Predicting Cancer Outcomes from Histology and Genomics Using Convolutional Networks

Presented by:
Haluk Dogan
https://haluk.github.io/
hdogan@vivaldi.net

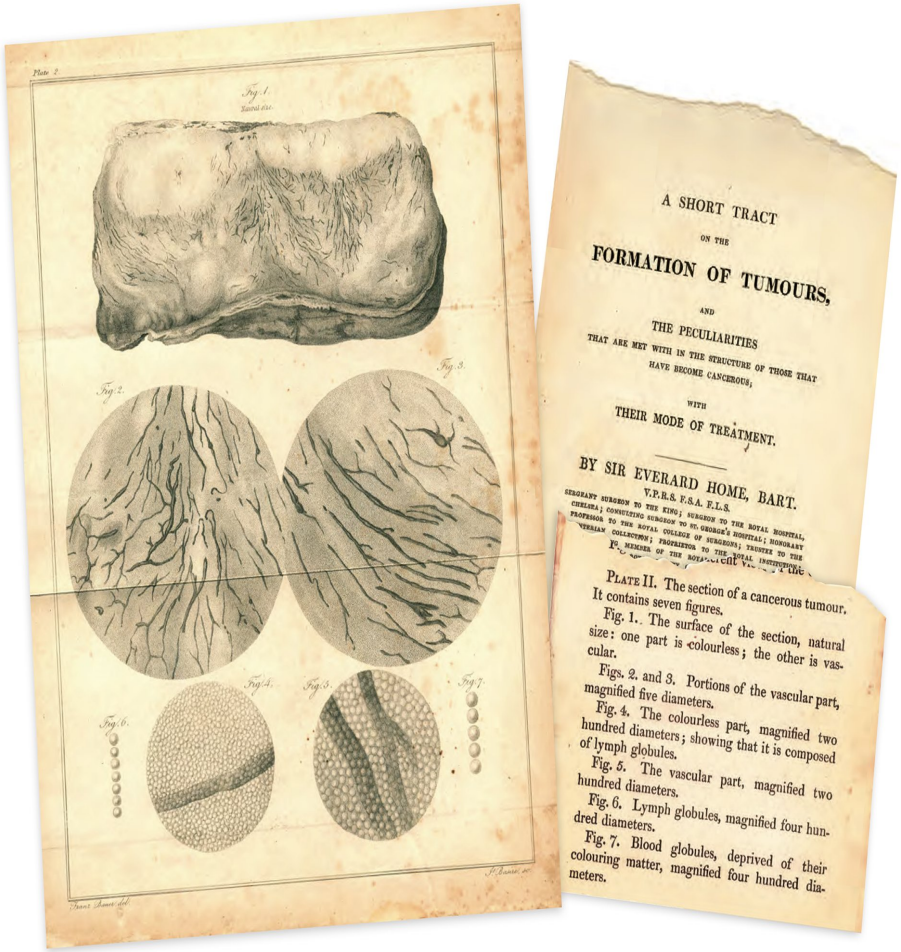
Some slides copied from

Prof. David Madigan, Columbia University

Department of Computer Science
University of Nebraska-Lincoln

March 6, 2020

Histology



A SHORT TRACT
ON THE
FORMATION OF TUMOURS,
AND
THE PECULIARITIES
THAT ARE MET WITH IN THE STRUCTURE OF THOSE THAT HAVE BECOME CANCEROUS;
WITH
THEIR MODE OF TREATMENT.

BY SIR EVERARD HOME, BART.

PLATE II. The section of a cancerous tumour. It contains seven figures.

Fig. 1. The surface of the section, natural size: one part is colourless; the other is vascular.

Figs. 2. and 3. Portions of the vascular part, magnified five diameters.

Fig. 4. The colourless part, magnified two hundred diameters; showing that it is composed of lymph globules.

Fig. 5. The vascular part, magnified two hundred diameters.

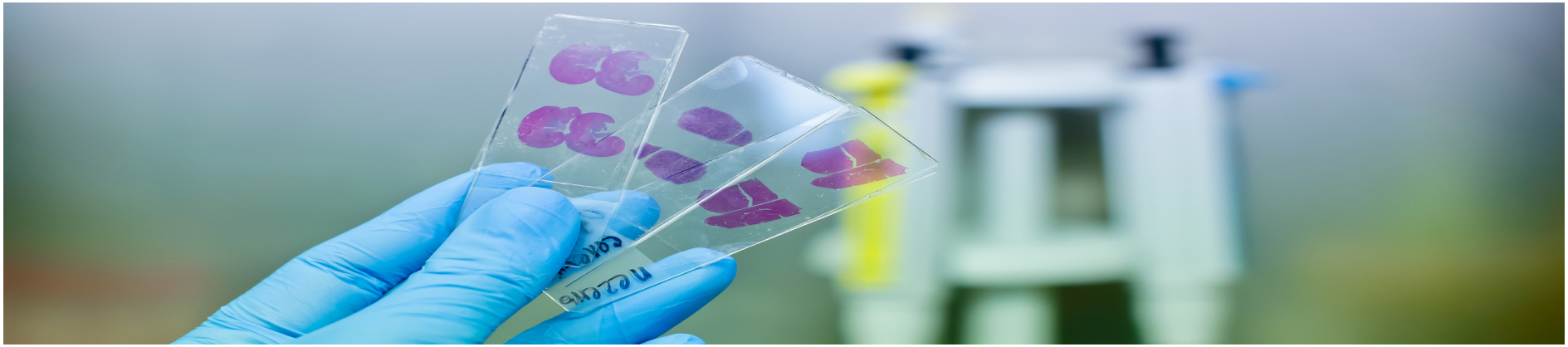
Fig. 6. Lymph globules, magnified four hundred diameters.

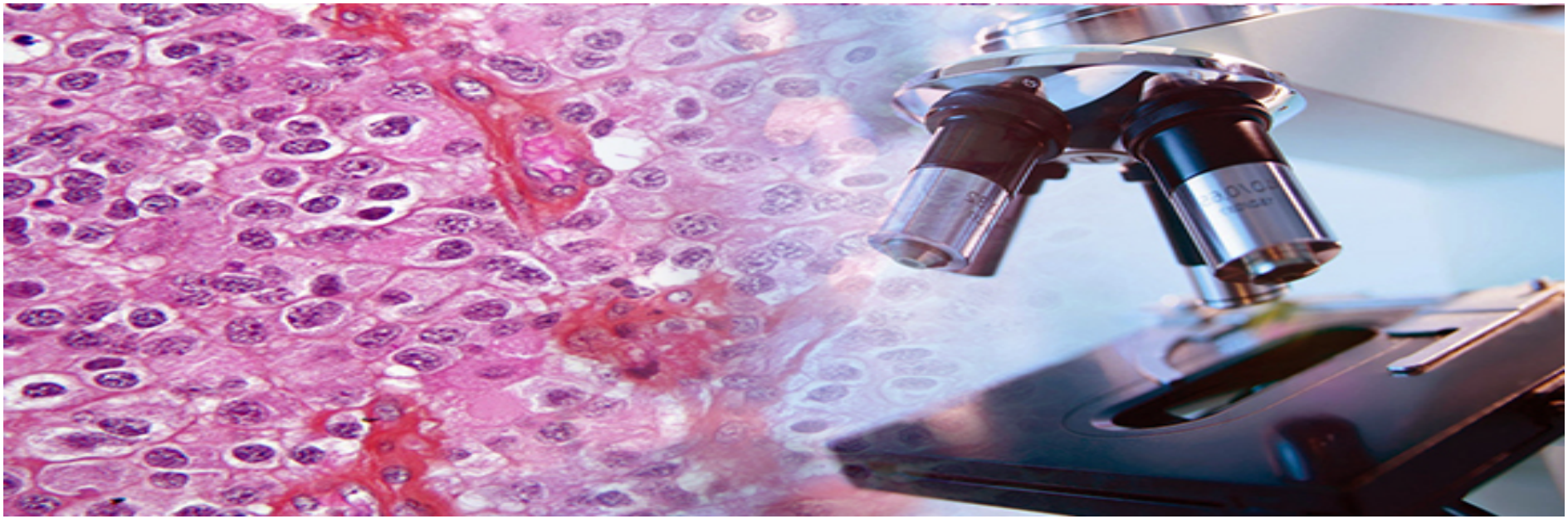
Fig. 7. Blood globules, deprived of their colouring matter, magnified four hundred diameters.

Everard Home in 1830

Science Meets Art

Histology (cont'd)





Histology (cont'd)

Anatomic pathologists evaluate histology to classify and grade lesions

Cancer diagnosis:

- characteristics
  - nuclear atypia
  - mitotic activity
  - cellular density
  - tissue architecture
- incorporating
  - cytologic details
  - higher-order patterns

Cancer prognostication:

- genomic biomarkers
  - genetic alterations
  - gene expression
  - epigenetic modifications

Time-to-Event Data Analysis

- Logistic regression finds associations between risk factors and **presence/absence** of a disease
- We are interested in how a risk factor or treatment affects **time** to disease or some other event
  - Time until tumor recurrence
  - Time until cardiovascular death after some treatment intervention
- Response is called **survival time**
  - Always ≥ 0
  - Usually continuous
  - Incompletely observed responses are censored
    - Survival time was at least equal to some time t
    - Right censoring
    - Left censoring

Time-to-Event Data Analysis (cont'd)

- Right censoring
  - When a subject leaves the study before an event occurs
  - Study ends before the event has occurred

Example

We consider patients in a clinical trial to study the effect of treatments on stroke occurrence. The study ends after 5 years. Those patients who have had no strokes by the end of the year are censored. If the patient leaves the study at time t_e, then the event occurs in (t_e, ∞)

- Left censoring: event of interest has already occurred before enrollment

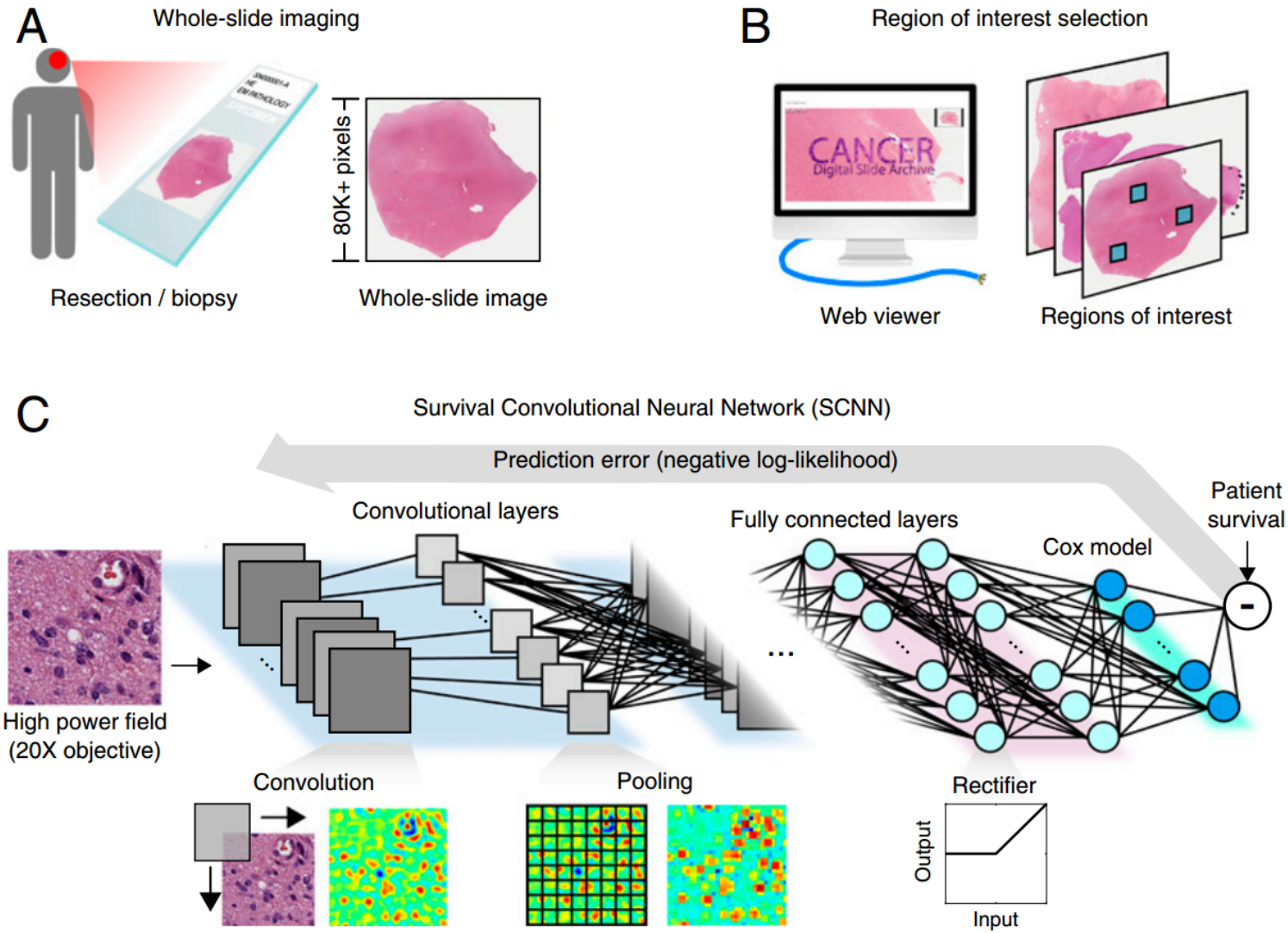
Time-to-Event Data Analysis (cont'd)

- Classification methods **can not**
  - use incomplete follow-up subjects
  - model the probability of survival at other times
- Cox regression model **can**
  - utilize all subjects in training
  - model their survival probabilities for a range of times with a single model

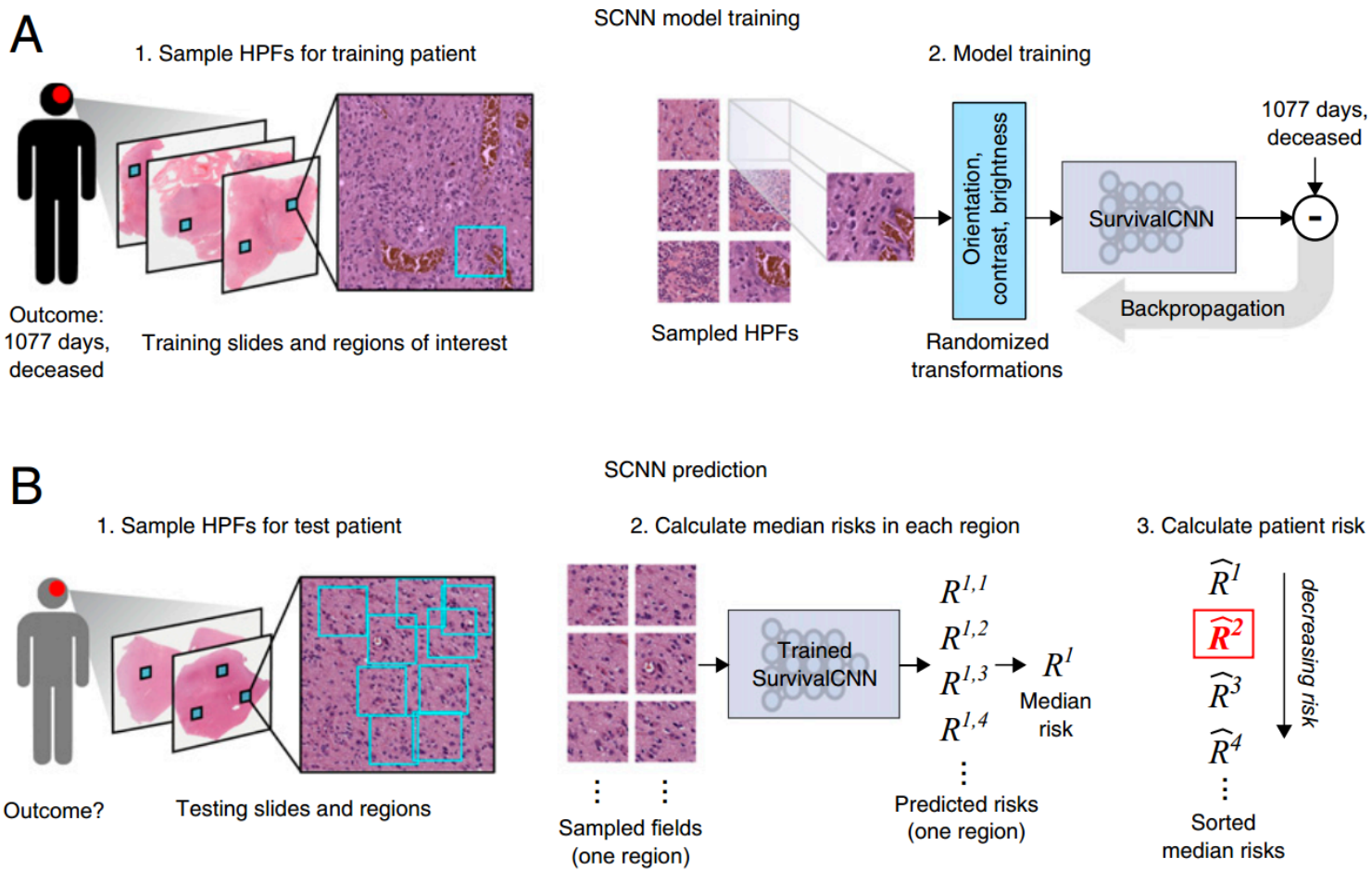
Proposed Method

1. Survival convolutional neural networks (SCNNs), which provide highly accurate prediction of time-to-event outcomes from histology images
2. Extend SCNN to integrate both histology images and genomic biomarkers into a unified prediction framework

Survival Convolutional Neural Networks



Survival Convolutional Neural Networks (cont'd)



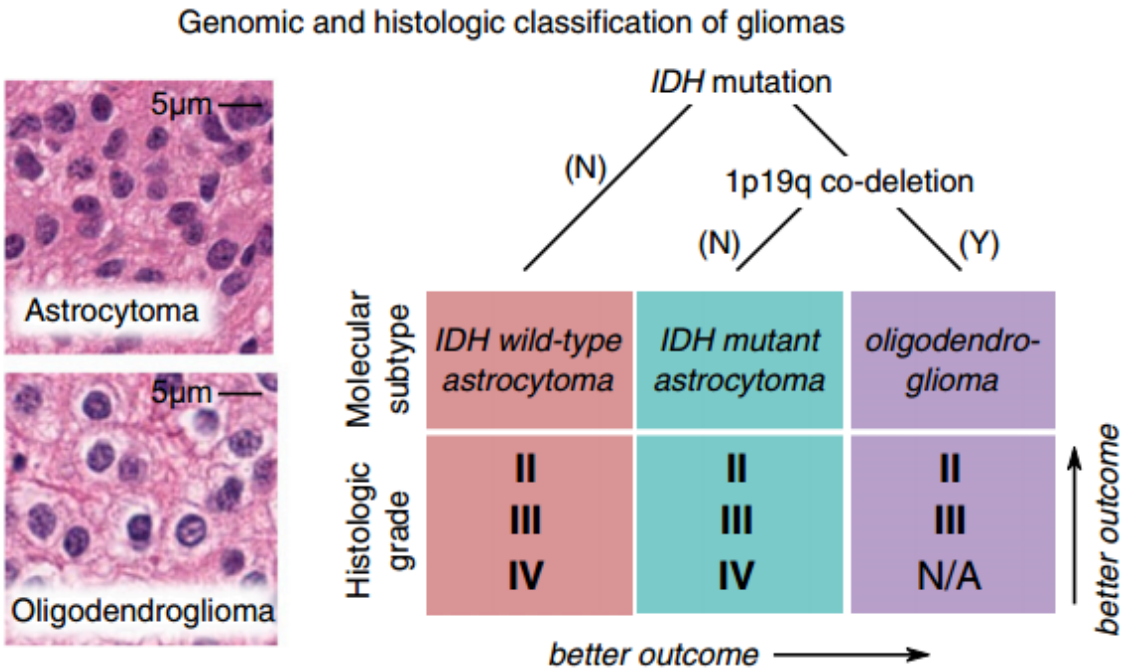
Harrell's C-index

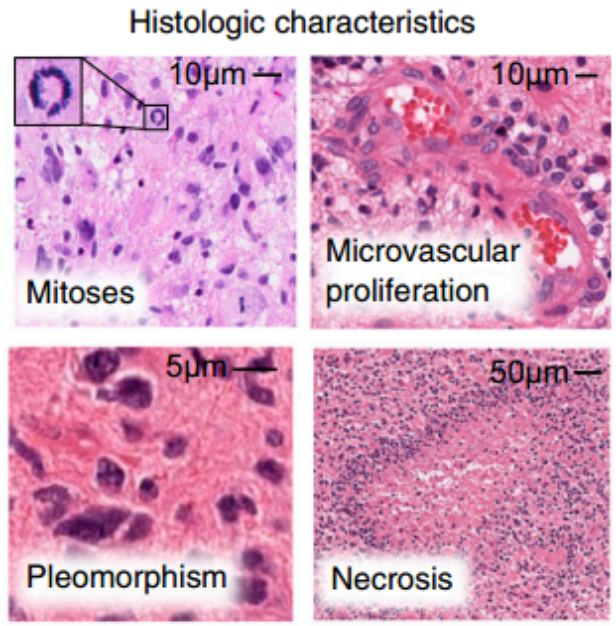
- Evaluates risk models in survival analysis, where data may be censored. Values:
  - near 0.5 are no better than a coin flip
  - near 1 indicate that the risk scores are good at determining which of two patients will have the disease first
  - near 0 indicate that the risk scores are worse than a coin flip
- Training a model:
  - n patients
  - $X_1, \ldots, X_p$ covariate variables
  - response variable T ("time-to-event")

$$c = \frac{\#\text{ concordant pairs}}{\#\text{ concordant pairs } + \#\text{ discordant pair}}$$

$$= \frac{\sum_{i \neq j} 1\{\eta_i < \eta_j\} 1\{T_i > T_j\} d_j}{\sum_{i \neq j} 1\{T_i > T_j\} d_j}$$

Experimental Design

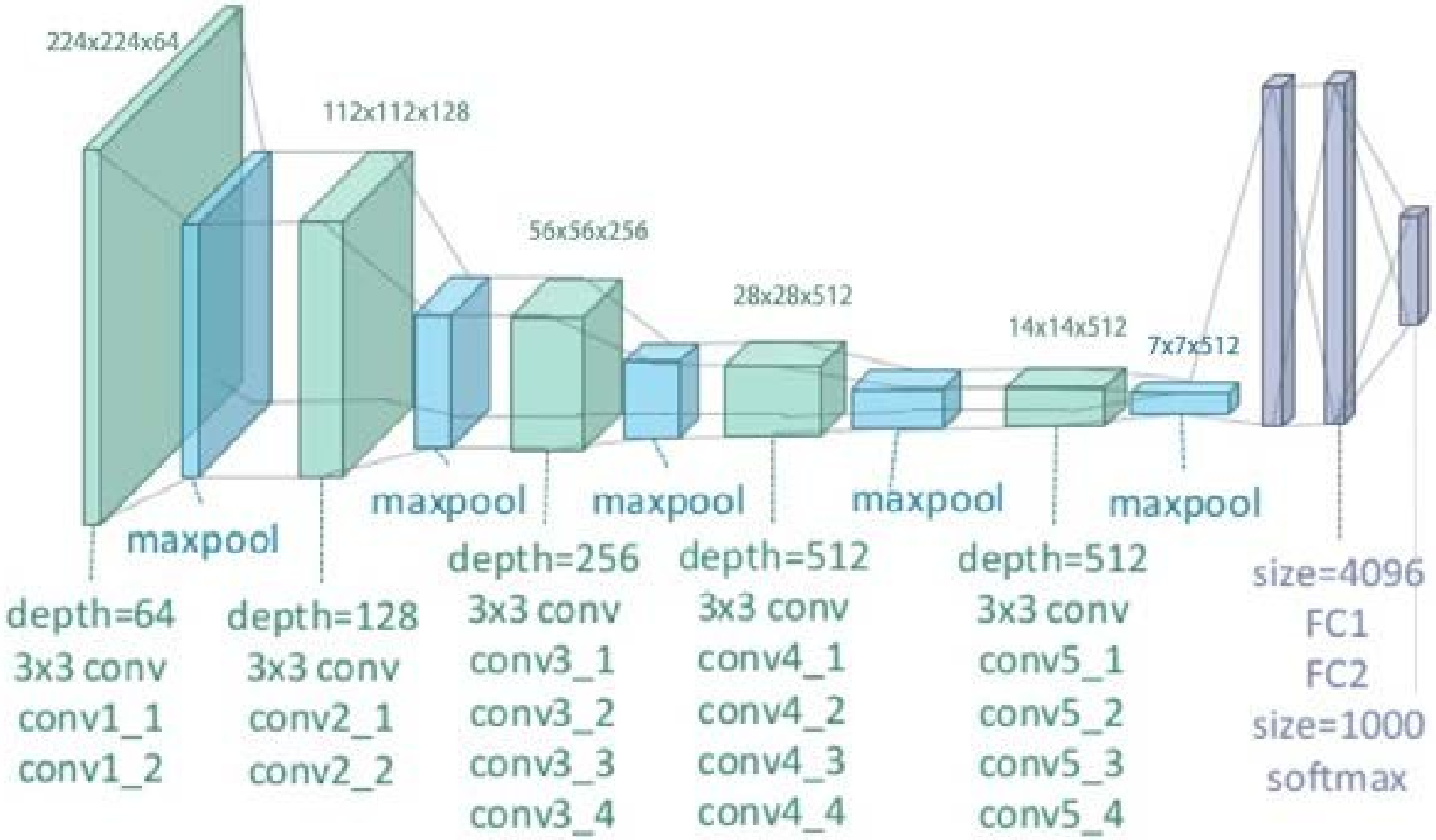




Data and Image Curation

- Brain cancer from TCGA (LGG and GBM cohorts). Overall survivals ranging from less than 1 to 14y or more
- Removed images containing
  - bubbles
  - section folds
  - pen markings
  - poor (over or under) staining
  - geographic necrosis
- After filtering: 1,061 WSI from 769 unique patients
- ROI images (1,024 $\times$ 1,024) were cropped at 20$\times$
- Color-normalized to a gold standard H&E calibration
- 256 $\times$ 256 patches were sampled from ROIs
- Randomized transformations applied to patches
  - tissue orientation (mirror) and color variations (contrast and brightness)

VGG19 Architecture



Risk Prediction

$R = \beta^T X$

- $\beta \in \mathbb{R}^{256\times 1}$ are weights
- $X \in \mathbb{R}^{256\times 1}$ are inputs
- Backpropagation: $L(\beta, X) = -\sum_{i\in U}\left(\beta^T X_i - \log\sum_{j\in\Omega_i} e^{\beta^T X_j}\right)$
  - U is the set of right-censored samples
  - Ω_i is the set of "at-risk" samples with event or follow-up times
  - $\Omega_i = \{j|Y_j \geq Y_i\}$
  - Y_i is the last follow up time of patient i
- Optimizer: Adagrad
  - initial accumulator = 0.1
  - initial learning rate = 0.001 (learning rate decay factor = 0.1)
  - variance scaling method is used to initialize model weights
  - decay rate = 4e-4
  - 100 epochs
  - Stochastic gradient training with minibatch size 14
  - Randomized minibatches
  - Regularization: Dropout in the last fully connected layer

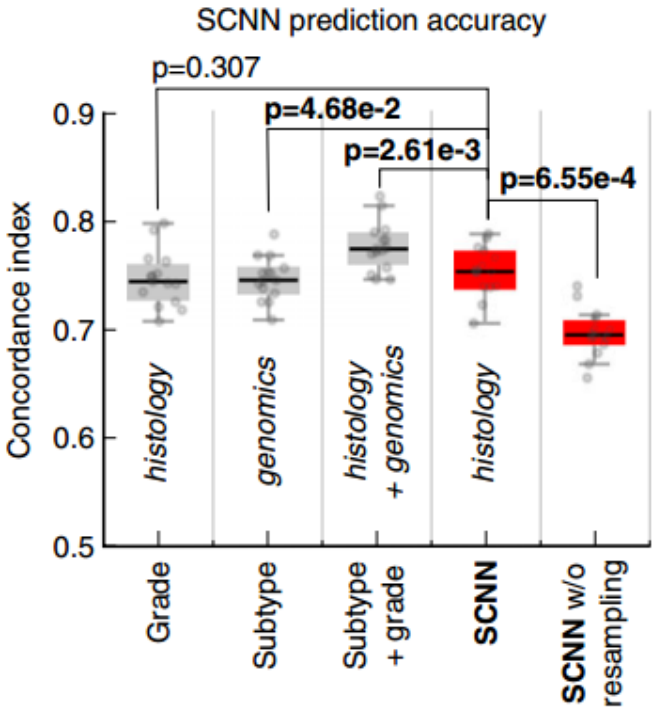
Testing and Model Averaging

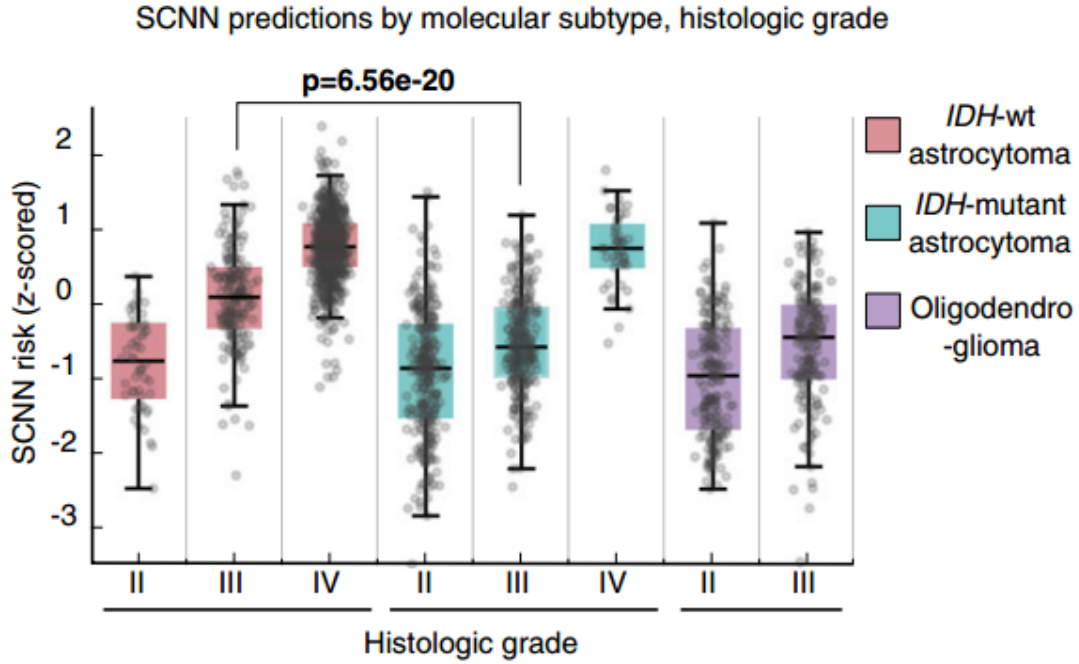
1. $R_m^{j,k}$ denotes the risk of k^{th} HPF in region j for patient m
2. $R_m^j = \text{median}_k\{R_m^{j,k}\}$ (to reject outlying risks)
3. $\widehat{R_m^1} > \widehat{R_m^2} > \widehat{R_m^3} \ldots$
4. $R_m^* = \widehat{R_m^2}$
   - robust to outliers or high risks that may occur due to some imaging or tissue-processing artifact
5. Model averaging: $\overline{R_m^*} = \frac{1}{5}\sum_{\gamma=96}^{100} R_{m(\gamma)}^*$

Model Performance

- Monte Carlo cross-validation (15 times)
  - 80% training
  - 20% testing
- Prediction accuracy was measured using Harrell's c index
  - concordance between predicted risk and actual survival for testing samples

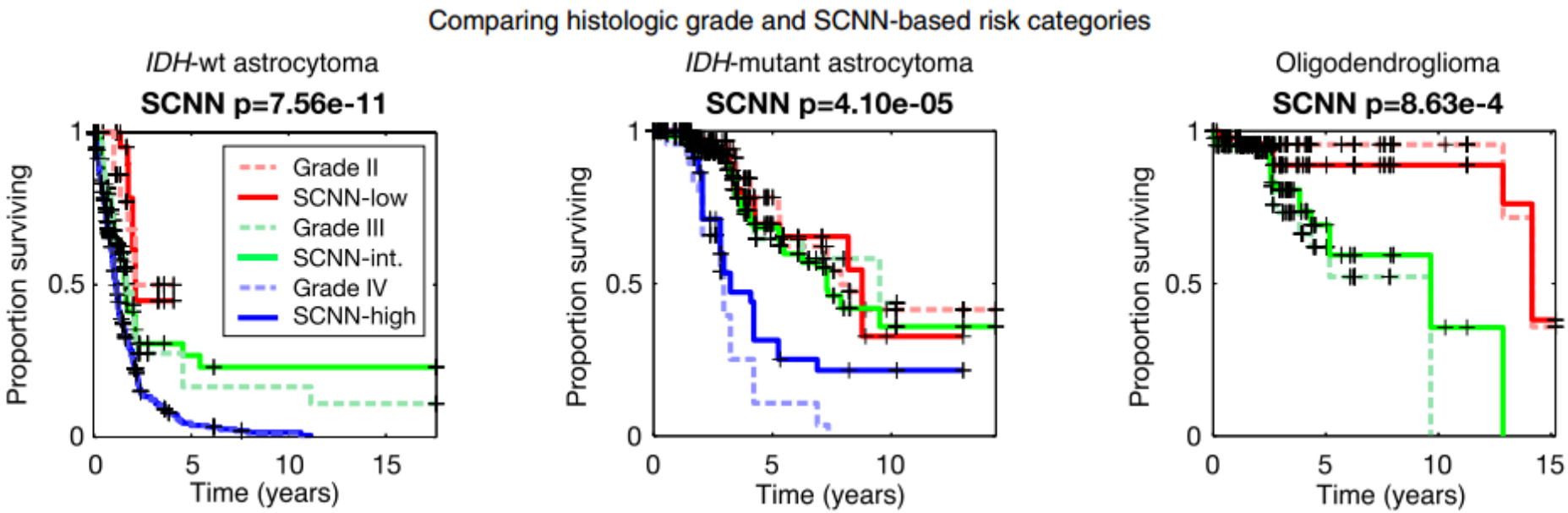
SCNN Performance



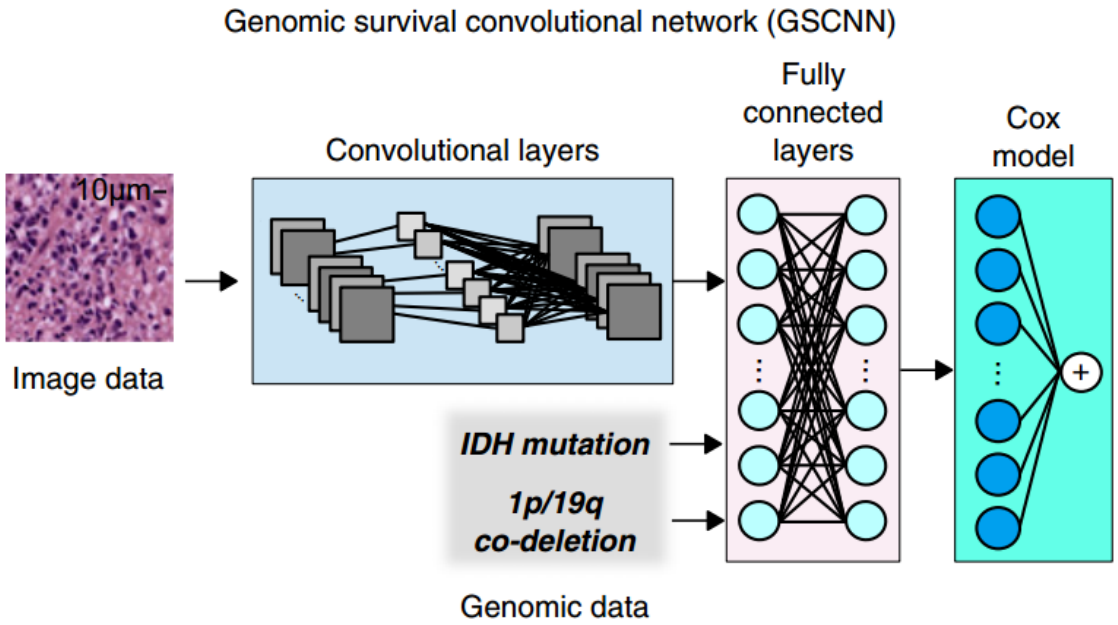


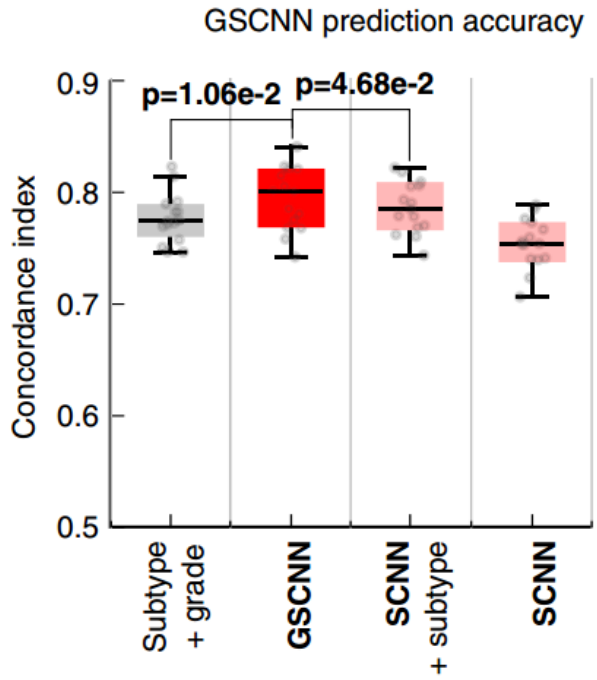
SCNN Performance (cont'd)

Kaplan-Meier plots

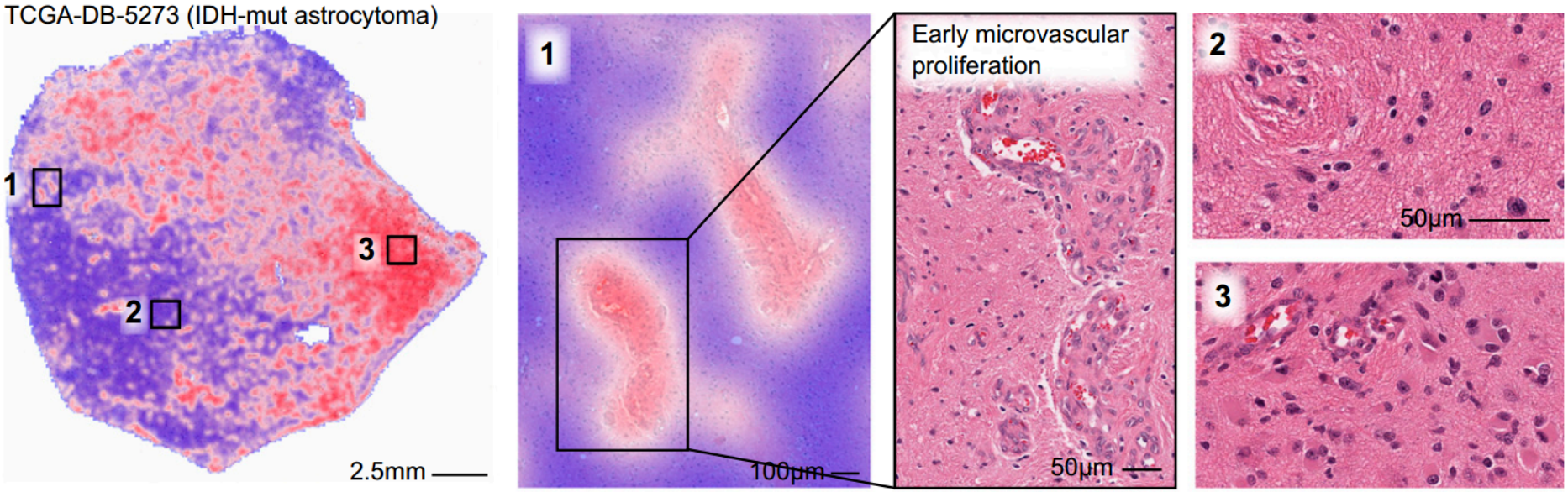


GSCNN Performance

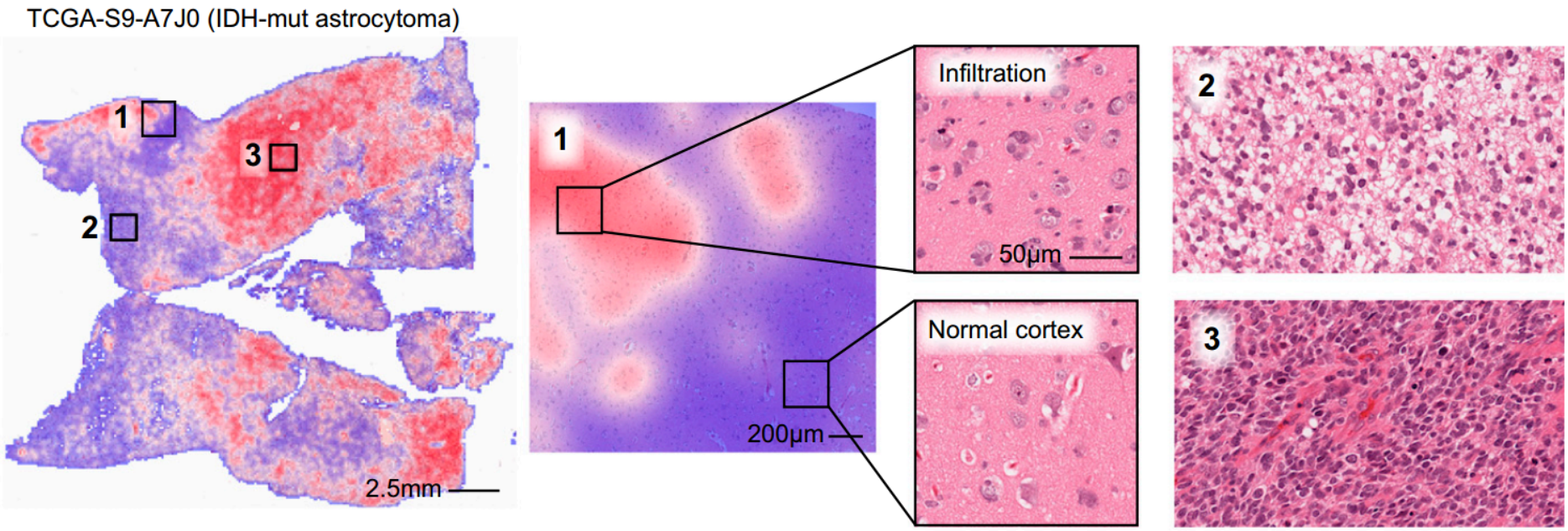




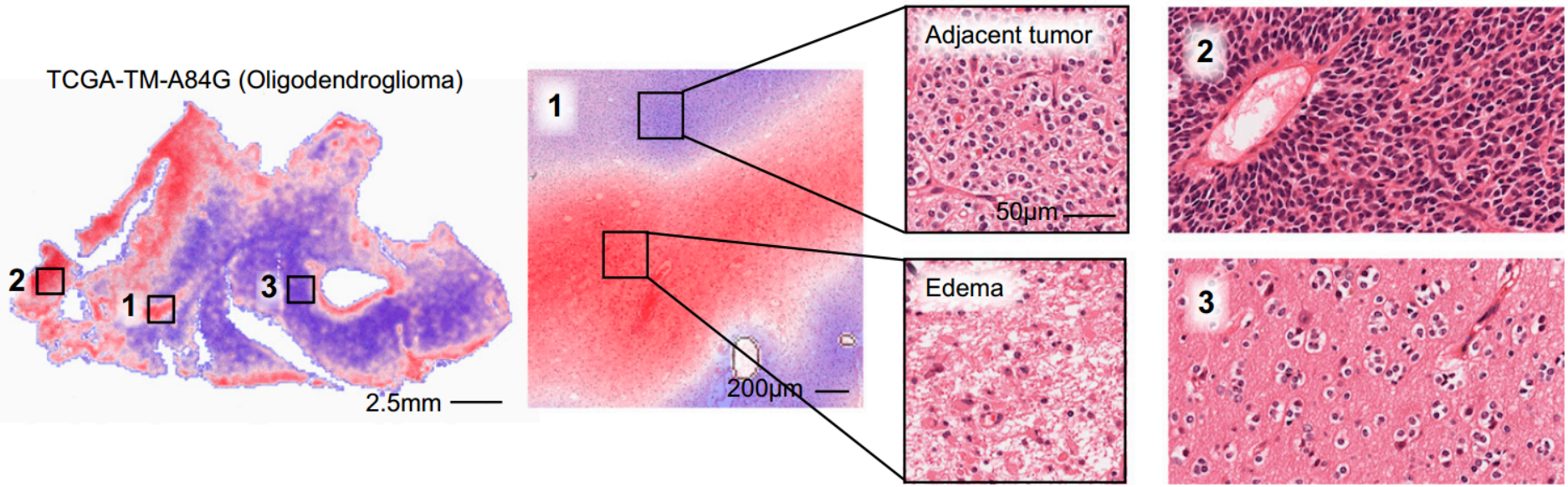
Heatmap Visualization



Heatmap Visualization (cont'd)



Heatmap Visualization (cont'd)



Conclusion

- They proposed a DL method for survival analysis by analyzing histology images
  - image transformations and sampling
  - risk filtering
- Extendend version combines histology images and genomic biomarkers
- Availability: https://github.com/CancerDataScience/SCNN

Questions

Questions?