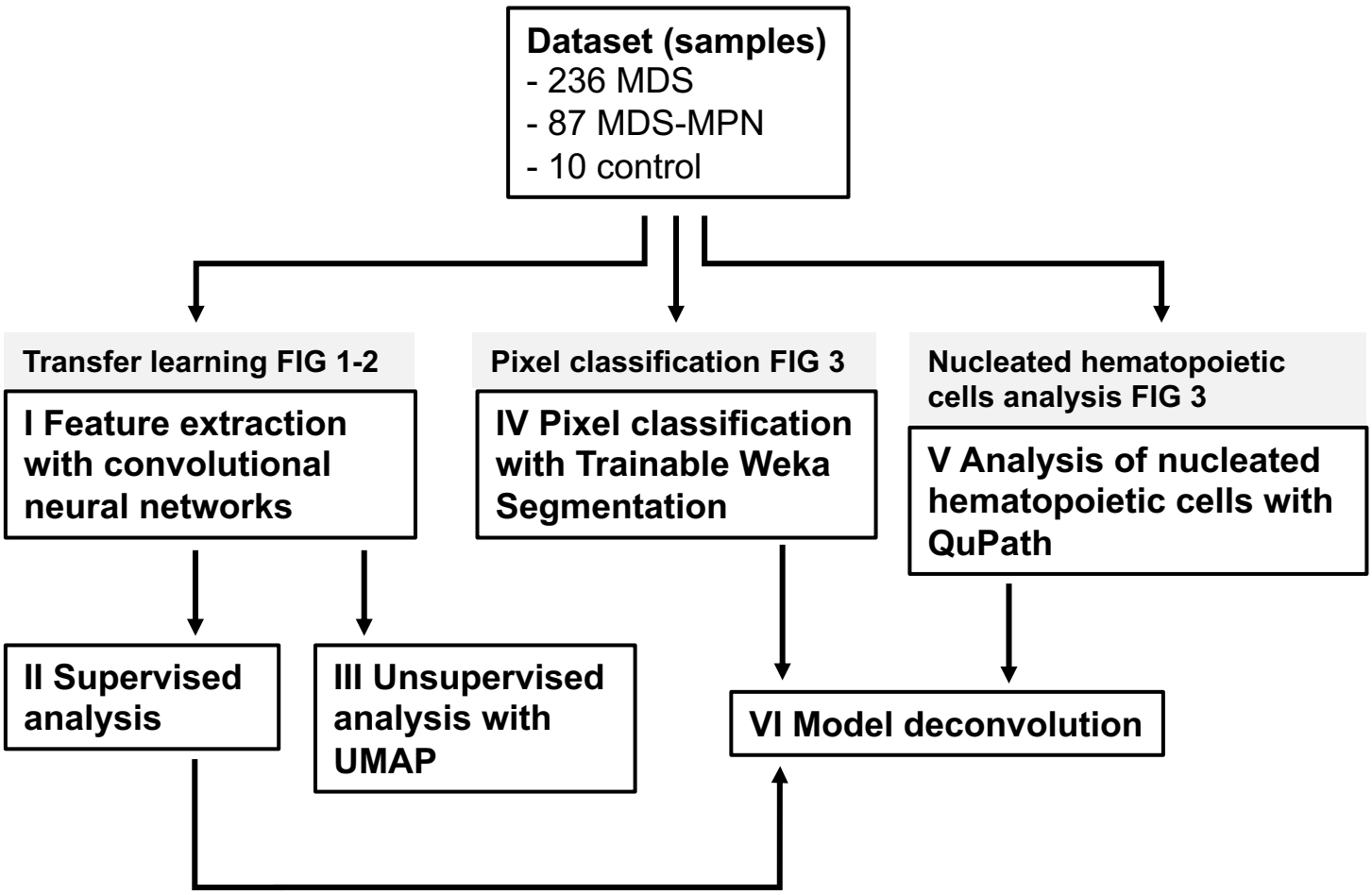
Supplementary Table 2. Catalog of pixel classification and intercellular metrics from segmented white blood cells.

Abbreviation	Definition
Stroma	Proportion of fibrous stroma and bone trabeculae from the bone marrow tissue microarray spot
WBC	Proportion of white blood cells from the bone marrow tissue microarray spot
RBC	Proportion of red blood cells from the bone marrow tissue microarray spot
Lipid droplet	Proportion of lipid droplets from the bone marrow tissue microarray spot
Nucleus_Area	Area of the nucleus from segmented cells
Nucleus_Perimeter	Perimeter of the nucleus from segmented cells
Nucleus_Circularity	Circularity of the nucleus from segmented cells
Nucleus_Max_caliper	Maximum caliper of the nucleus from segmented cells
Nucleus_Min_caliper	Minimum caliper of the nucleus from segmented cells
Nucleus_Eccentricity	Eccentricity of the nucleus from segmented cells
Nucleus_Hematoxylin_OD_mean	Optical density mean of deconvoluted hematoxylin stain of the nucleus from segmented cells
Nucleus_Hematoxylin_OD_sum	Optical density sum of deconvoluted hematoxylin stain of the nucleus from segmented cells
Nucleus_Hematoxylin_OD_std_dev	Optical density standard deviation of deconvoluted hematoxylin stain of the nucleus from segmented cells
Nucleus_Hematoxylin_OD_max	Optical density maximum value of deconvoluted hematoxylin stain of the nucleus from segmented cells
Nucleus_Hematoxylin_OD_min	Optical density minimum value of deconvoluted hematoxylin stain of the nucleus from segmented cells
Nucleus_Hematoxylin_OD_range	Optical density range of deconvoluted hematoxylin stain of the nucleus from segmented cells
Nucleus_Eosin_OD_mean	Optical density mean of deconvoluted eosin stain of the nucleus from segmented cells
Nucleus_Eosin_OD_sum	Optical density sum of deconvoluted eosin stain of the nucleus from segmented cells
Nucleus_Eosin_OD_std_dev	Optical density standard deviation of deconvoluted eosin stain of the nucleus from segmented cells
Nucleus_Eosin_OD_max	Optical density maximum value of deconvoluted eosin stain of the nucleus from segmented cells
Nucleus_Eosin_OD_min	Optical density minimum value of deconvoluted eosin stain of the nucleus from segmented cells
Nucleus_Eosin_OD_range	Optical density range of deconvoluted hematoxylin stain eosin the nucleus from segmented cells
NucleustoCell_area_ratio	Ratio of the nucleus area to the cell area from segmented cells
Cell_Area	Area of segmented cells
Cell_Perimeter	Perimeter of segmented cells
Cell_Circularity	Circularity of segmented cells
Cell_Max_caliper	Maximum caliperof segmented cells
Cell_Min_caliper	Minimum caliper of segmented cells
Cell_Eccentricity	Eccentricity of segmented cells
Cell_Hematoxylin_OD_mean	Optical density mean of deconvoluted hematoxylin stain of segmented cells
Cell_Hematoxylin_OD_std_dev	Optical density standard deviation of deconvoluted hematoxylin stain of segmented cells
Cell_Hematoxylin_OD_max	Optical density maximum value of deconvoluted hematoxylin stain of segmented cells
Cell_Hematoxylin_OD_min	Optical density minimum value of deconvoluted hematoxylin stain of segmented cells

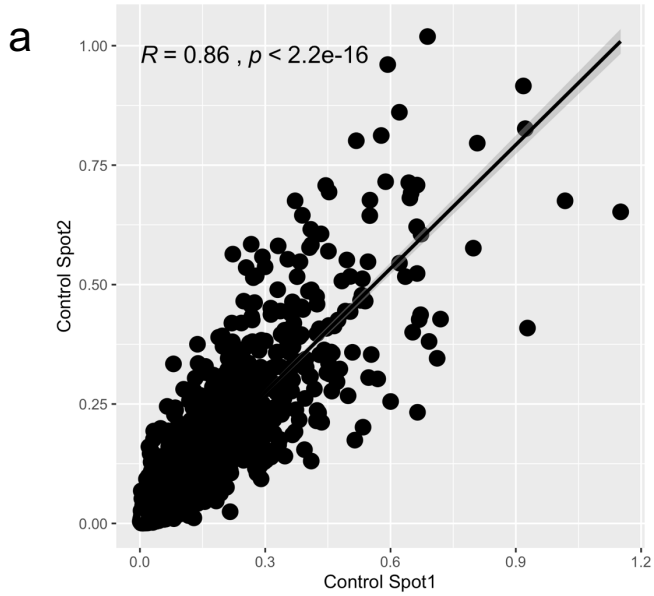
Cell_Eosin_OD_mean	Optical density mean of deconvoluted eosin stain of segmented cells
Cell_Eosin_OD_std_dev	Optical density standard deviation of deconvoluted eosin stain of segmented cells
Cell_Eosin_OD_max	Optical density maximum value of deconvoluted eosin stain of segmented cells
Cell_Eosin_OD_min	Optical density minimum value of deconvoluted eosin stain of segmented cells
Cytoplasm_Hematoxylin_OD_mean	Optical density mean of deconvoluted hematoxylin stain of the cytoplasm from segmented cells
Cytoplasm_Hematoxylin_OD_std_dev	Optical density standard deviation of deconvoluted hematoxylin stain of the cytoplasm from segmented cells
Cytoplasm_Hematoxylin_OD_max	Optical density maximum value of deconvoluted hematoxylin stain of the cytoplasm from segmented cells
Cytoplasm_Hematoxylin_OD_min	Optical density minimum value of deconvoluted hematoxylin stain of the cytoplasm from segmented cells
Cytoplasm_Eosin_OD_mean	Optical density mean of deconvoluted eosin stain of the cytoplasm from segmented cells
Cytoplasm_Eosin_OD_std_dev	Optical density standard deviation of deconvoluted eosin stain of the cytoplasm from segmented cells
Cytoplasm_Eosin_OD_max	Optical density maximum value of deconvoluted eosin stain of the cytoplasm from segmented cells
Cytoplasm_Eosin_OD_min	Optical density minimum value of deconvoluted eosin stain of the cytoplasm from segmented cells

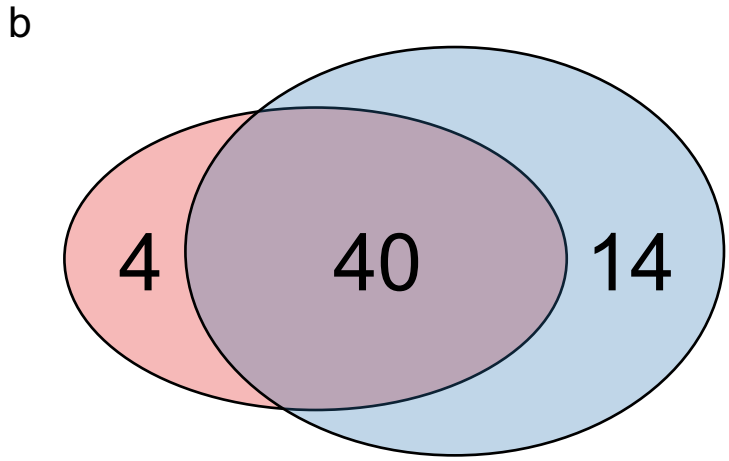
Supplementary Table 3. R Packages used in the analysis

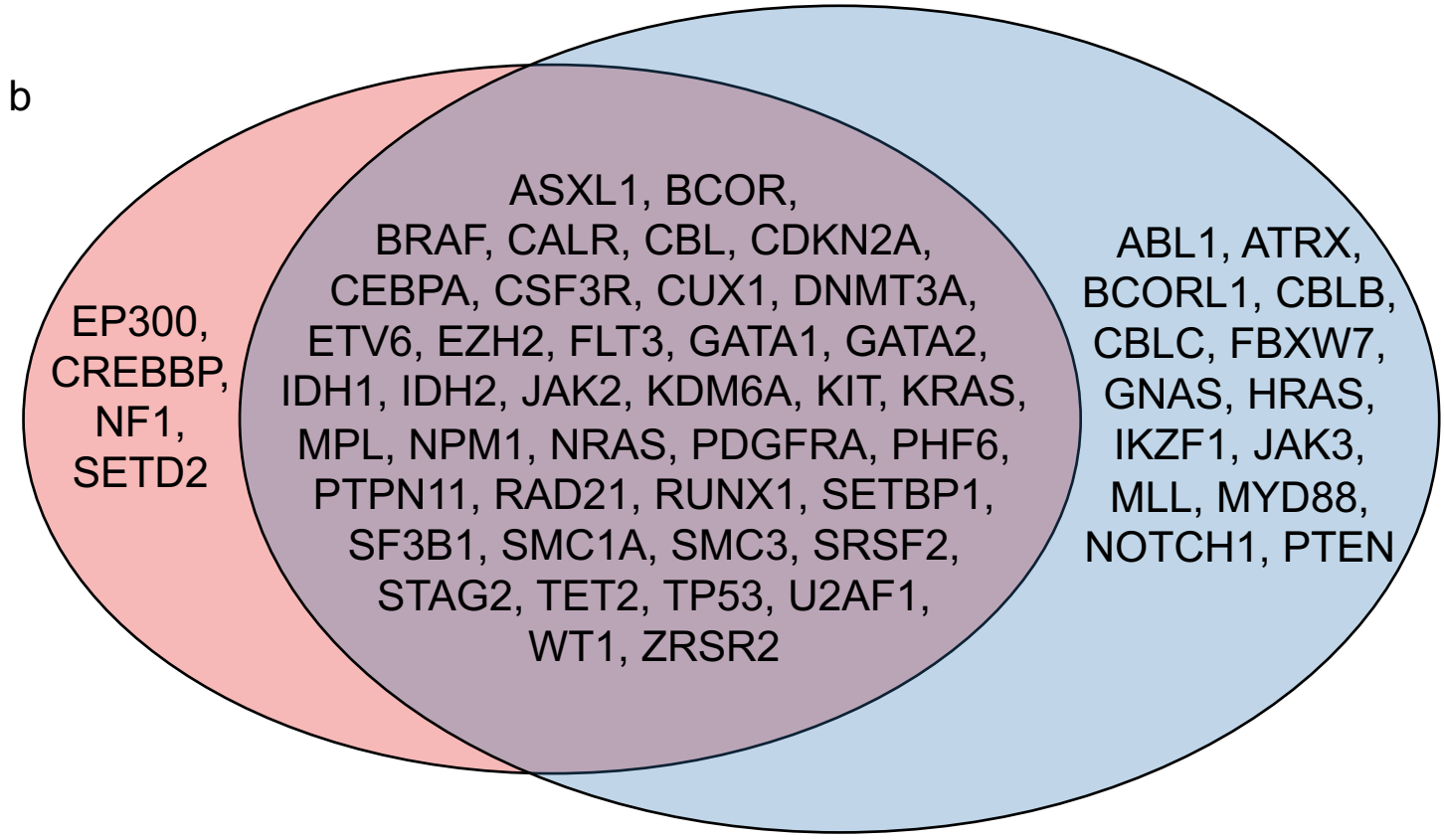
ComplexHeatmap 2.3.2	dendextend 1.9.0	ggdendro 0.1-20	corrplot 0.84	Hmisc 4.2-0	Formula 1.2-3	lattice 0.20-38	ggpubr 0.2
magrittr 1.5	ROCR 1.0-7	gplots 3.0.1	plotROC 2.2.1	survAUC 1.0-5	survival 2.43-1	pROC 1.13.0	biglasso 1.3-7
ncvreg 3.11-1	bigmemory 4.5.33	glmnet 2.0-16	foreach 1.4.7	writexl 1.0	reshape2 1.4.3	NbClust 3.0	factoextra 1.0.5
raster 2.8-4	sp 1.3-1	ggimage 0.2.4	RColorBrewer 1.1-2	readxl 1.1.0	uwot 0.0.0.9010	Matrix 1.2-15	Rphenograph 0.99.1
igraph 1.2.4.1	forcats 0.3.0	stringr 1.4.0	dplyr 0.8.3	purrr 0.3.3	readr 1.3.1	tidyr 1.0.0	tibble 2.1.3
ggplot2 3.2.1	tidyverse 1.2.1	keras 2.2.4	snakecase 0.9.2	rvcheck 0.1.3	lubridate 1.7.4	base64enc 0.1-3	class 7.3-14
circlize 0.4.8	backports 1.1.5	plyr 1.8.4	lazyeval 0.2.2	splines 3.5.1	tfruns 1.4	digest 0.6.23	htmltools 0.4.0
viridis 0.5.1	magick 2.0	gdata 2.18.0	fansi 0.4.0	checkmate 1.9.4	cluster 2.0.7-1	modelr 0.1.2	RcppParallel 4.4.4
colorspace 1.4-1	rvest 0.3.2	ggrepel 0.8.1	haven 2.0.0	xfun 0.10	crayon 1.3.4	jsonlite 1.6	bigmemory.sri 0.1.3
zeallot 0.1.0	iterators 1.0.12	glue 1.3.1	gtable 0.3.0	GetoptLong 0.1.8	kernlab 0.9-27	shape 1.4.4	prabclus 2.2-6
DEoptimR 1.0-8	scales 1.0.0	mvtnorm 1.0-8	Rcpp 1.0.3	viridisLite 0.3.0	htmlTable 1.13.2	clue 0.3-57	gridGraphics 0.3-0
reticulate 1.12	foreign 0.8-71	mclust 5.4.2	stats4 3.5.1	htmlwidgets 1.5.1	httr 1.4.1	fpc 2.1-11.1	acepack 1.4.1
modeltools 0.2-22	pkgconfig 2.0.3	flexmix 2.3-14	nnet 7.3-12	ggplotify 0.0.3	tidyselect 0.2.5	rlang 0.4.4	munsell 0.5.0
cellranger 1.1.0	tools 3.5.1	cli 2.0.0	generics 0.0.2	broom 0.5.2	yaml 2.2.0	knitr 1.25	robustbase 0.93-3
caTools 1.17.1.1	RANN 2.6.1	nlme 3.1-137	whisker 0.3-2	xml2 1.2.2	compiler 3.5.1	rstudioapi 0.10	png 0.1-7
stringi 1.4.3	trimcluster 0.1-2.1	tensorflow 1.13.1	vctrs 0.2.1	pillar 1.4.2	lifecycle 0.1.0	GlobalOptions 0.1.1	data.table 1.12.6
bitops 1.0-6	R6 2.4.1	latticeExtra 0.6-28	KernSmooth 2.23-15	gridExtra 2.3	codetools 0.2-15	MASS 7.3-51.1	gtools 3.8.1
assertthat 0.2.1	rjson 0.2.20	withr 2.1.2	parallel 3.5.1	diptest 0.75-7	hms 0.5.2	rpart 4.1-13	



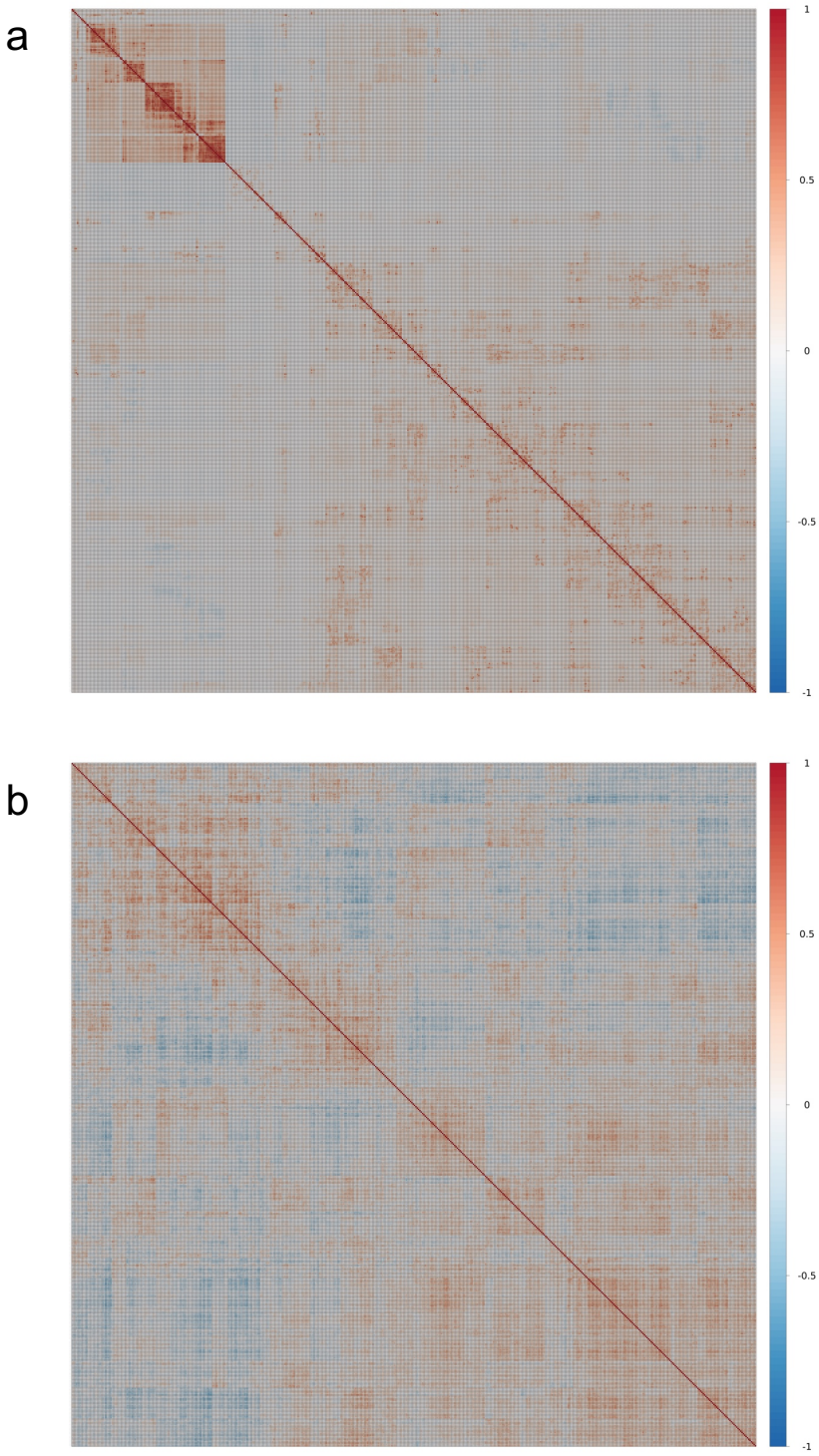
Supplementary Figure 1. Flowchart of the study design and the occurrence of each analysis in the manuscript.
I In transfer learning, a prior algorithm or its configurations are reused to develop a new machine learning algorithm. Here, we repurposed settings from the ImageNet dataset commonly used for training and evaluating image classification models.
In feature extraction, morphological patterns are identified and quantitated from H&E-stained bone marrow images with ImageNet-pretrained convolutional neural networks (VGG16 and Xception).
II In supervised analysis, clinical variables are predicted with numerical representation of tissue patterns (morphological features) using penalized regression models.
III In unsupervised analysis, bone marrow images are grouped based on their similarity. K-means analysis is used to group tissue microarray (TMA) images into clusters based on their morphological features. TMA images are split into 500 smaller images (tiles) to examine in more detail tissue patterns. Image tiles are grouped with Phenograph clustering analysis based on their morphological features.
IV In pixel classification, we classify each image pixel (0.22 mm) as a part of nucleated hematopoietic cells, stroma, lipid droplets or red blood cells. First, tissue areas of 10 training images are annotated into these four groups. Then, we train a Weka Segmentation algorithm (based on pixel classification) to identify these areas from all other images.
V Nucleated hematopoietic cells are identified and segmented by their nuclear border. Nucleus size, circularity and H&E dye intensity are quantitated from individual nuclei using the QuPath software.
VI In model deconvolution, we correlate results of pixel classification and nucleated hematopoietic cell analysis to the prediction probabilities of the supervised analysis. This helps us to understand which tissue and cell characteristics are associated with morphological patterns (combination of individual visual features) seen in certain patient subgroups.



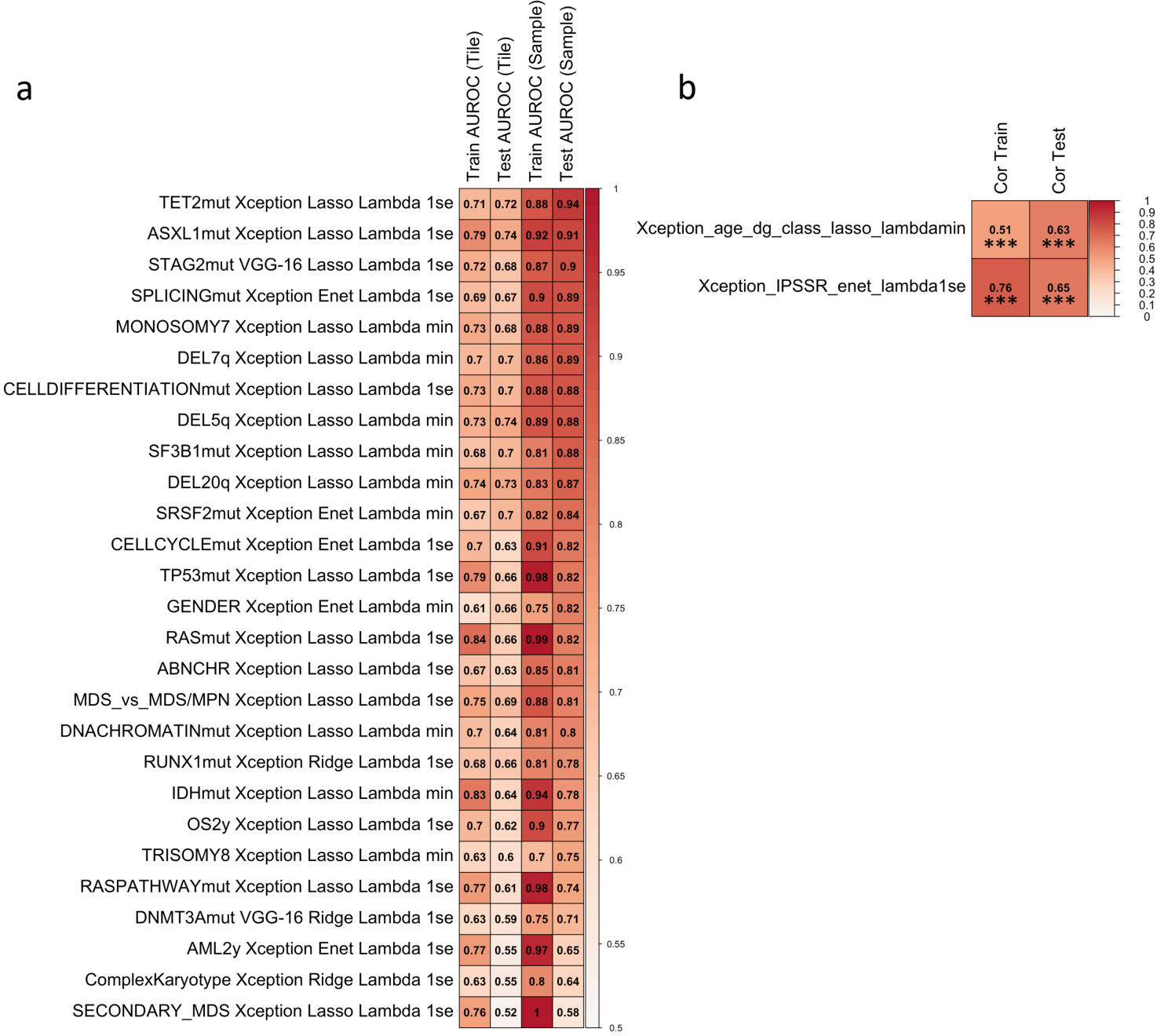




Supplementary Figure 2. (a) Scatter plot of visual features in duplicate control spots. The R correlation coefficient value has been computed with Spearman correlation. (b) Venn diagram visualizing number and (c) gene names included in myeloid amplicon panels used in the study. Gene names on red background are part of the HUSLAB myeloid panel (108 samples) and blue background of the Illumina TruSight panel (30 samples). Genes included in both panels were analyzed in the study.

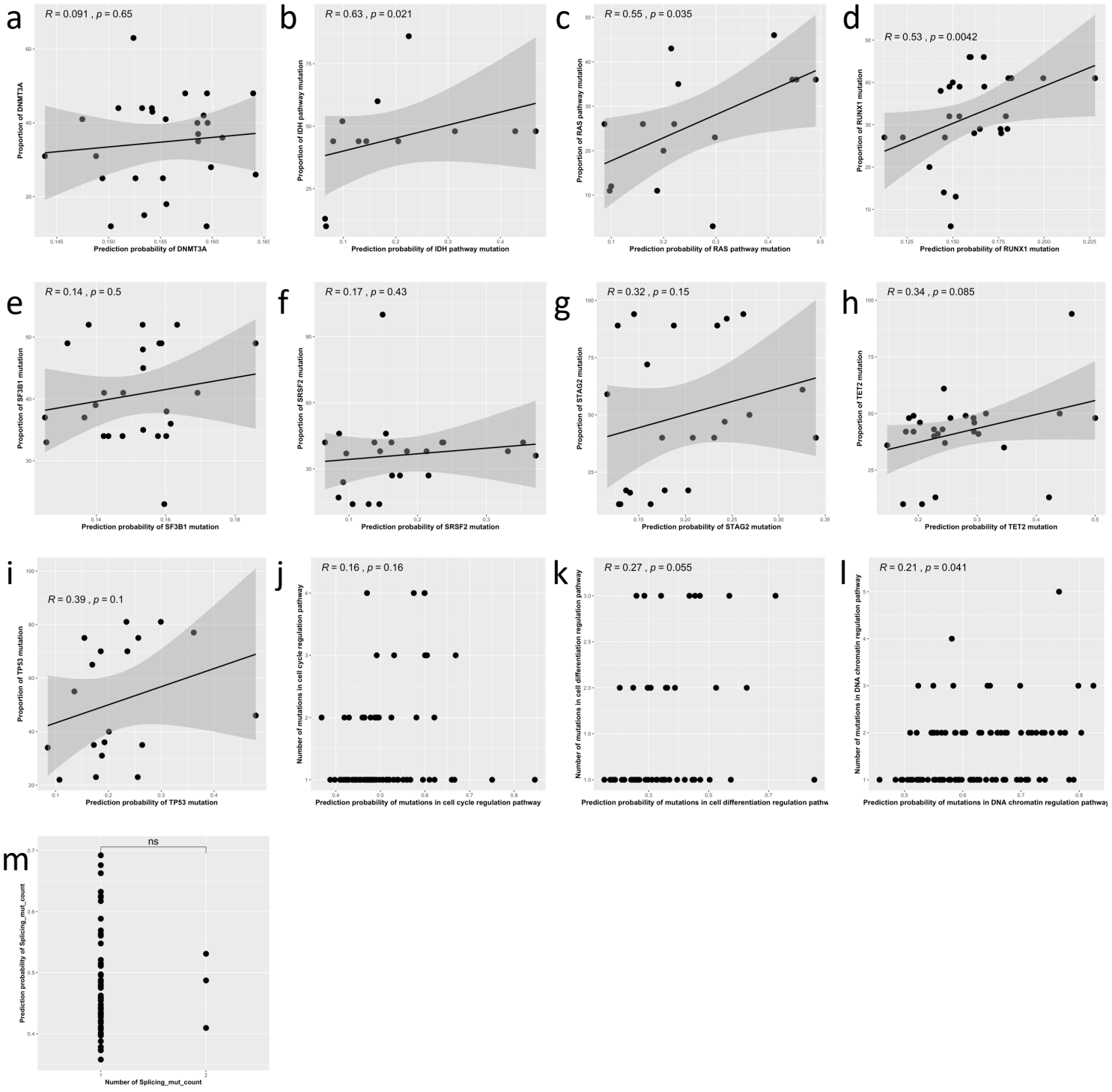


Supplementary Figure 3. (a) Correlation plot of the visual features extracted from the VGG16 and (b) Xception neural network configured with the ImageNet weights. For visualization purposes 25% (n=512) of the features used in the analysis are plotted here.

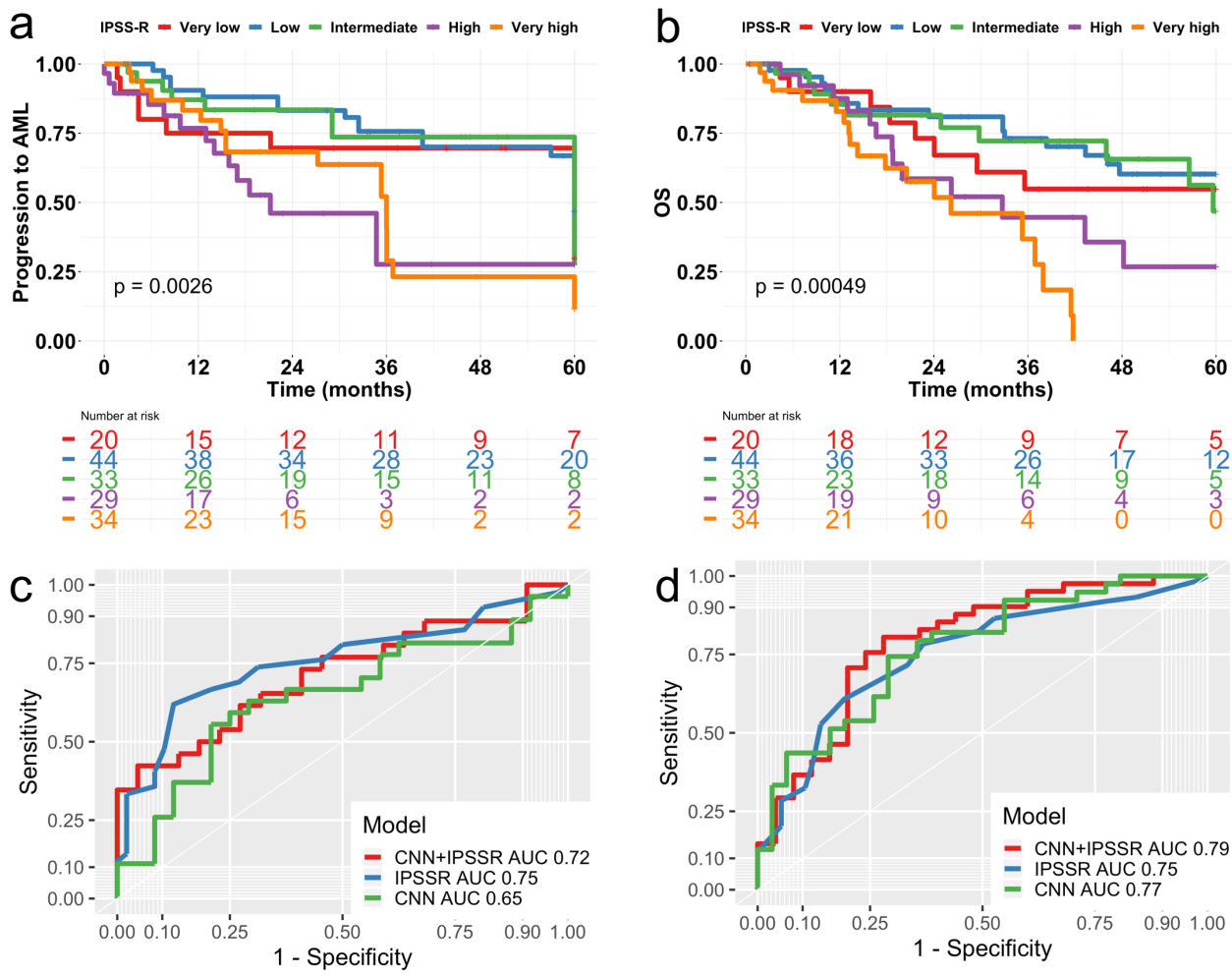


Supplementary Figure 4. Heatmap displaying area under the receiver operating characteristic curve (AUROC) values of (a) elastic net-regularized logistic regression models. Rows represent individual models and row names the predicted end point, the convolution neural network (Xception or VGG-16), elastic net alpha penalization (Lasso, Ridge or Enet) and lambda value (Lambda min or 1se). The left-most column shows the number of samples included in the analysis. The following two columns inform the AUROC values of the models in the training (2/3) and test (1/3) dataset in the tile-level images and the last two columns at the tissue microarray (TMA) spot-level. The color grading reflects AUROC value [0.5-1.0]. (b) Similar plot for elastic net-regularized linear regression models. Rows represent individual models and row names the predicted end point, the convolution neural network (Xception or VGG-16), elastic net alpha penalization (Lasso, Ridge or Enet) and Lambda value (Lambda min or 1se). The left-most column shows the correlation value in the training dataset and right-mostin the test dataset. The color grading reflects Spearman correlation values [0.0-1.0]. All correlations were statistically significant (*** p<0.001).

Abbreviations. SPLICINGmut: spliceosome mutations. CELLDIFFERENTIATIONmut: mutation in genes regulating cell differentiation. CELLCYCLEmut: mutation in genes regulating cellcycle. RASmut: mutation in *NRAS* or *KRAS*. ABNCHR: presence of any abnormal chromosome. IDHmut: mutation in *IDH1* or *IDH2*. OS2y: overall survival event within 2 years of follow-up. RASPATHWAYmut: mutation in genes regulating RAS pathway. AML2y: progression to AML within 2 years of follow-up.



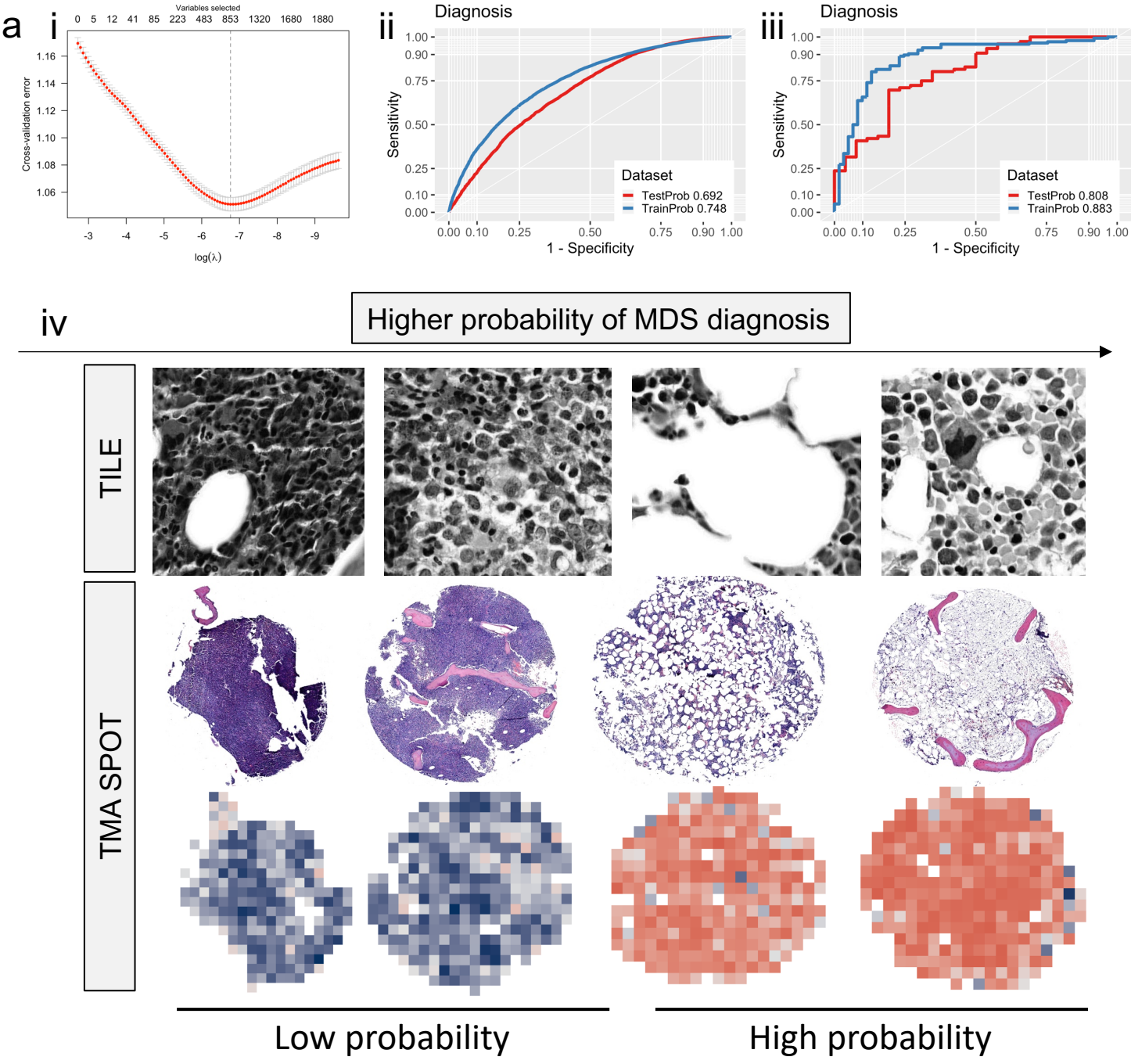
Supplementary Figure 5. Association between predicted and detected mutation occurrence. Linear regression between prediction probability of (a) *DNMT3A*, (b) *IDH1/IDH2*, (c) *KRAS/NRAS*, (d) *RUNX1*, (e) *SF3B1*, (f) *SRSF2*, (g) *STAG2*, (h) *TET2* and (i) *TP53* mutation (x-axis) and its detected variant allele frequency (y-axis). (h) Linear regression for predicted (j) cell cycle, (k) cell differentiation and (l) DNA chromatin structure regulation pathway mutation probability and detected frequency of altered genes in respective pathway. The R correlation coefficient value has been computed with Spearman correlation. (m) Wilcoxon comparison for predicted spliceosome regulation pathway mutation probability and detected frequency of altered genes in the *RAS* pathway.



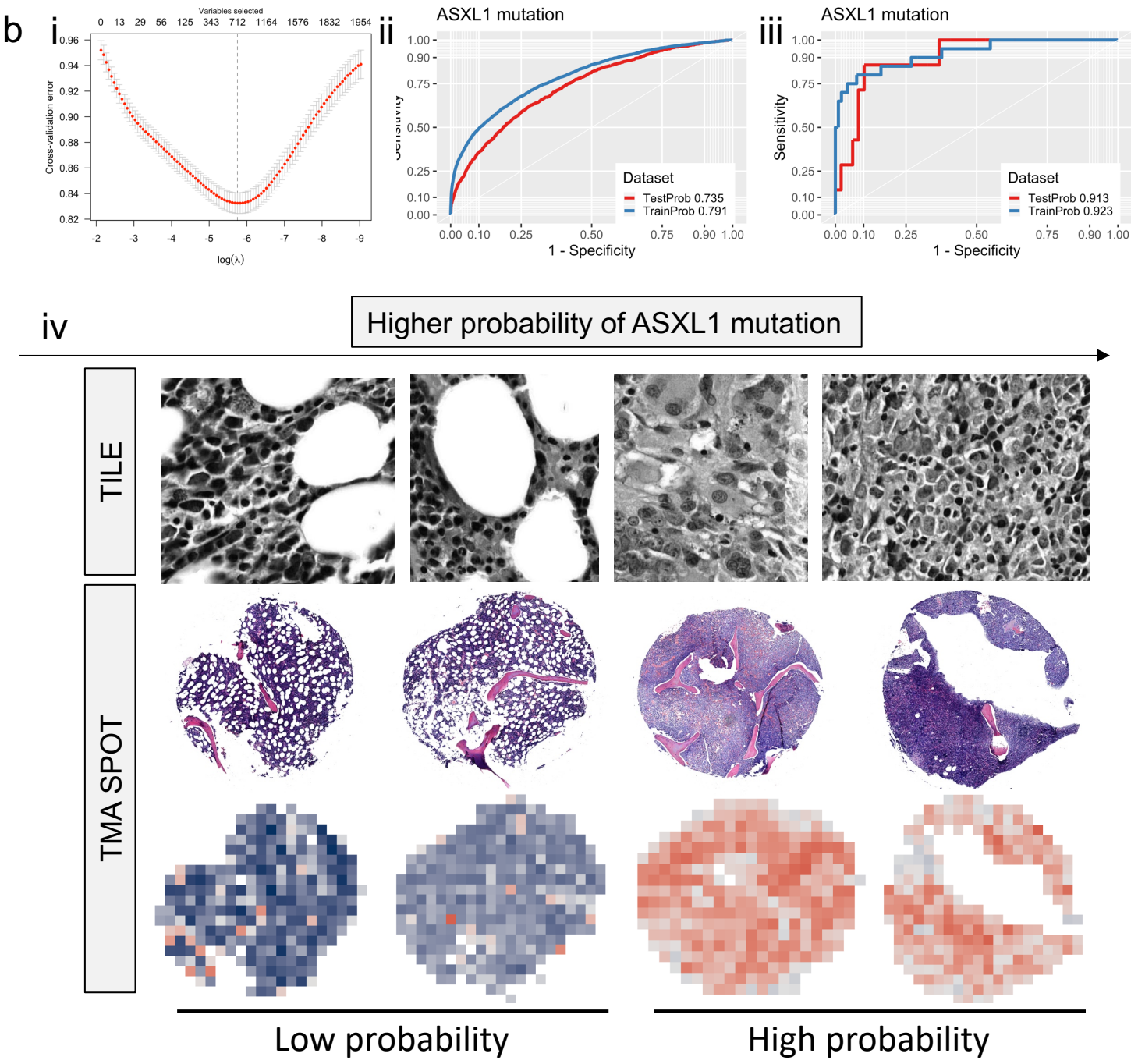
Supplementary Figure 6. (a) Progression to AML and (b) overall survival stratified with IPSSR. Kaplan-Meier curves are compared with the log-rank test. (c-d) Receiver operating characteristic (ROC) curves for models developed with features extracted from convolutional neural networks (CNN), Revised International Prognostic Scoring System (IPSSR) or their combination (CNN+IPSSR) to predict (c) progression to AML in 2 years or (d) overall survival in 2 years. AUROCs have been computed for CNN and CNN+IPSSR using only the test set and for IPSSR using both training and validation sets.

Supplementary Figure 7. Model reports for prediction of (a) MDS vs. MDS/MPN, (b) *ASXL1* mutation, (c) chromosome 5q deletion, (d) abnormal karyotype, (e) progression to AML in 2 years, (f) mutation in genes regulating cell cycle, (g) mutation in genes regulating cell differentiation, (h) complex karyotype, (i) chromosome 7q deletion, (j) chromosome 20q deletion, (k) mutation in genes regulating DNA chromatin structure, (l) *DNMT3A* mutation, (m) female gender, (n) *IDH1/IDH2* mutation, (o) chromosome 7q monosomy, (p) overall survival in 2 years, (q) *KRAS/NRAS* mutation, (r) mutation in genes regulating *RAS* pathway, (s) *RUNX1* mutation, (t) secondary MDS, (u) *SF3B1* mutation, (v) mutation in genes regulating splicesome, (w) *SRSF2* mutation, (x) *STAG2* mutation, (y) *TET2* mutation, (z) *TP53* mutation, and (aa) chromosome 8 trisomy.

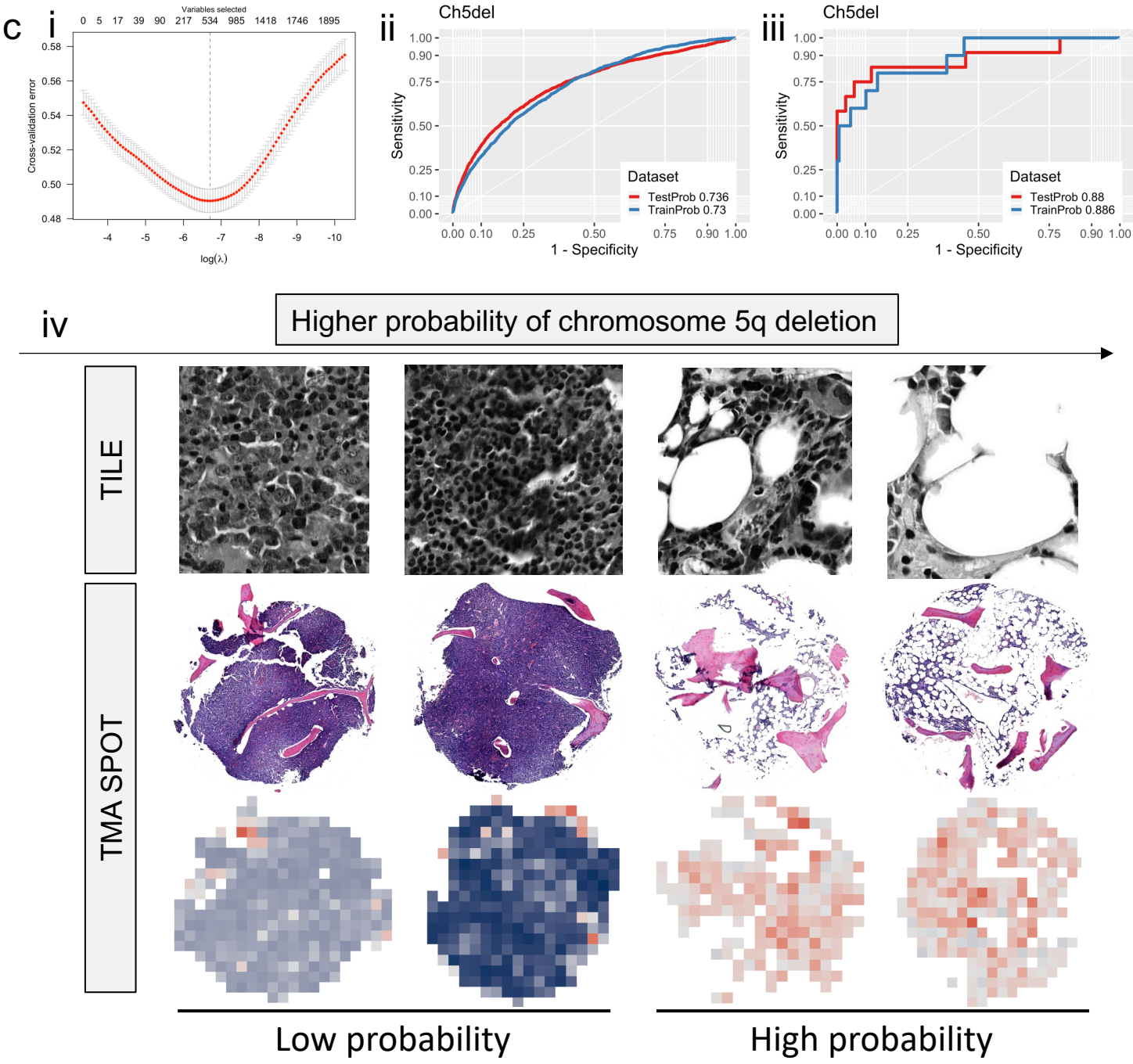
Reports include plots for (i) Variable numbers of the elastic net regression models with least crossvalidation error and best-performing penalization level. (ii) The plot with highest area under the receiver operating characteristic (AUROC) curve to evaluate tile-level label and (iii) TMA spot level label prediction. (iv) Two representative tile-level and TMA spot level images and activation maps visualizing the prediction probability.



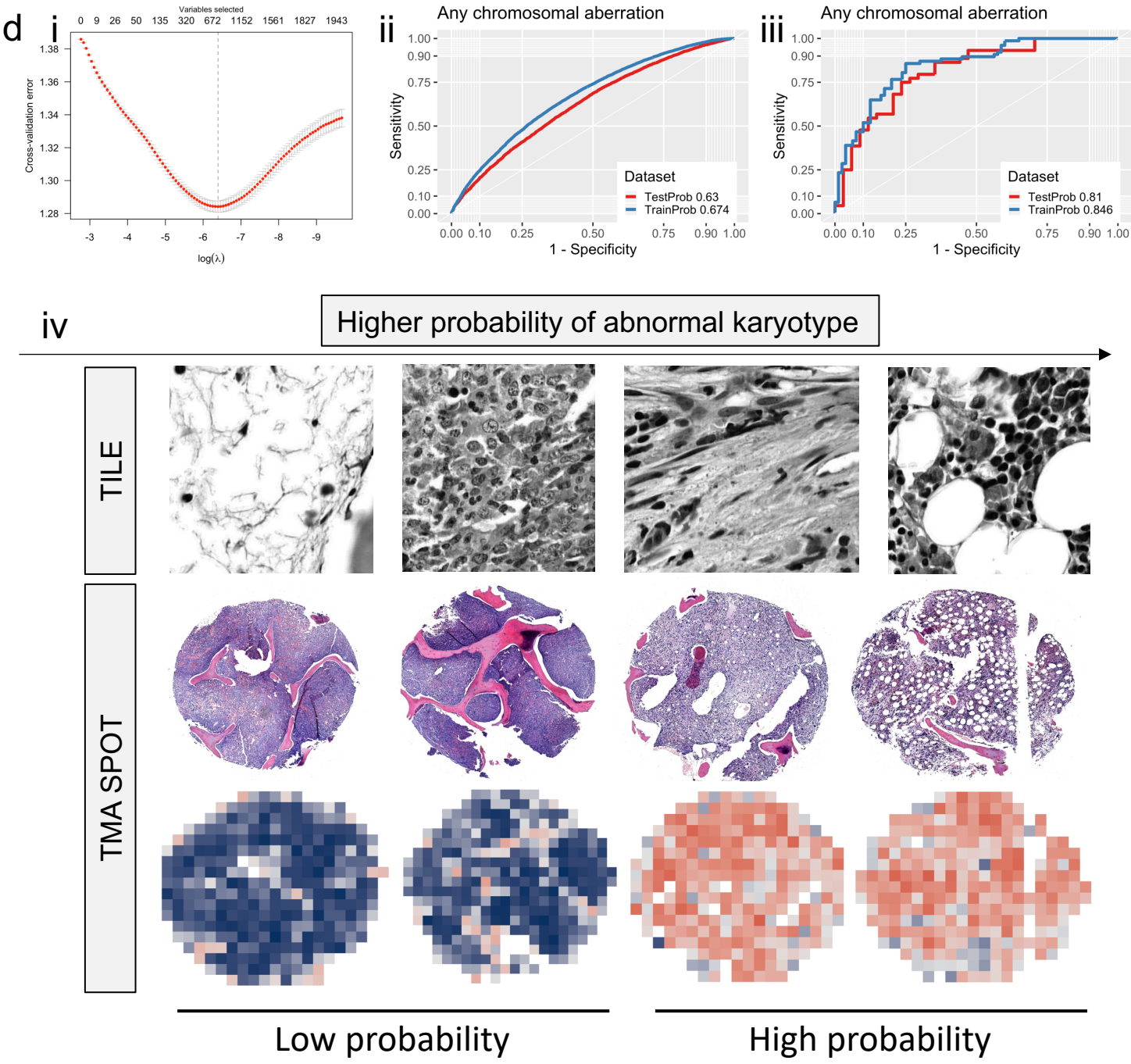
Supplementary Figure 7a.



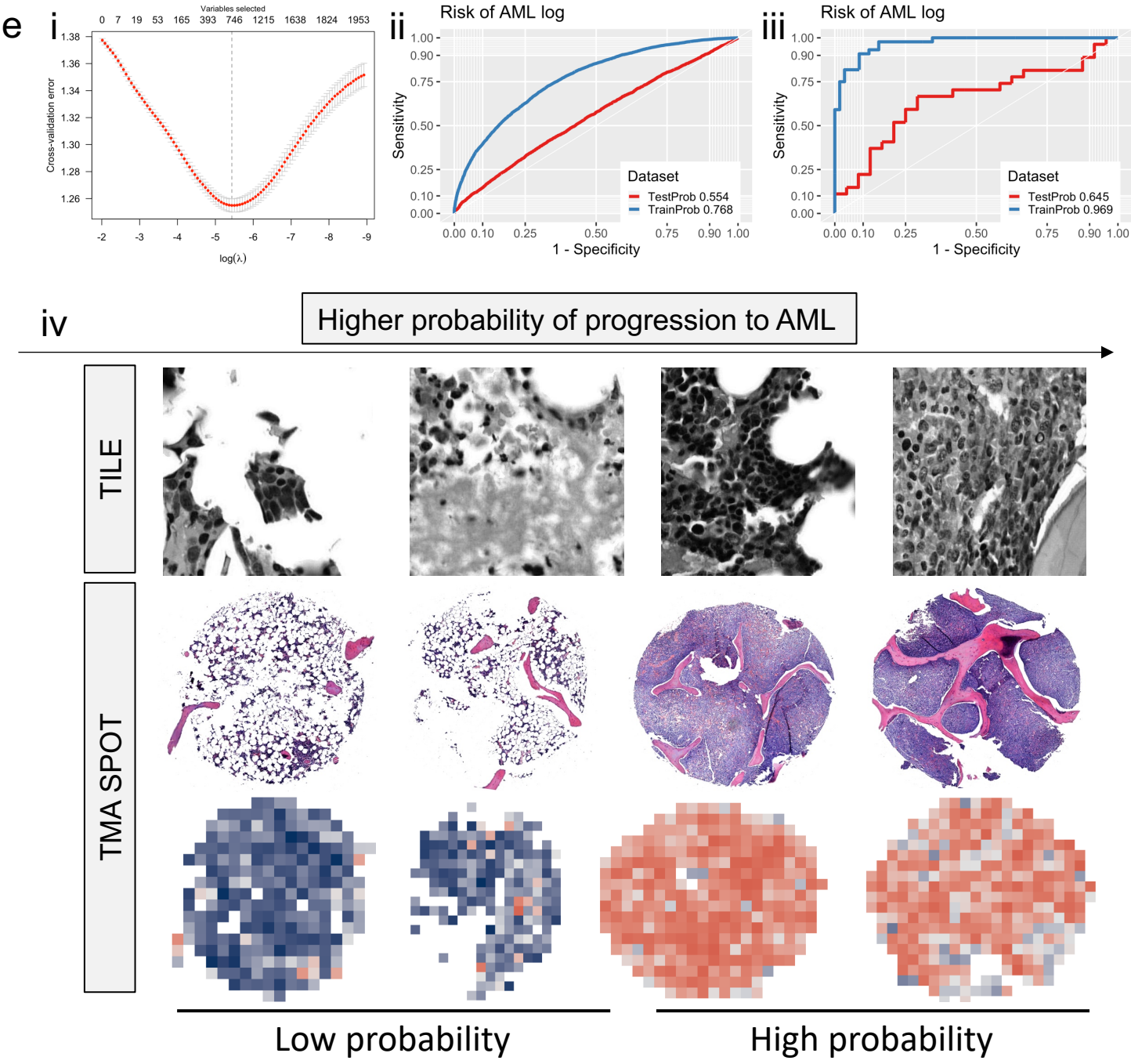
Supplementary Figure 7b.



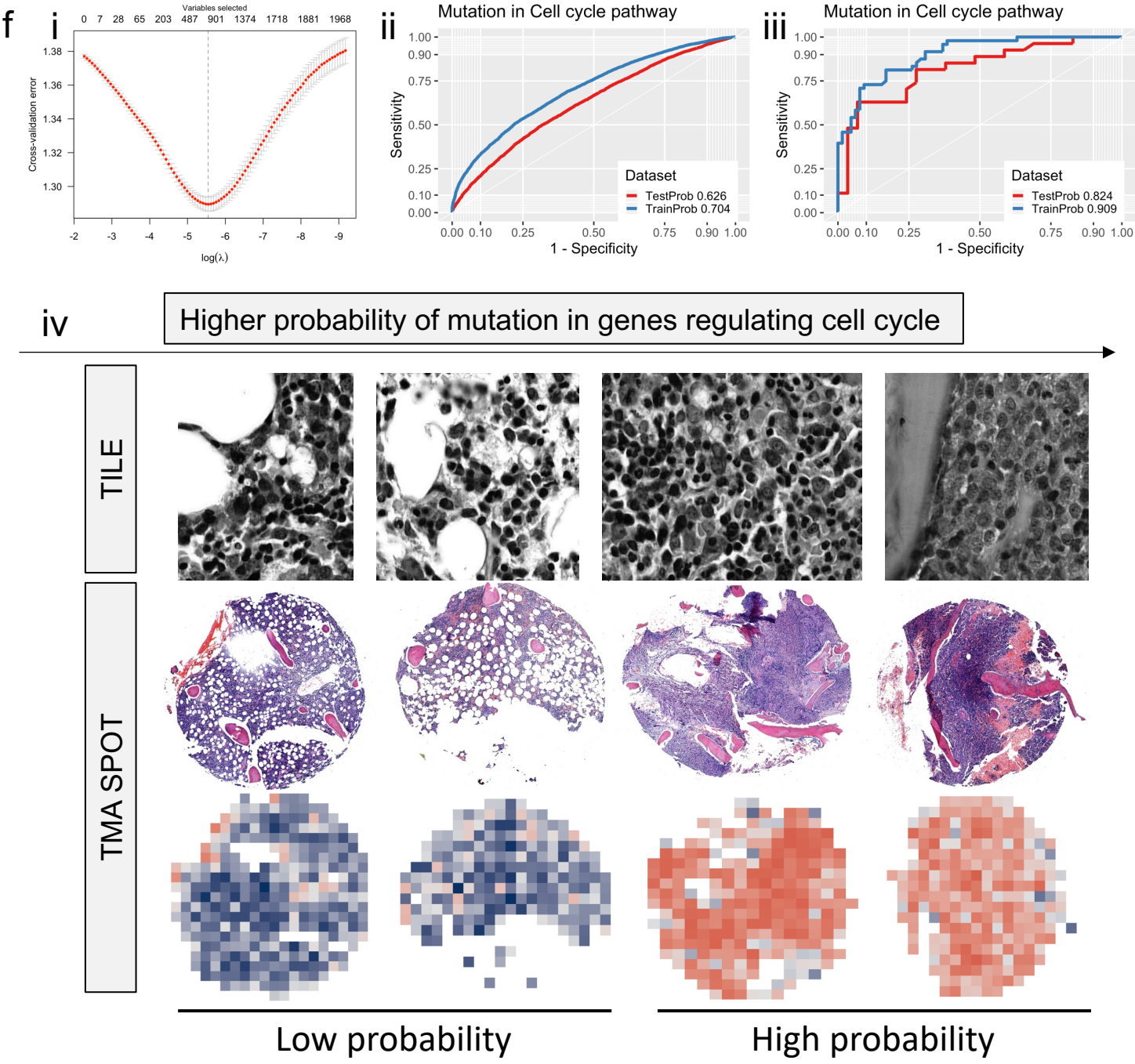
Supplementary Figure 7c.



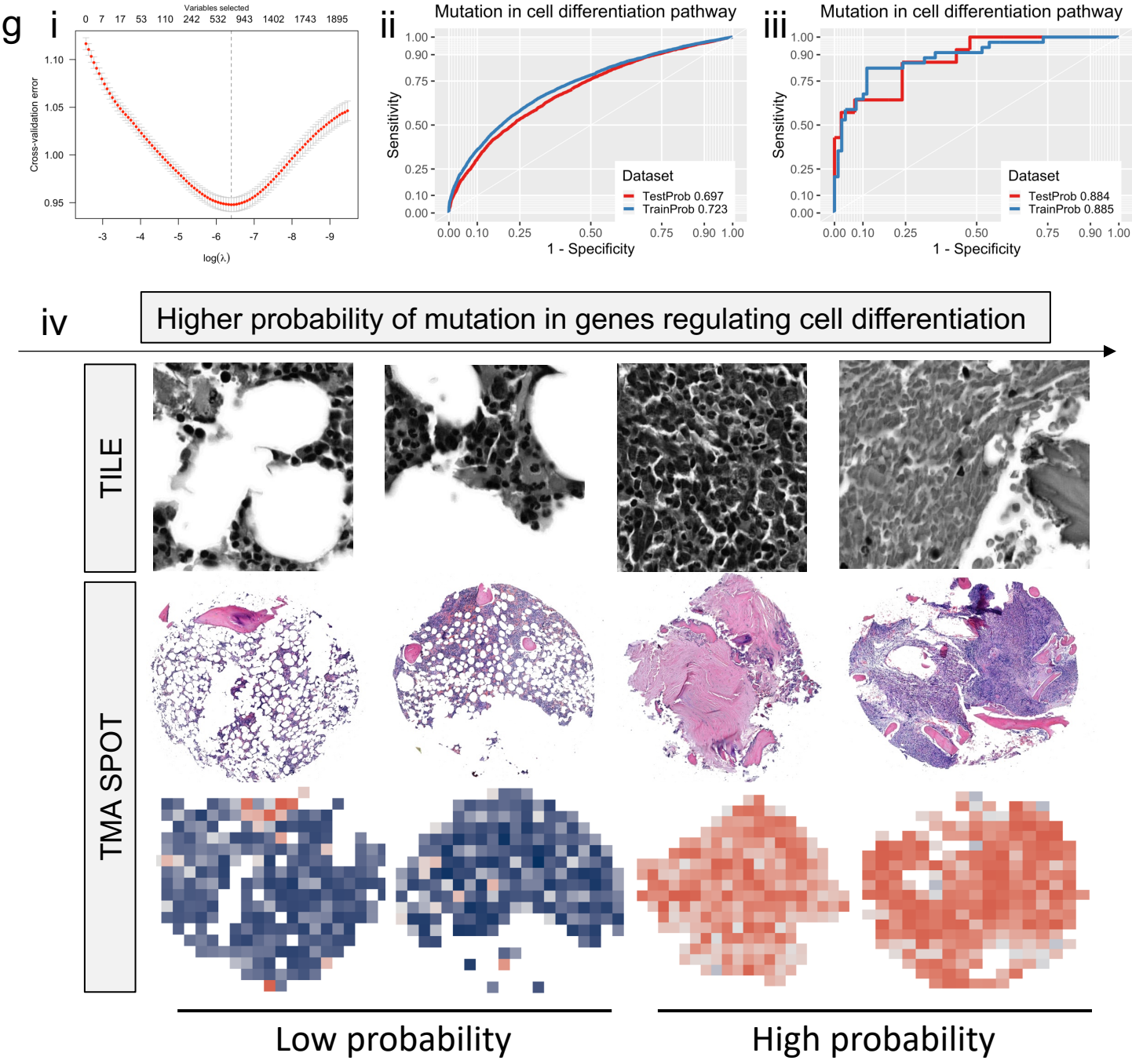
Supplementary Figure 7d.



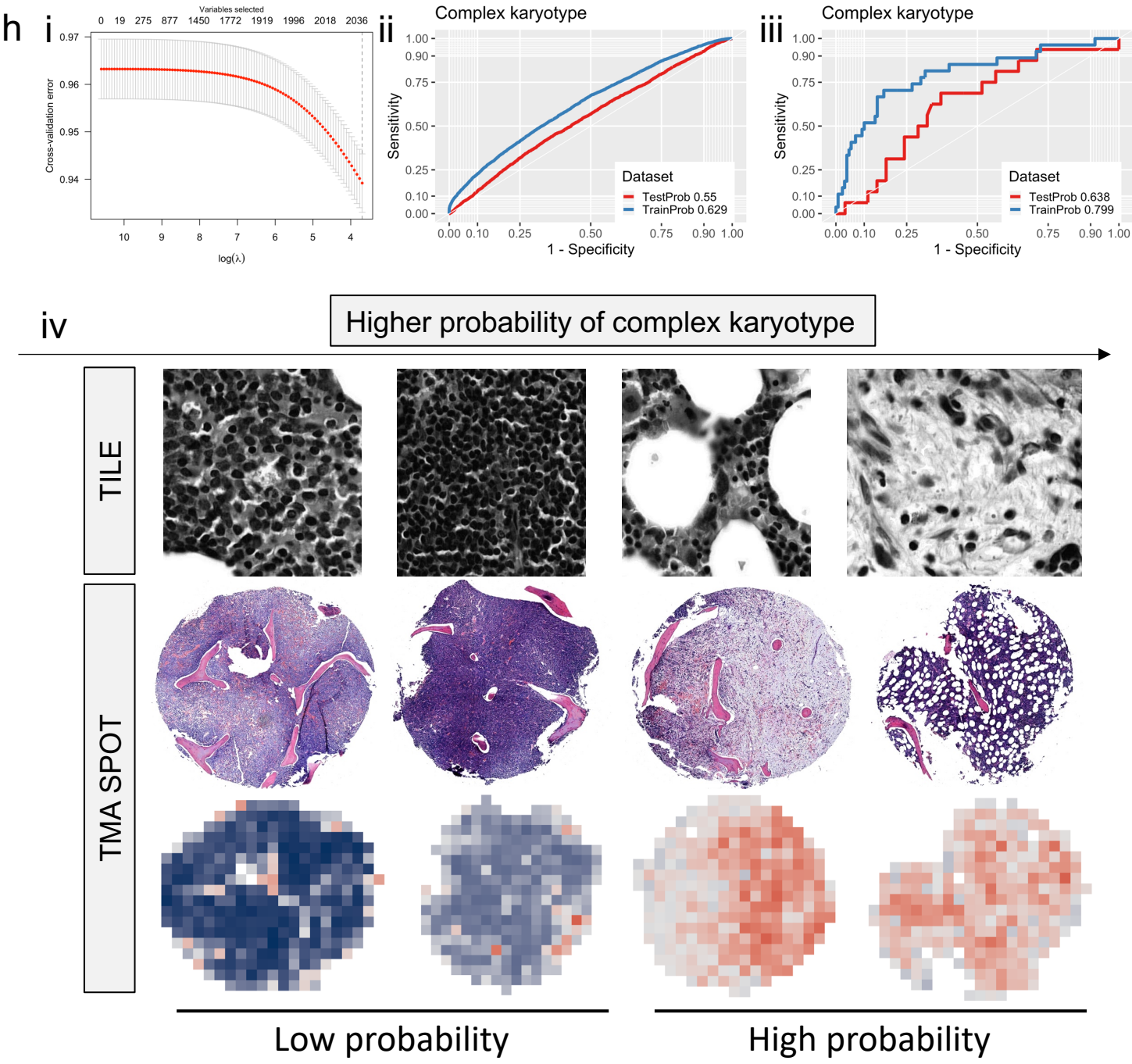
Supplementary Figure 7e.



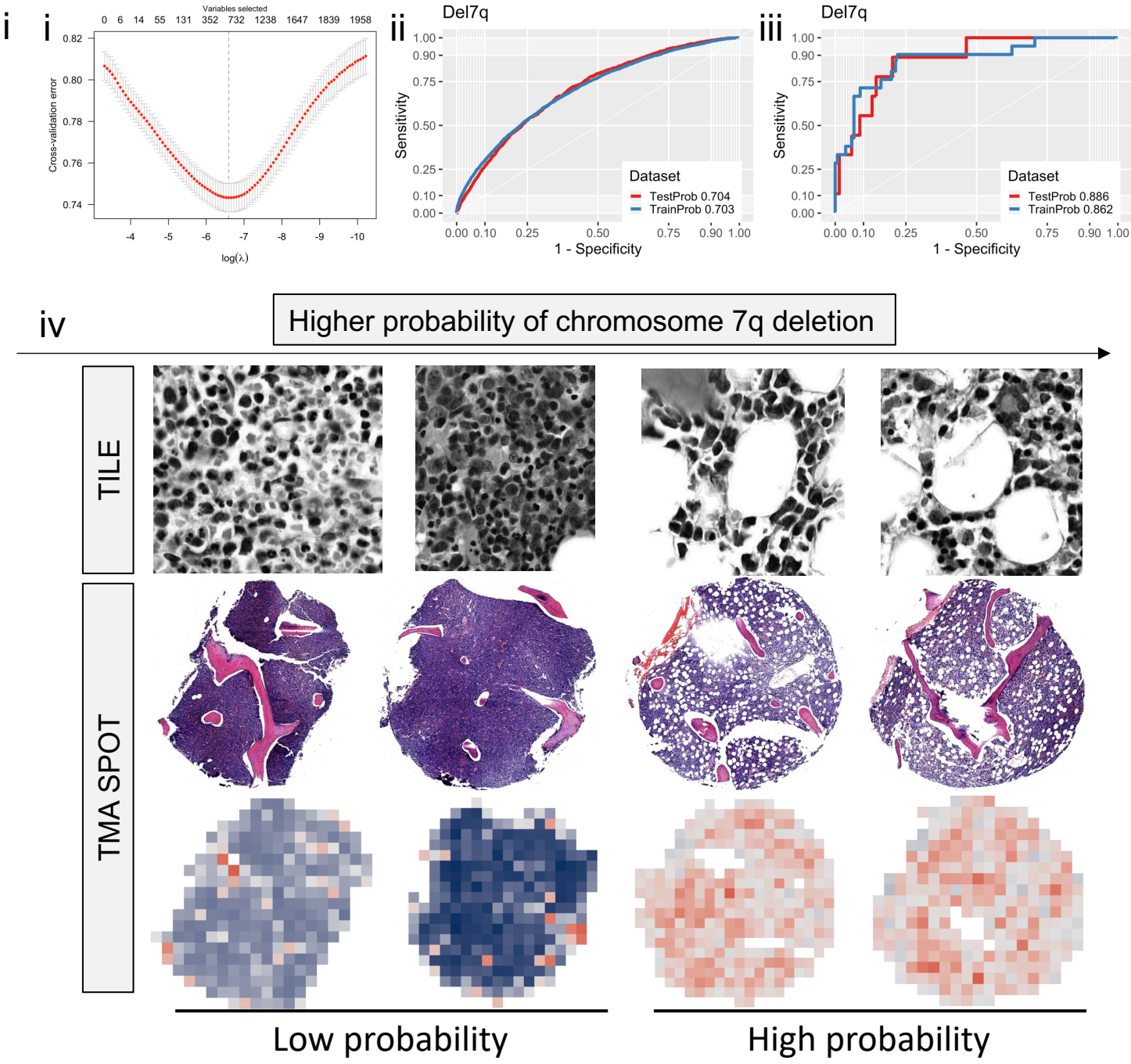
Supplementary Figure 7f.



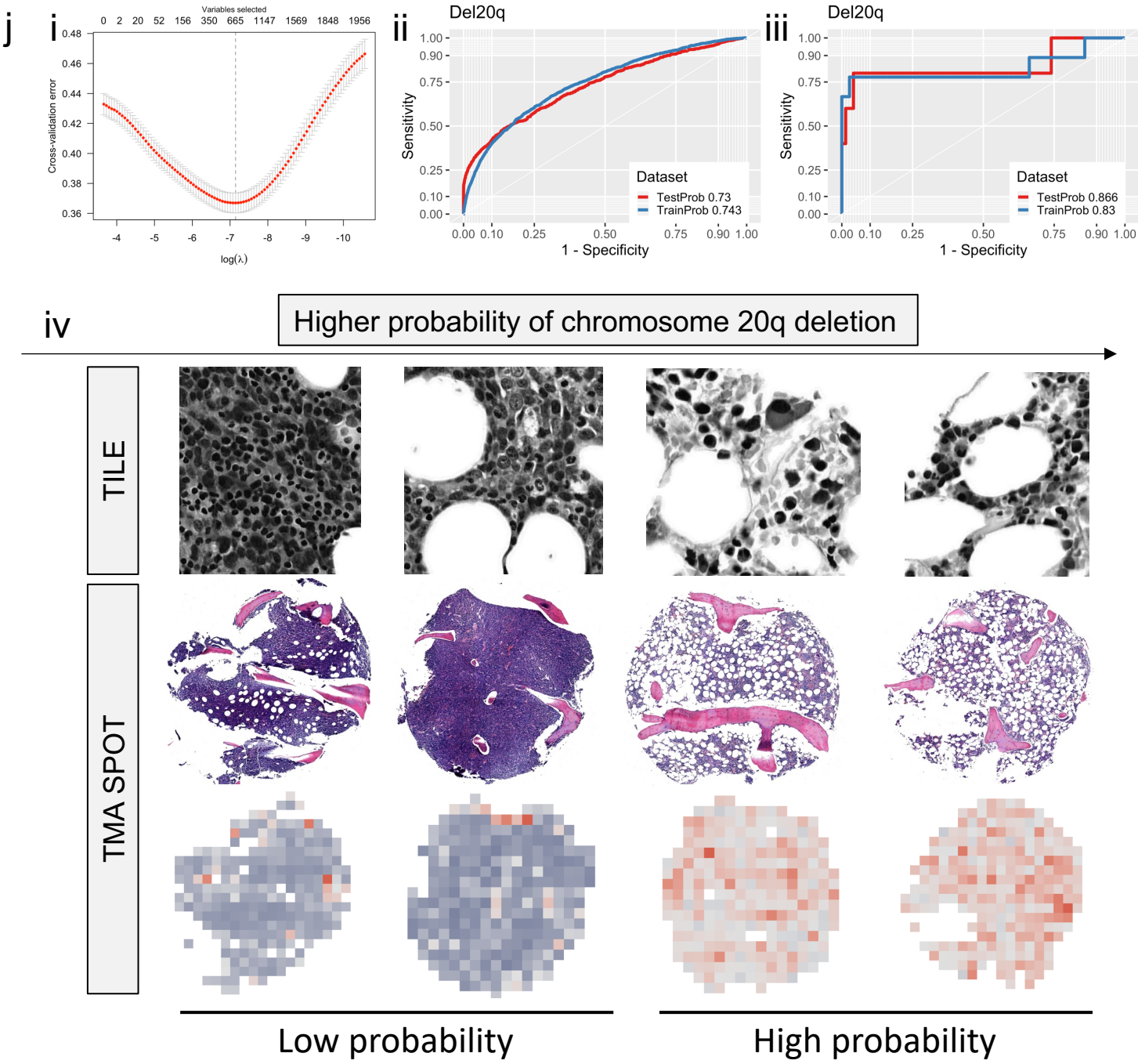
Supplementary Figure 7g.



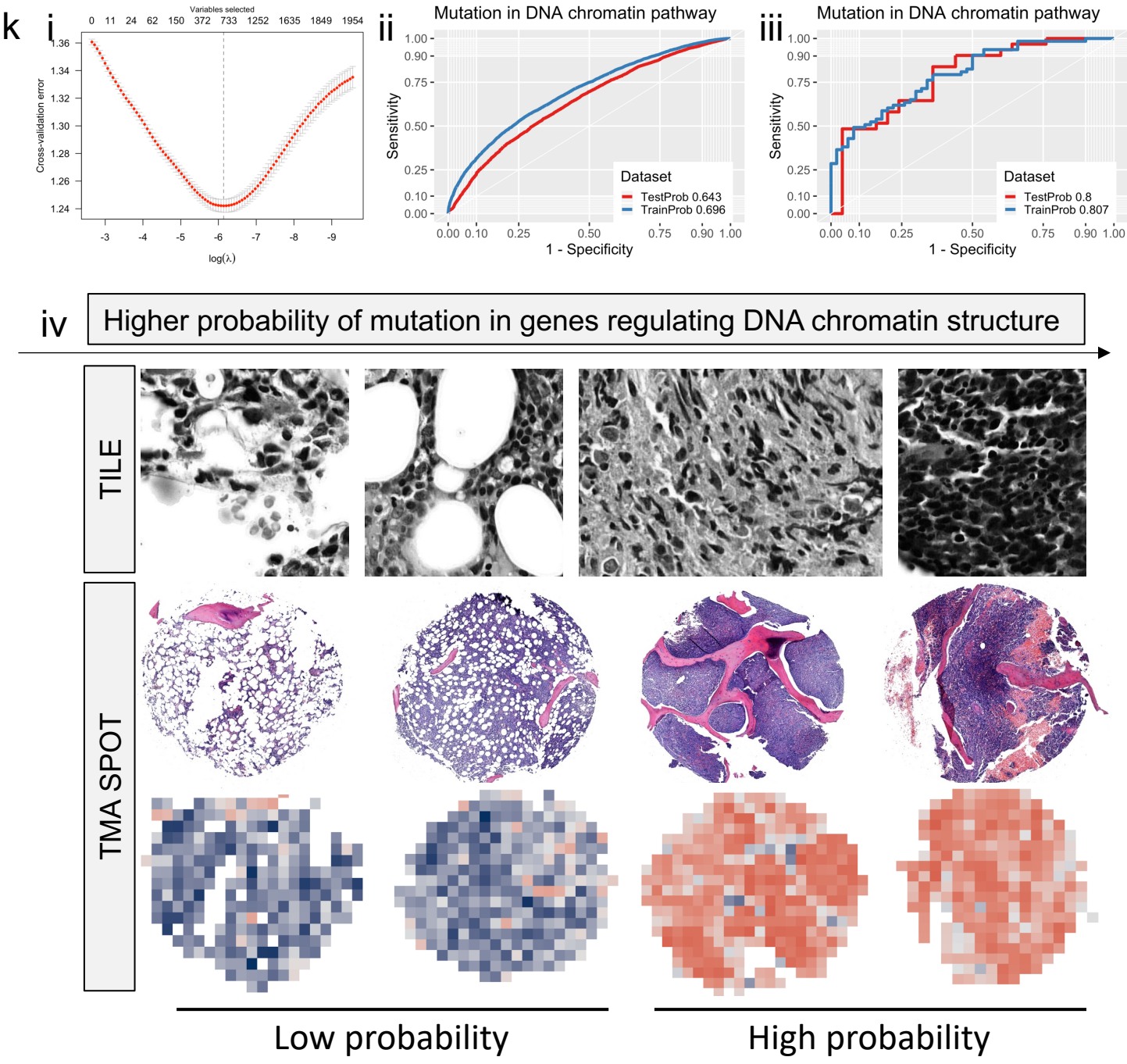
Supplementary Figure 7h.



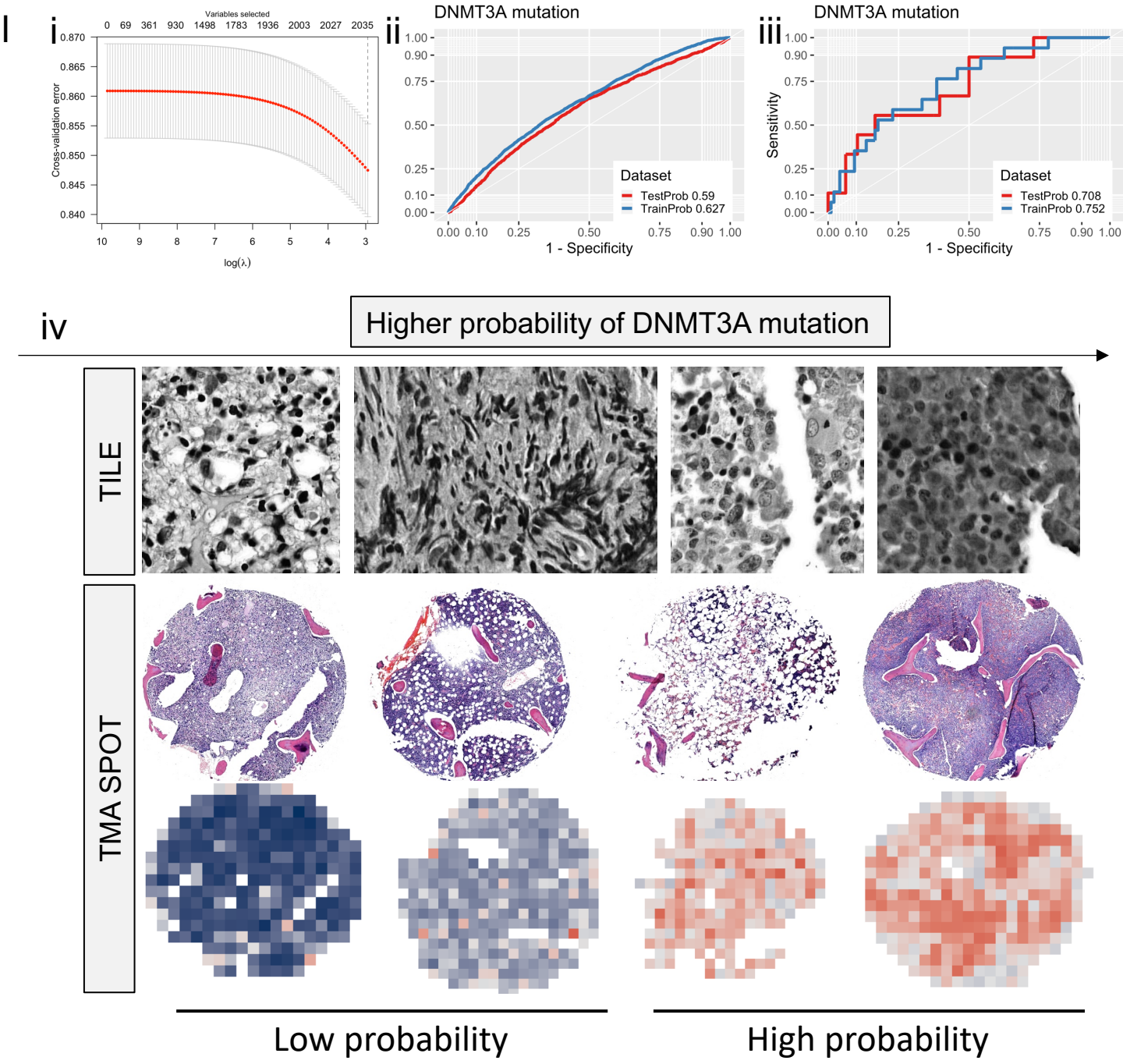
Supplementary Figure 7i.



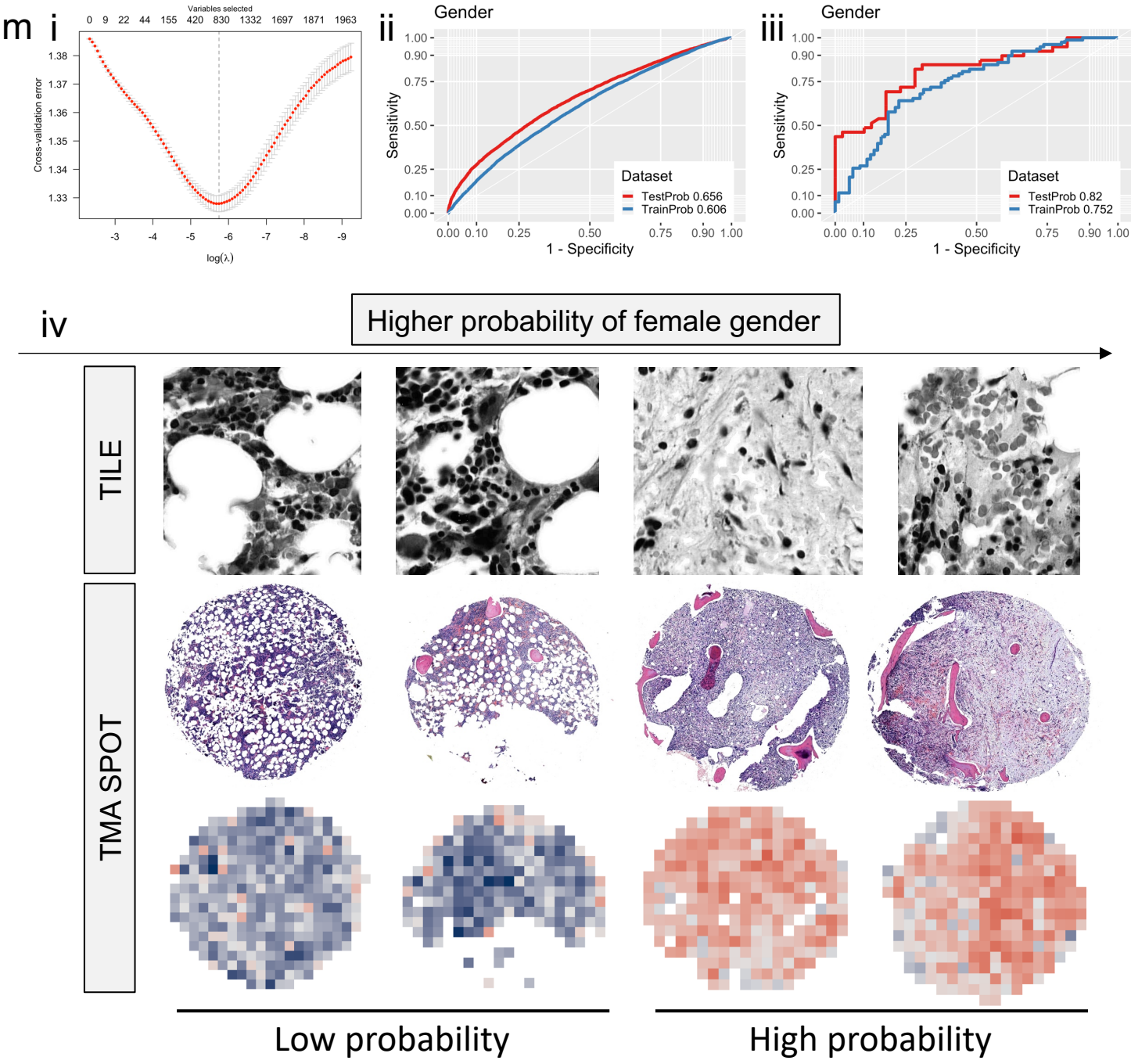
Supplementary Figure 7j.



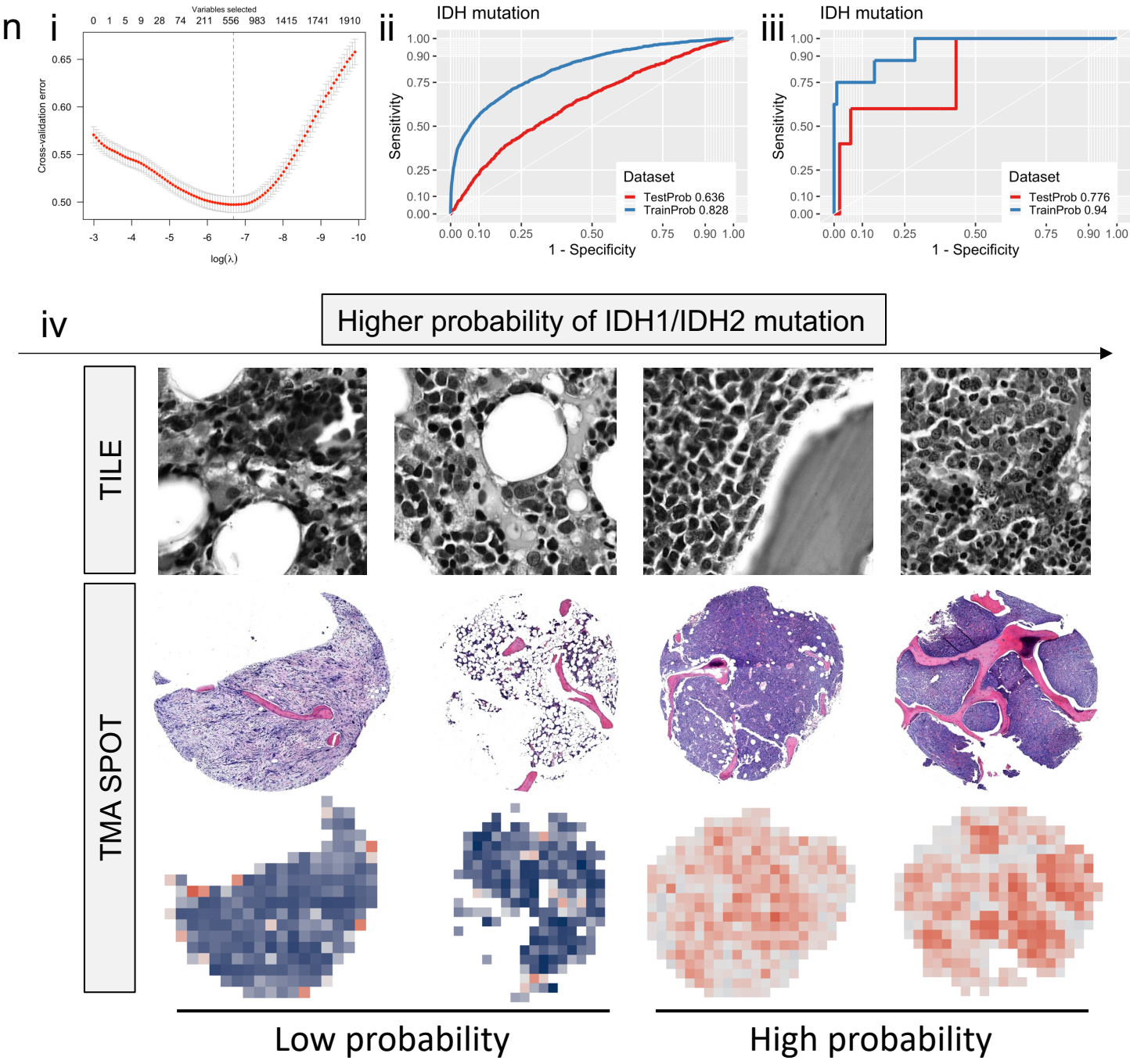
Supplementary Figure 7k.



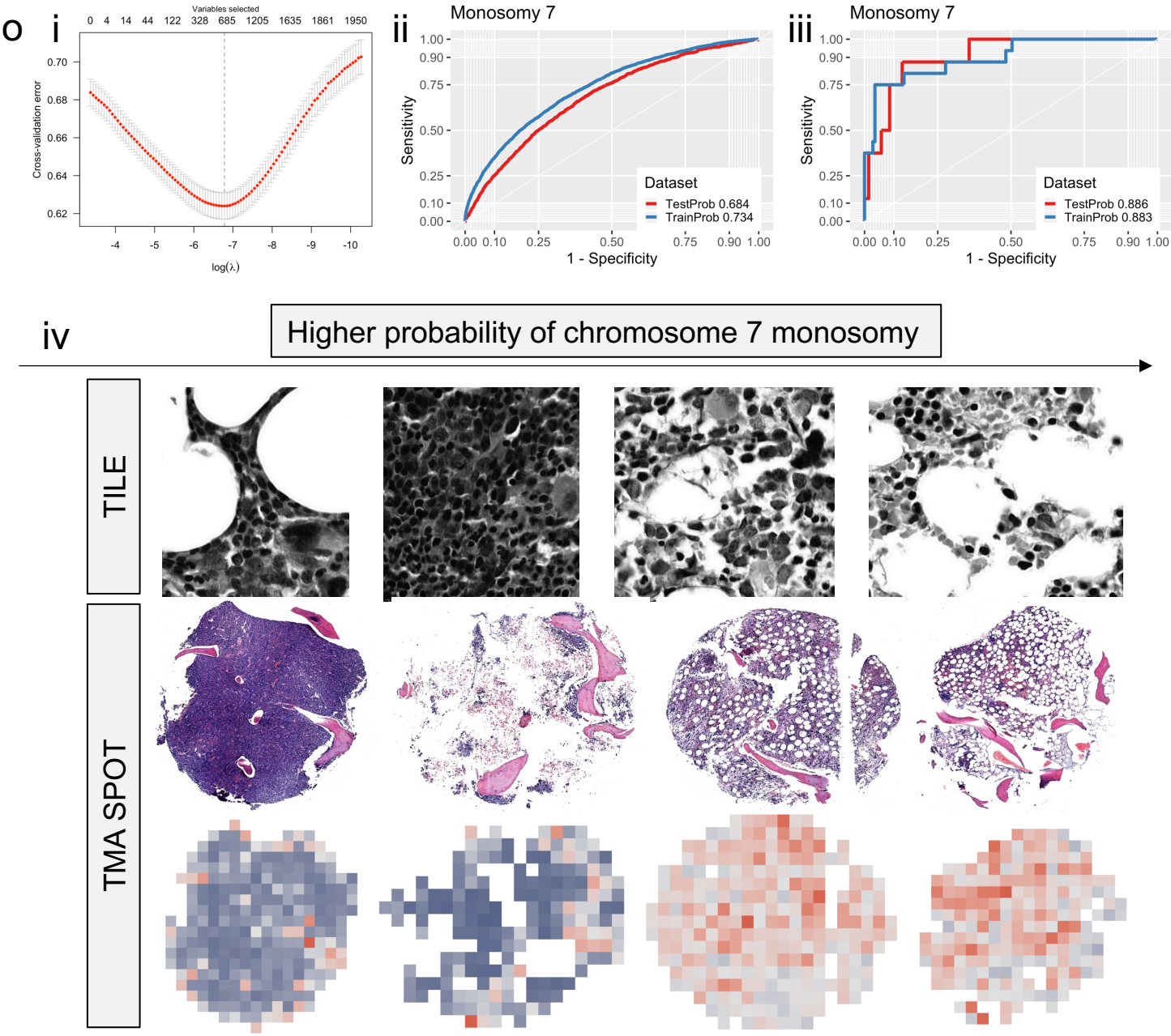
Supplementary Figure 7l.



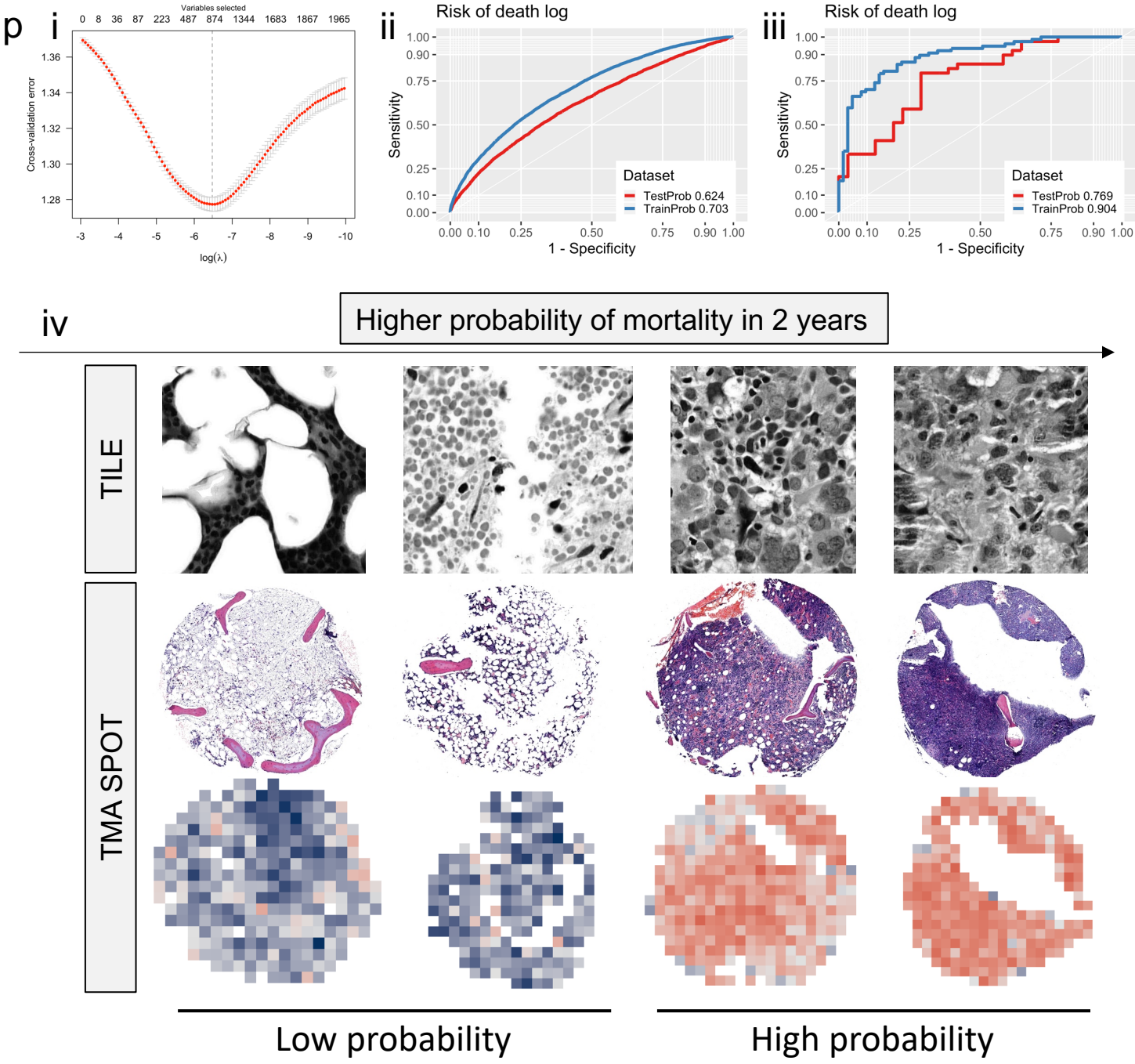
Supplementary Figure 7m.



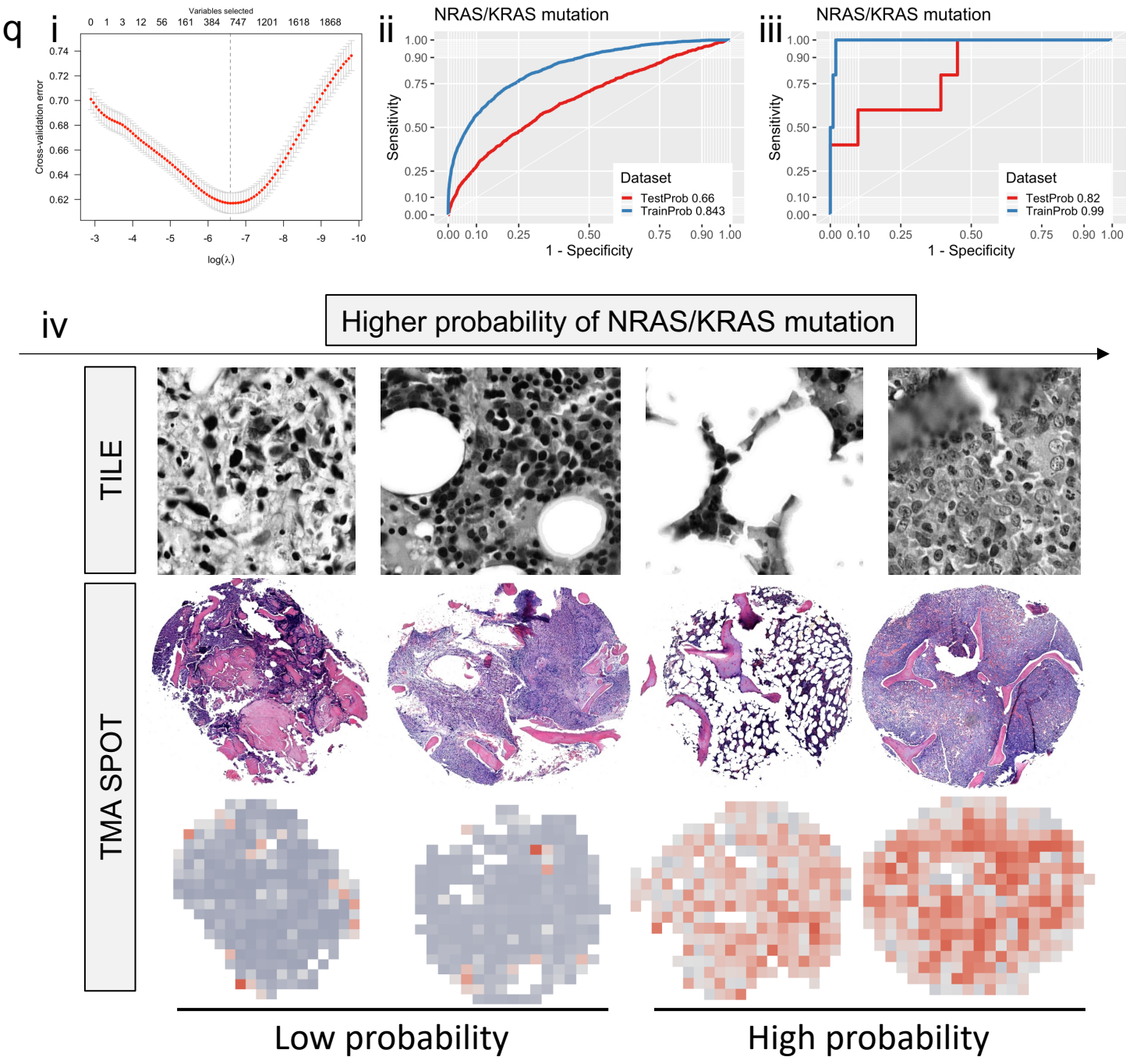
Supplementary Figure 7n.



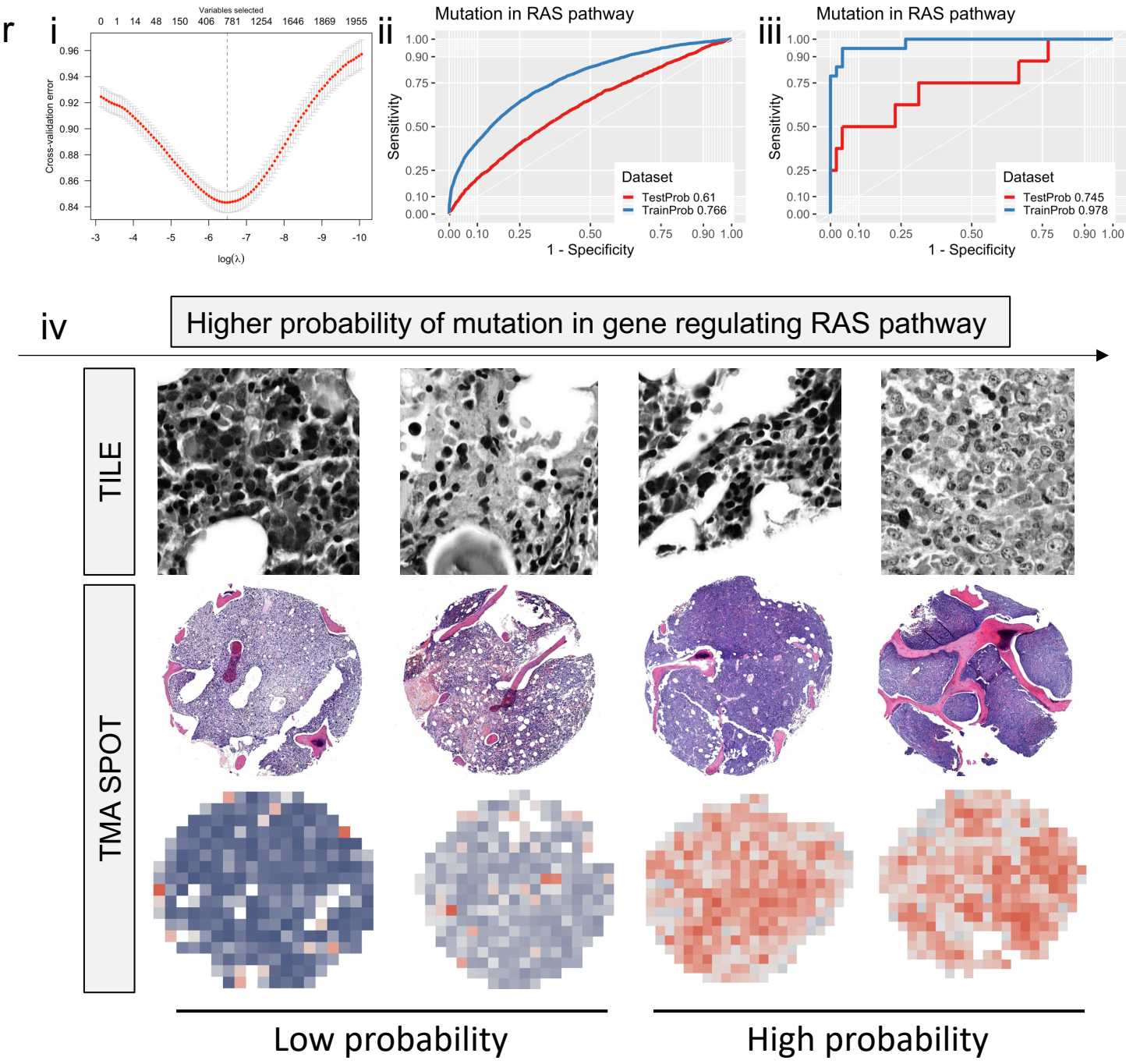
Supplementary Figure 7o.



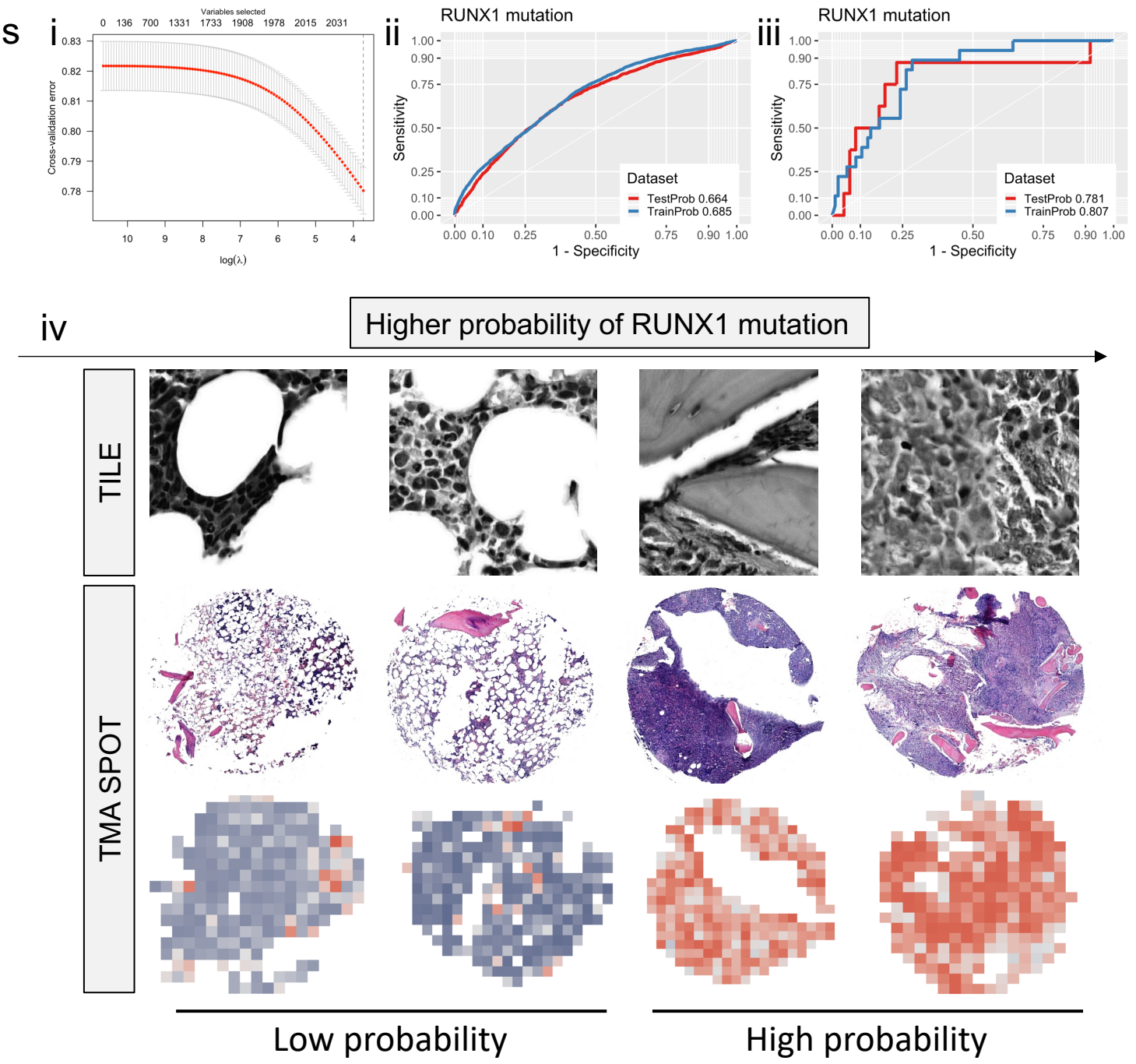
Supplementary Figure 7p.



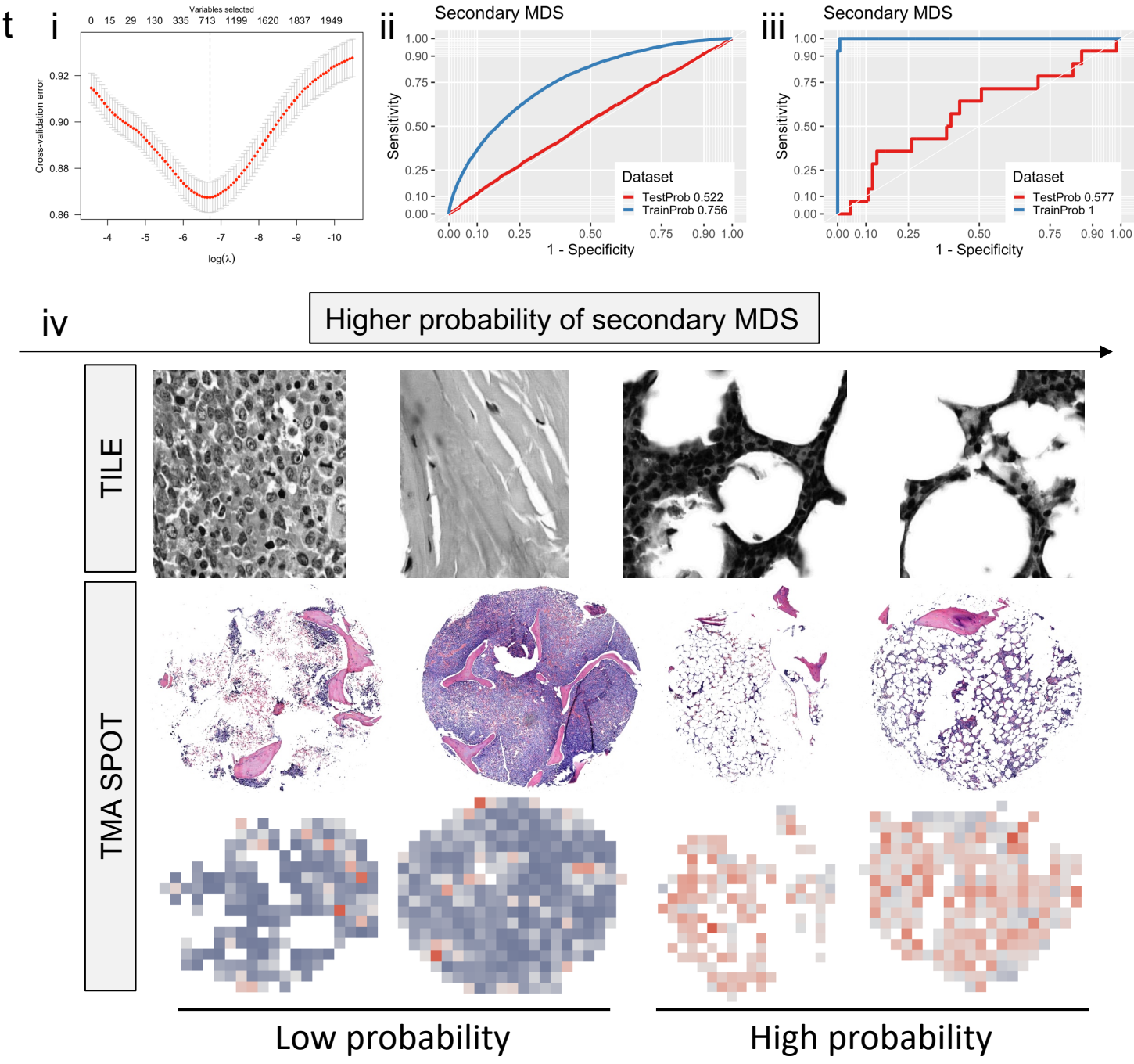
Supplementary Figure 7q.



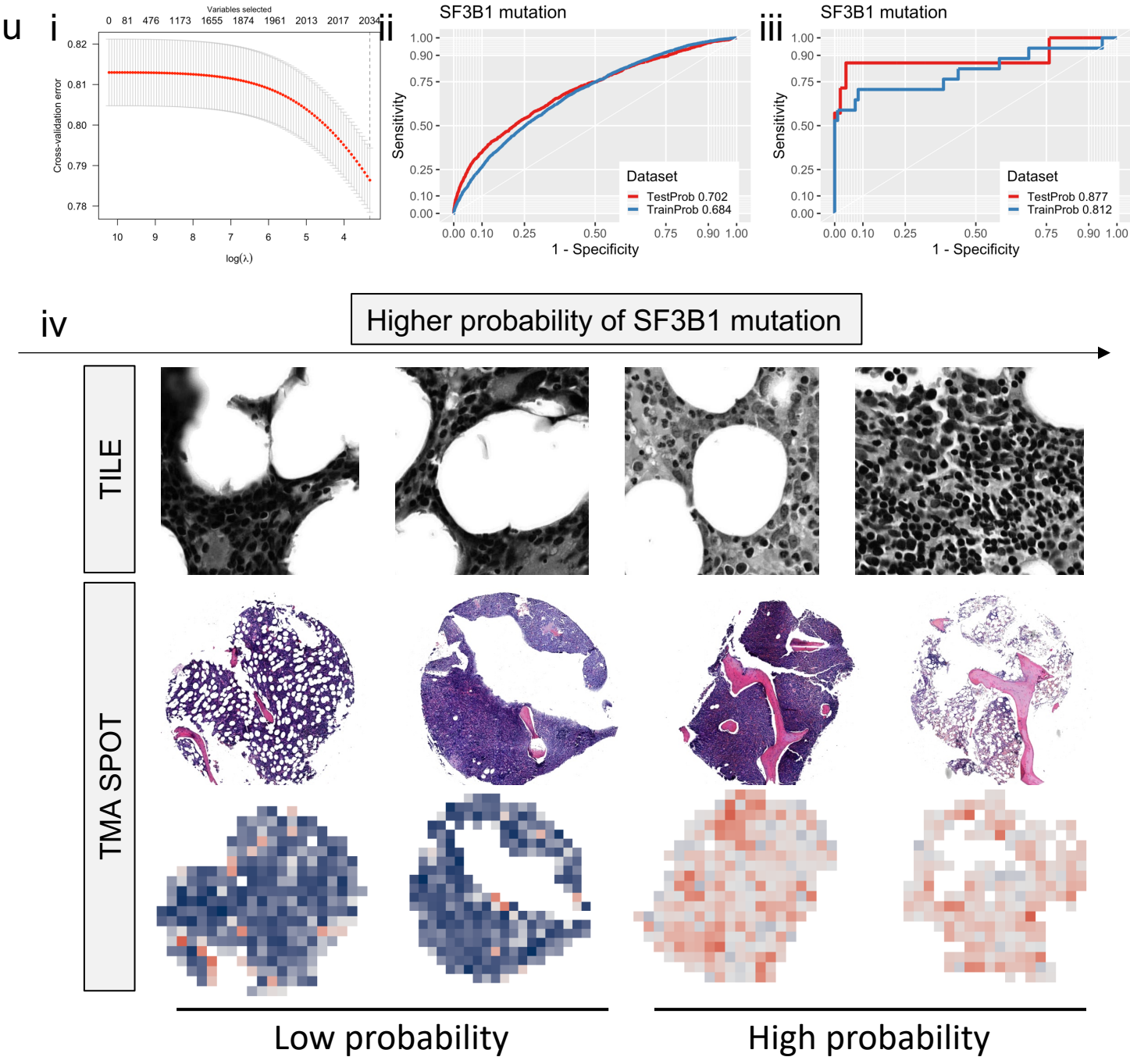
Supplementary Figure 7r.



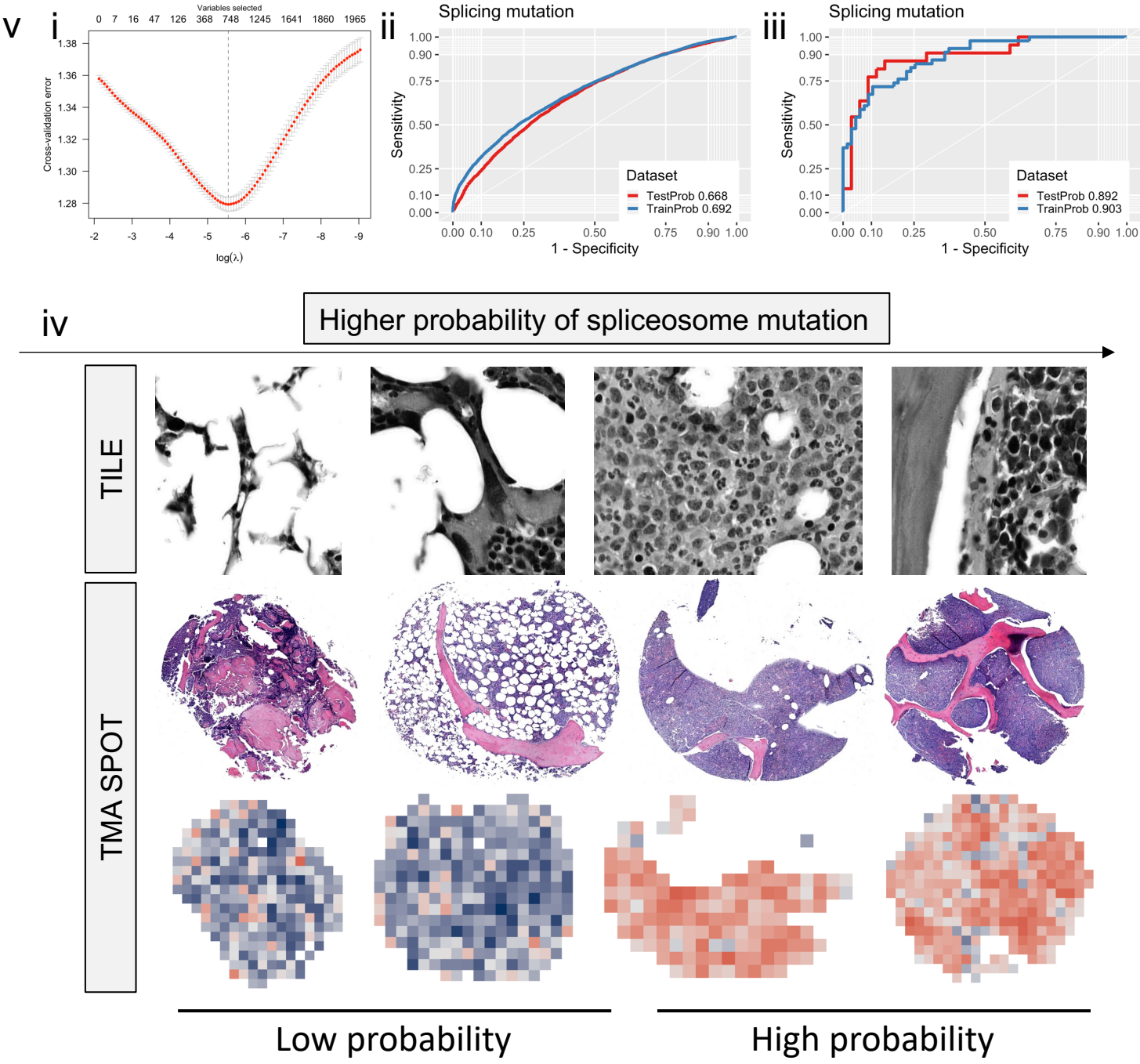
Supplementary Figure 7s.



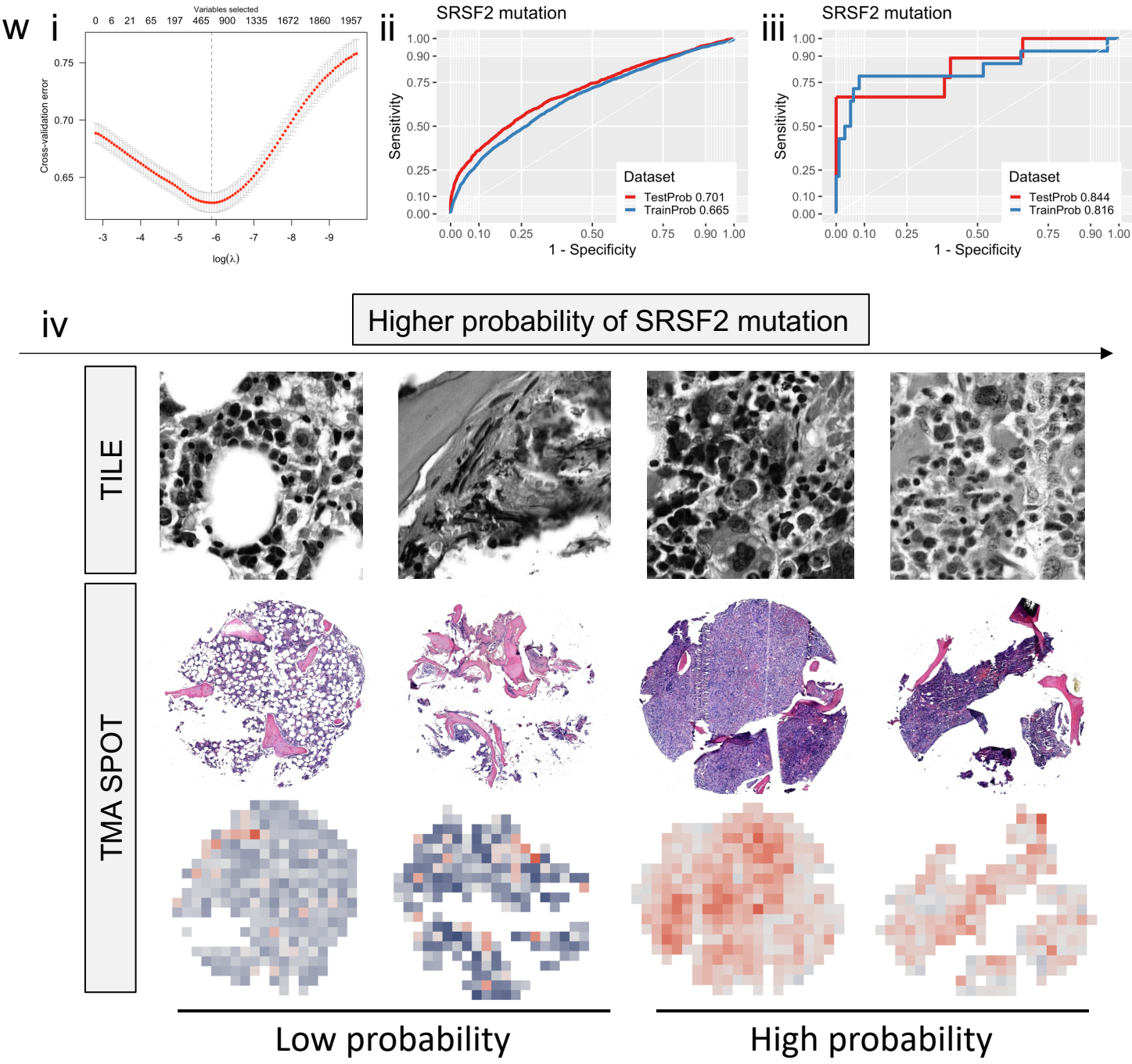
Supplementary Figure 7t.



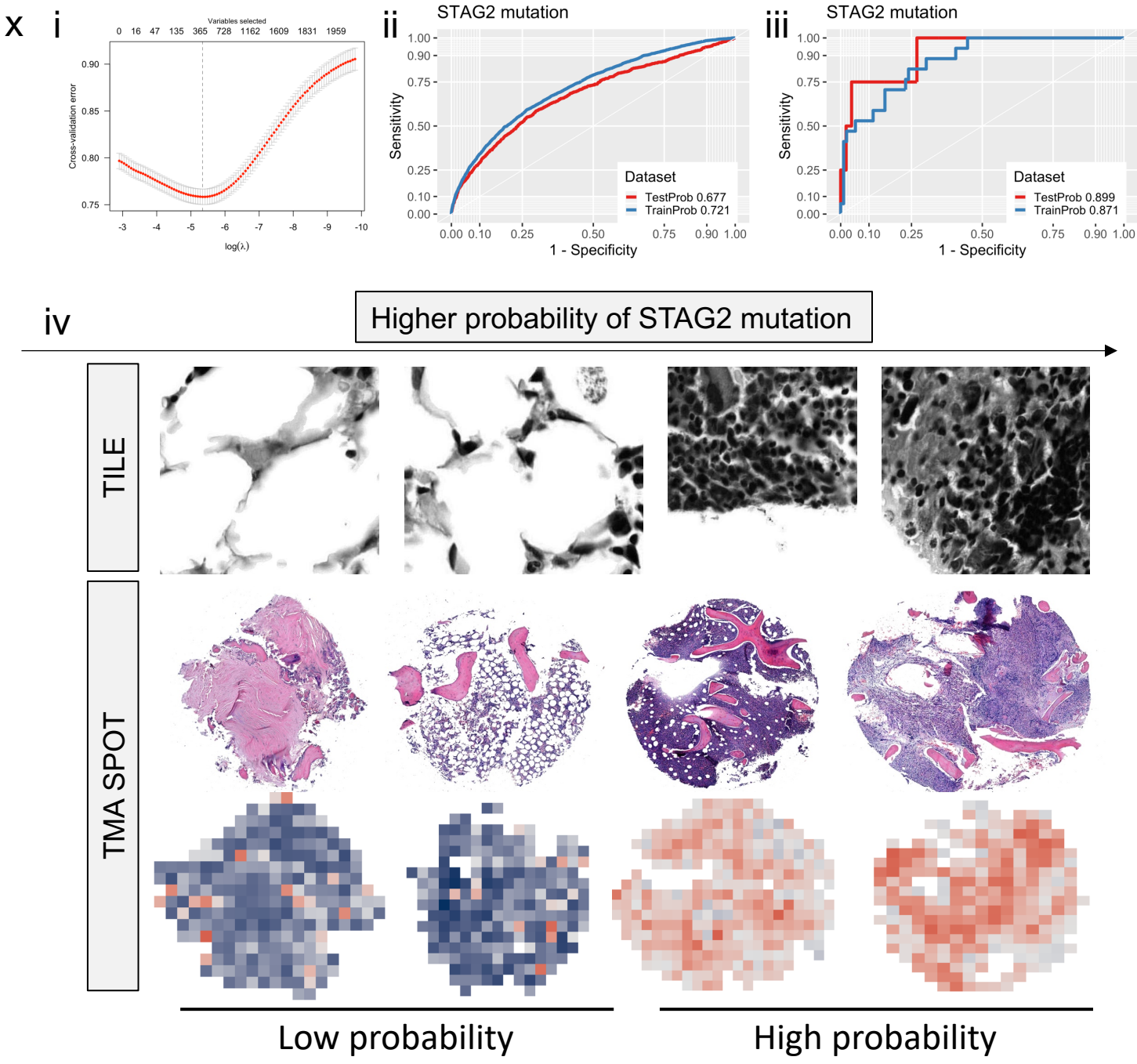
Supplementary Figure 7u.



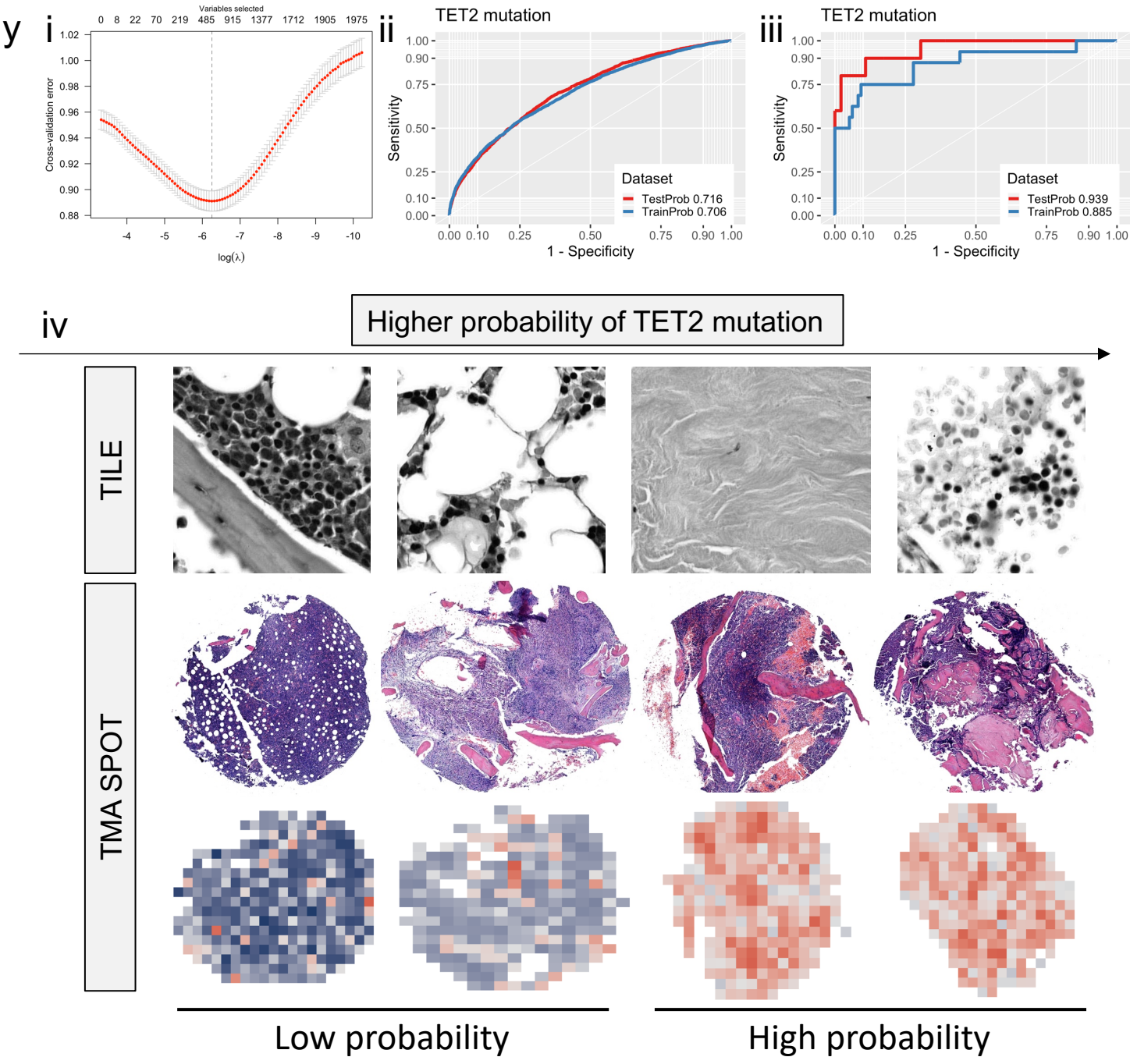
Supplementary Figure 7v.



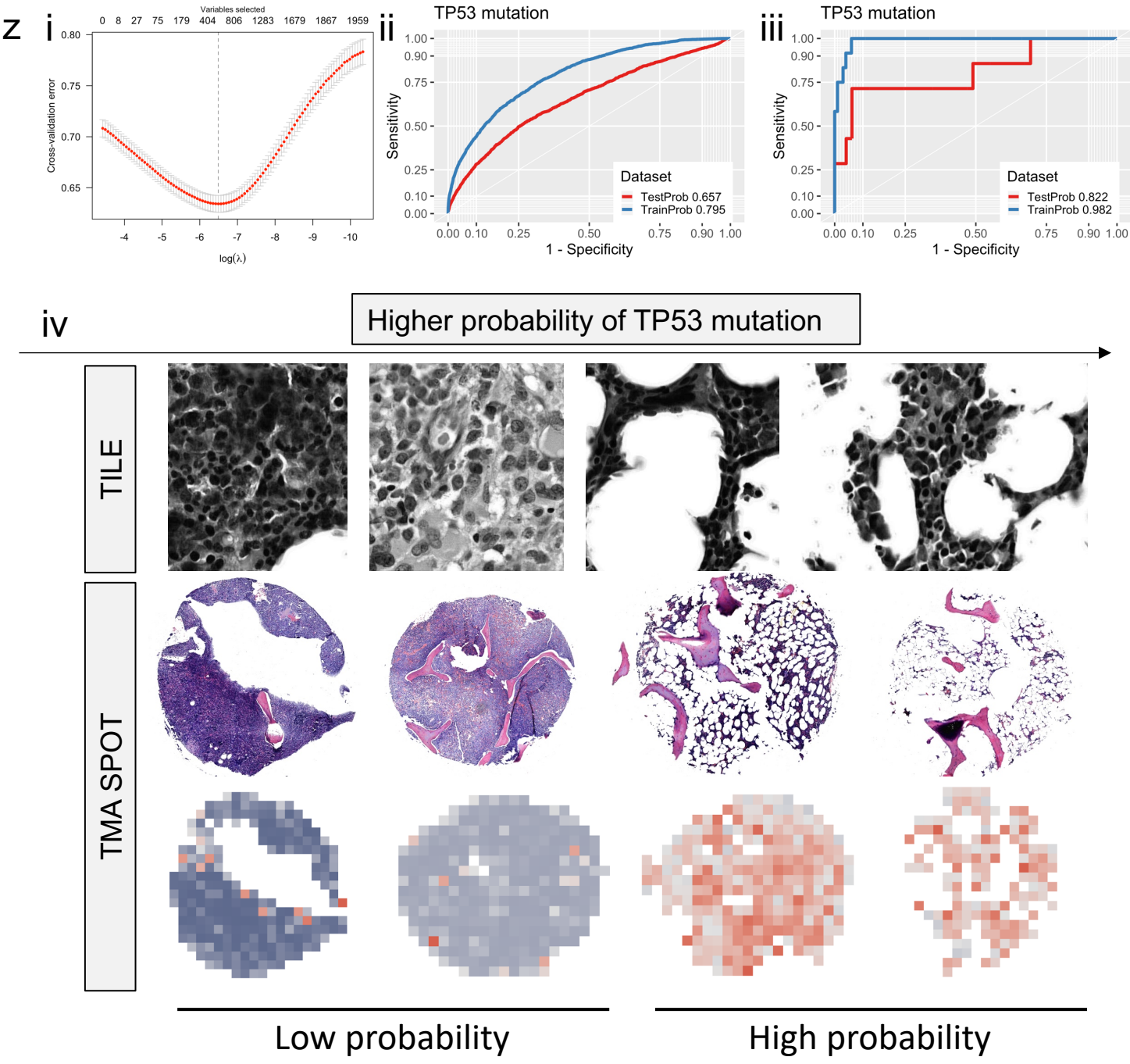
Supplementary Figure 7w.



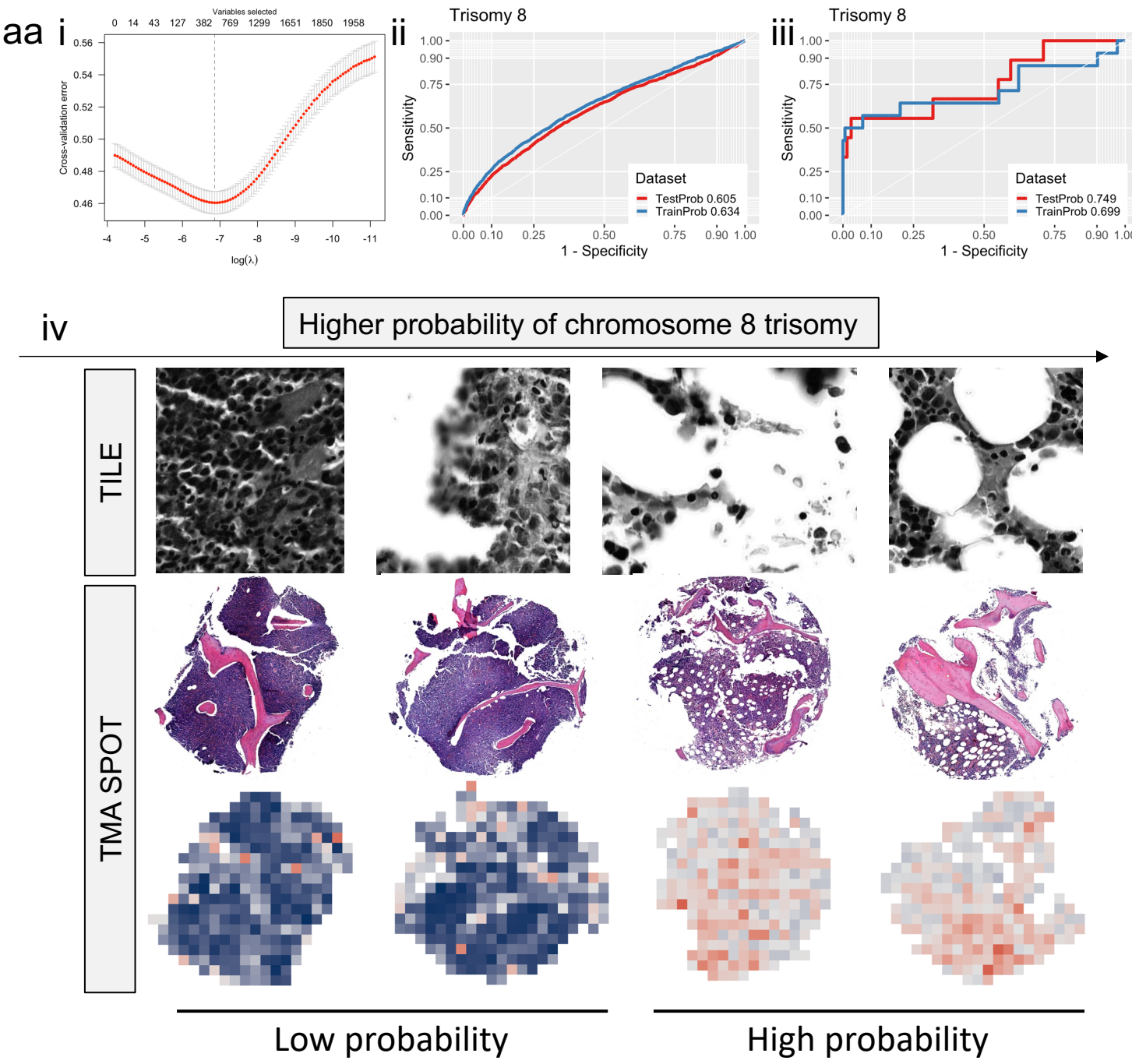
Supplementary Figure 7x.



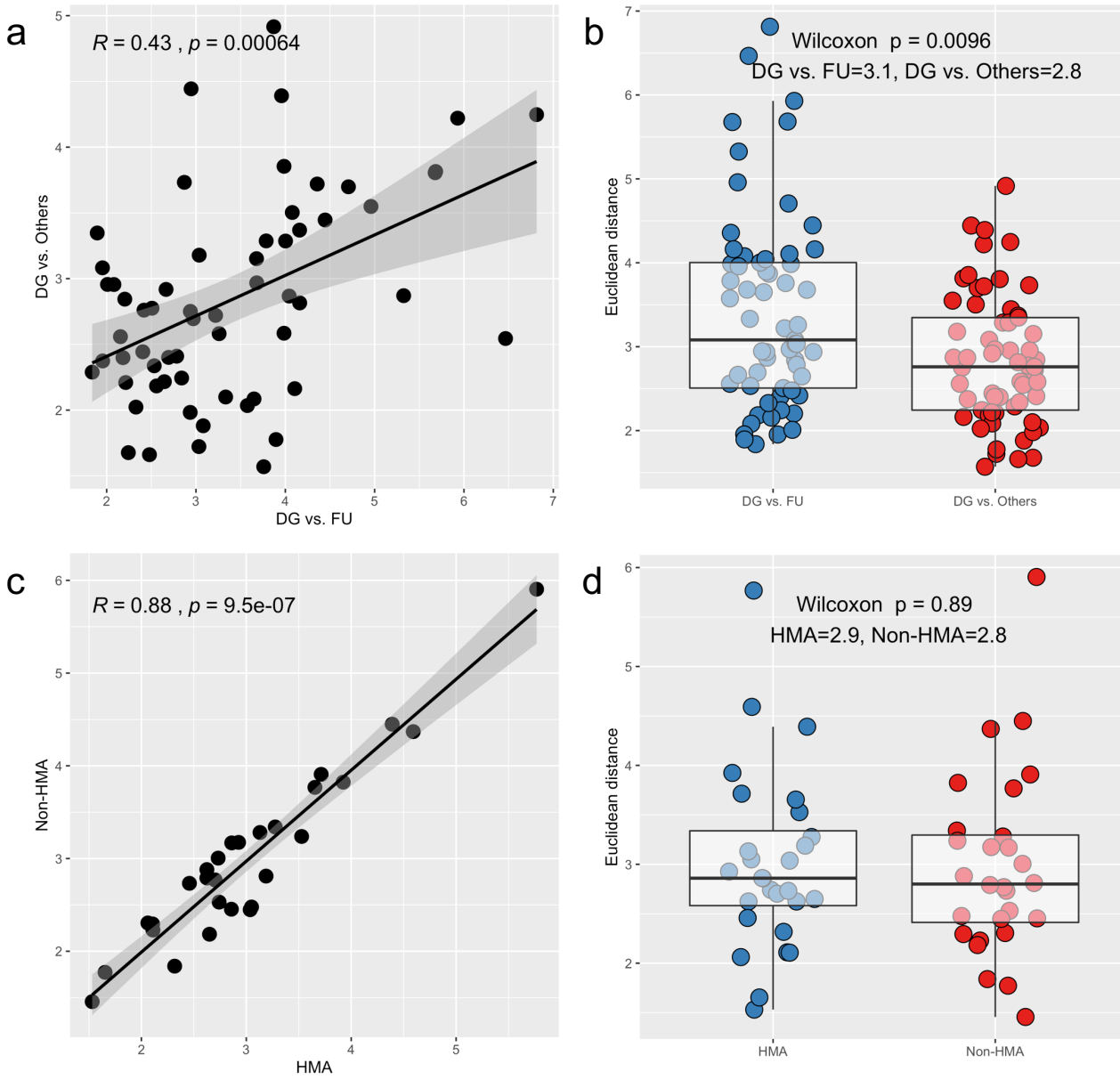
Supplementary Figure 7y.



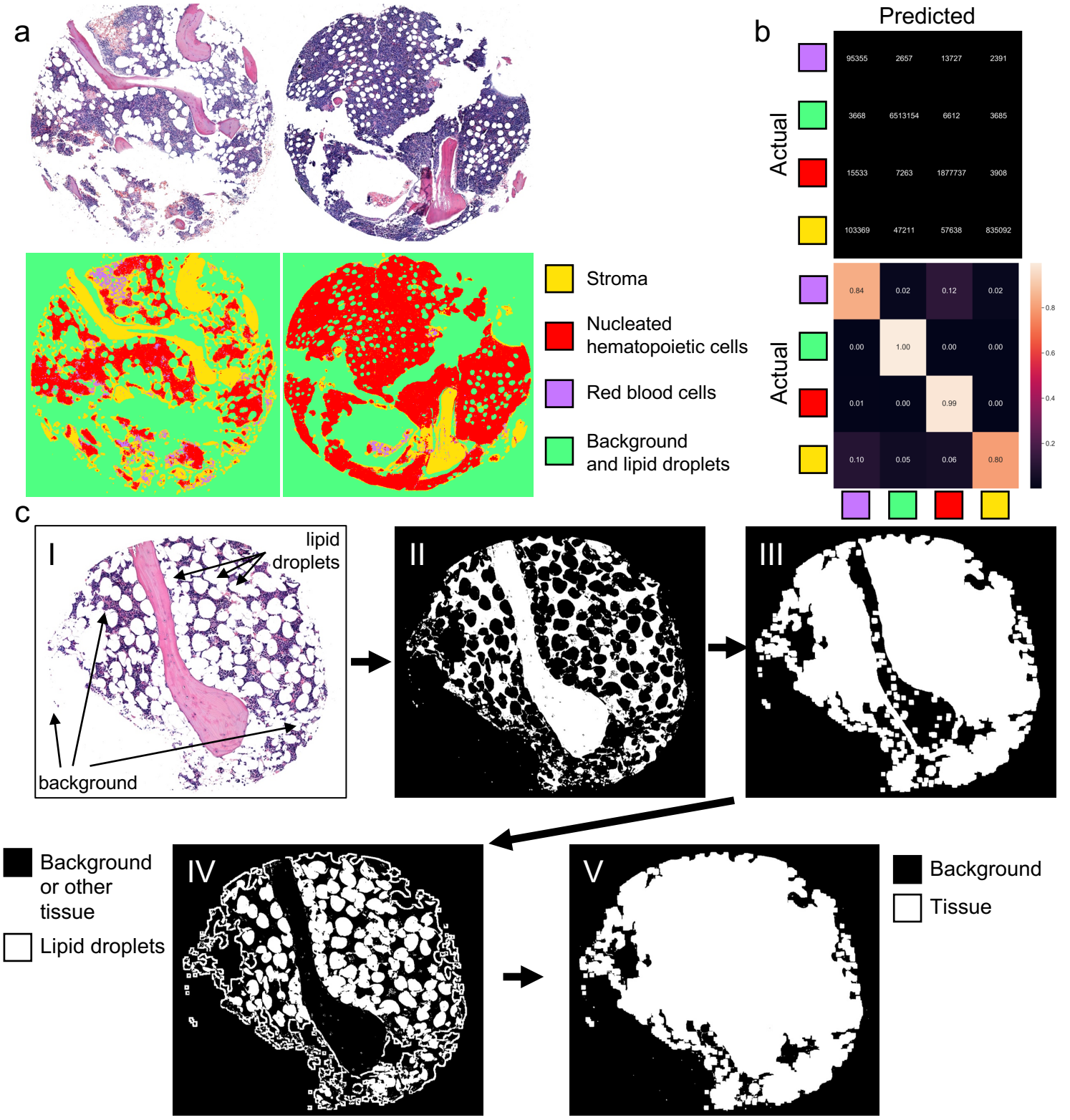
Supplementary Figure 7z.



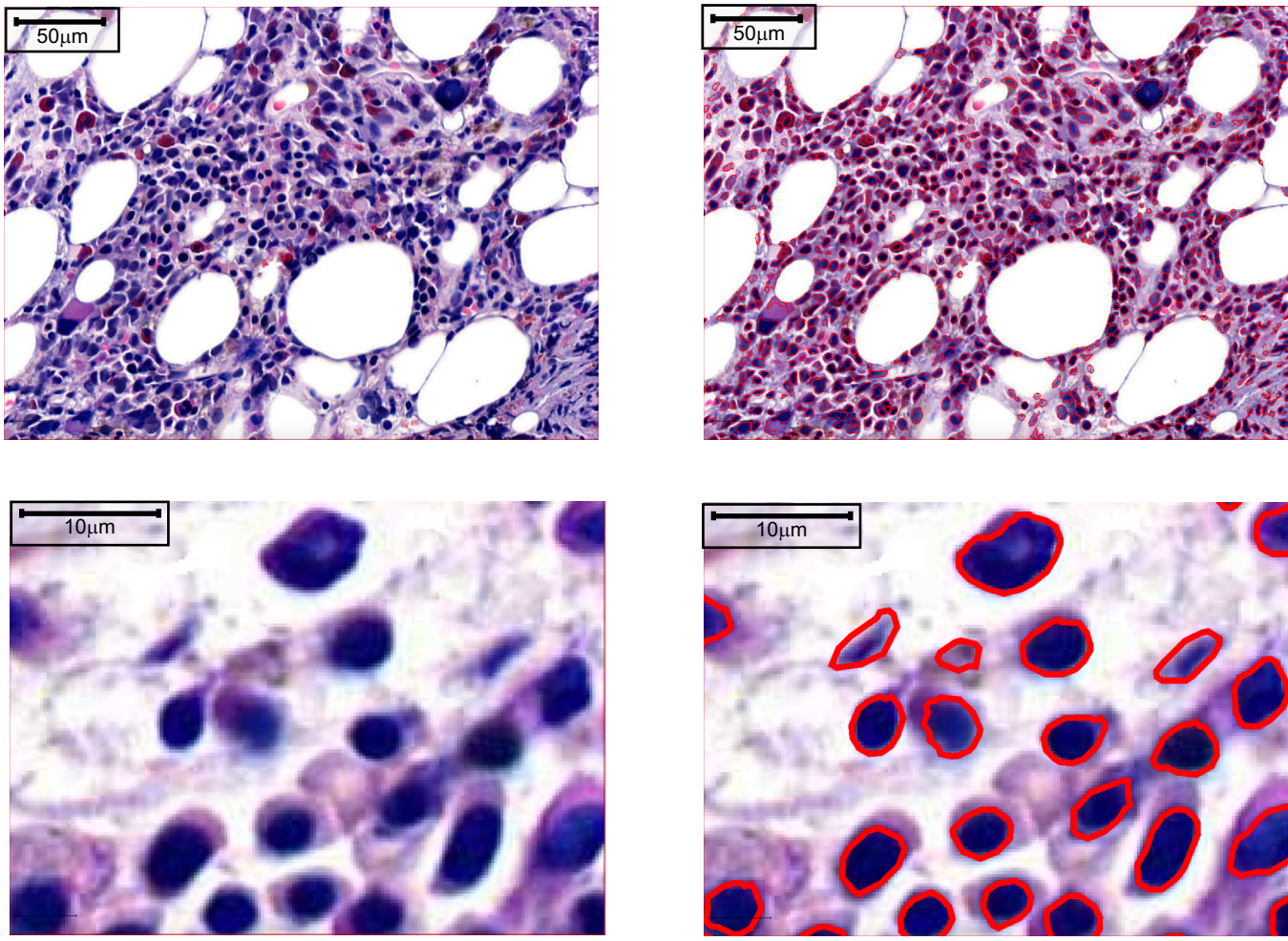
Supplementary Figure 7aa.



Supplementary Figure 8. (a) Scatter plot of the Euclidean distance between the diagnose sample of an index patient and the follow-up sample of the index patient (x-axis) compare with a sample from a non-index patient (y-axis). The Euclidean distance represents the similarity of the visual features of two samples. The R correlation coefficient value has been computed with Spearman correlation. (b) Scatter plot and Wilcoxon Rank-Sum test to compare Euclidean distance as in previous figure. Boxplots define the interquartile ranges and median values by diagnoses. (c) Scatter plot of the Euclidean distance between the samples taken during administration of a hypomethylating agent (HMA) to samples taken before or after HMA (Non-HMA). (d) Scatter plot and Wilcoxon Rank-Sum test to compare Euclidean distance as in previous figure.



Supplementary Figure 9. (a) Example results of pixel classification. Upper row shows H&E-stained bone marrow (BM) biopsy cores. Lower row shows results of pixel classification. (b) Confusion matrix of the classification accuracy of the pixel classification. Upper plot shows absolute pixel classification results of the 10 manually annotated test set images and the lower plot their proportions. (c) Tissue mask creation. I. H&E-stained tissue microarray core of BM biopsy. II. Binary image of image I. III. The binary image mask has been dilated and empty holes filled. IV. Filled holes in image III are visualized here with white color. V. The final image mask consists of combined white areas in images II and IV.



Supplementary Figure 10. Hematoxylin and eosin-stained bone marrow tissue (left) and results from cell segmentation where red-colored boundaries delimit individual nuclei (right) at two different magnification.