ETH zürich

Deep learning for automated grading of prostate cancer histopathology images

Eirini Arvaniti

Applied Machine Learning Days 2019

AI & Health

ETH Zurich
Institute of Molecular Systems Biology

PhD supervisor: Manfred Claassen

University Hospital Zurich
Institute of Pathology and Molecular Pathology

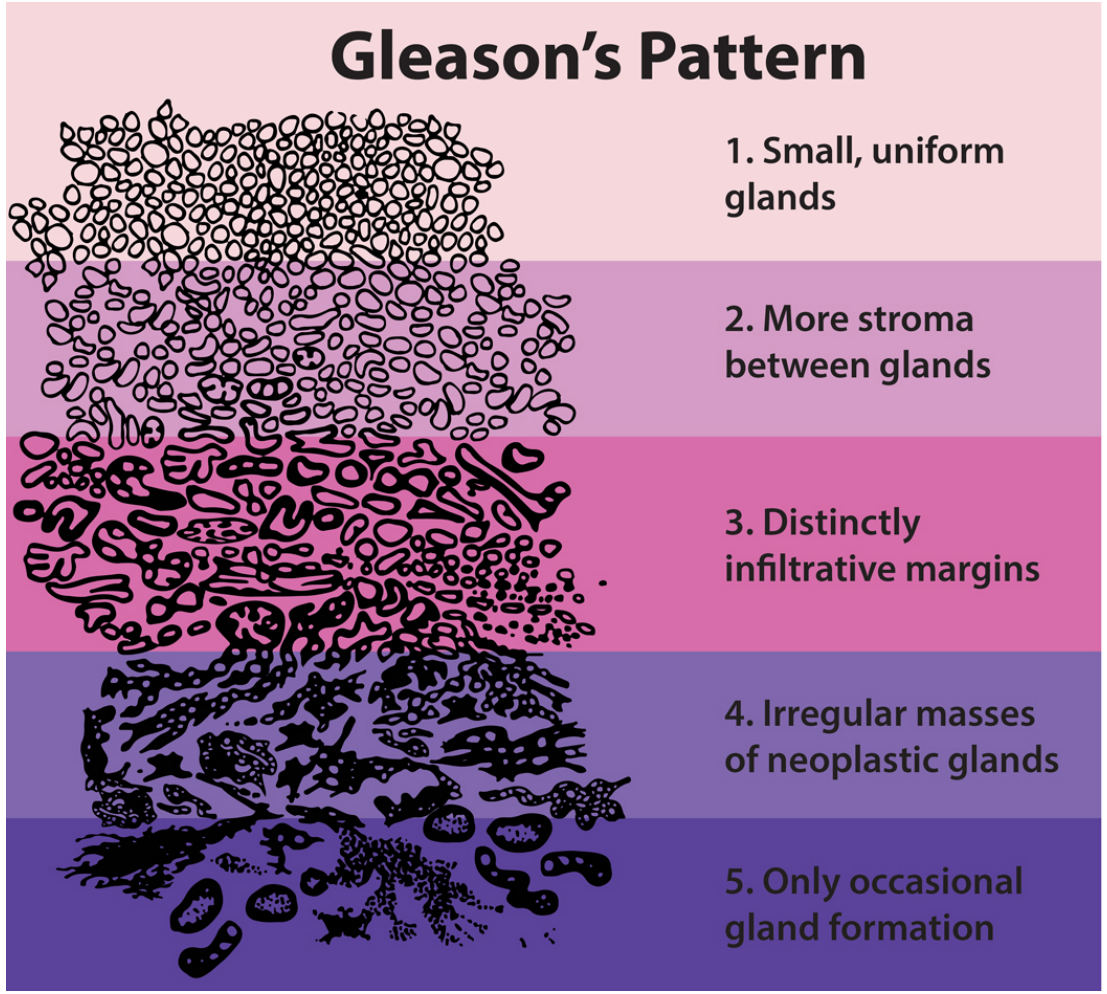
Peter Wild

Kim Fricker

Jan Rüschoff

Acknowledgements

Gleason grading system for prostate cancer



Well differentiated

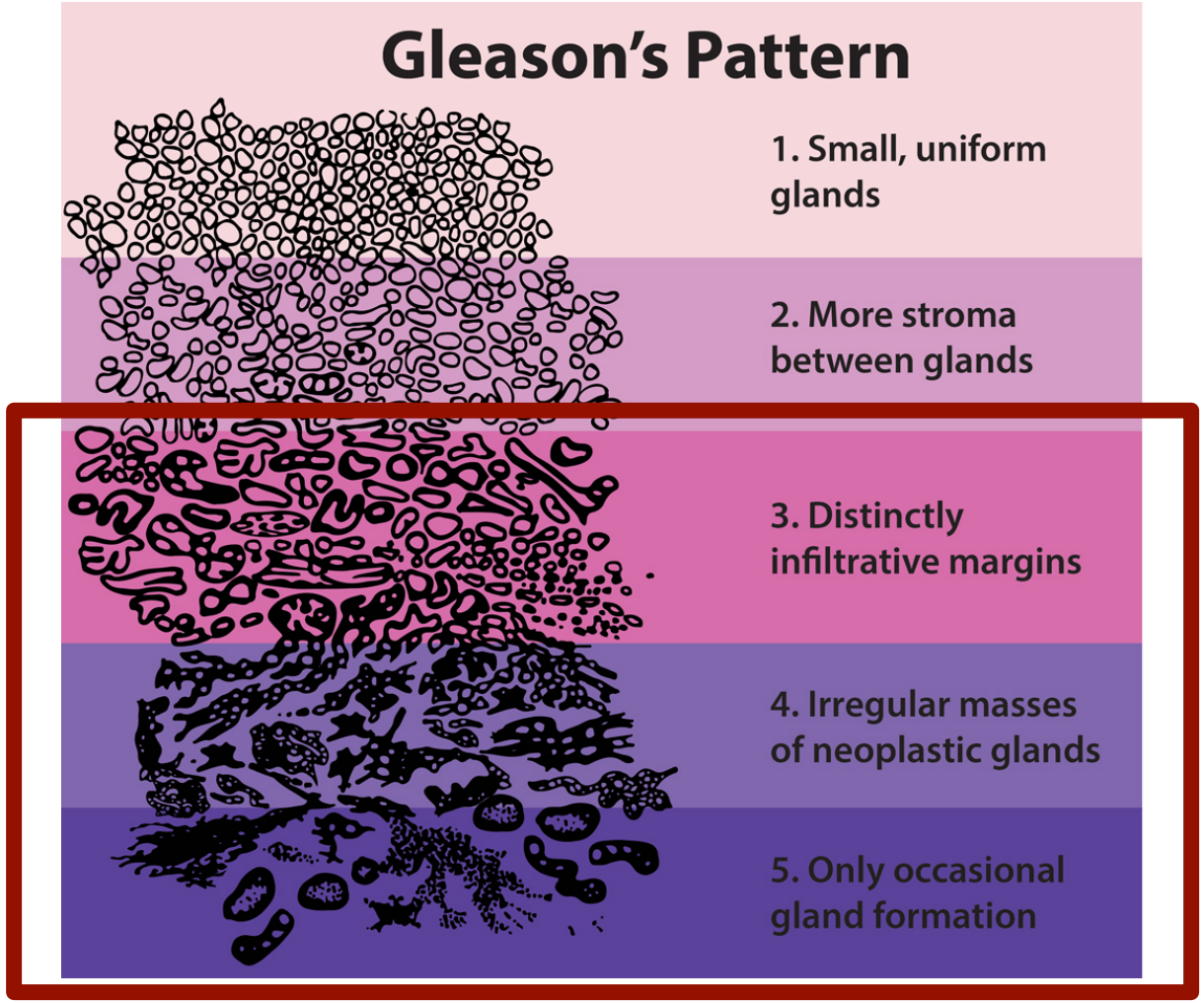
↓

Moderately differentiated

↓

Poorly differentiated/ Anaplastic

Gleason grading system for prostate cancer



Well differentiated

↓

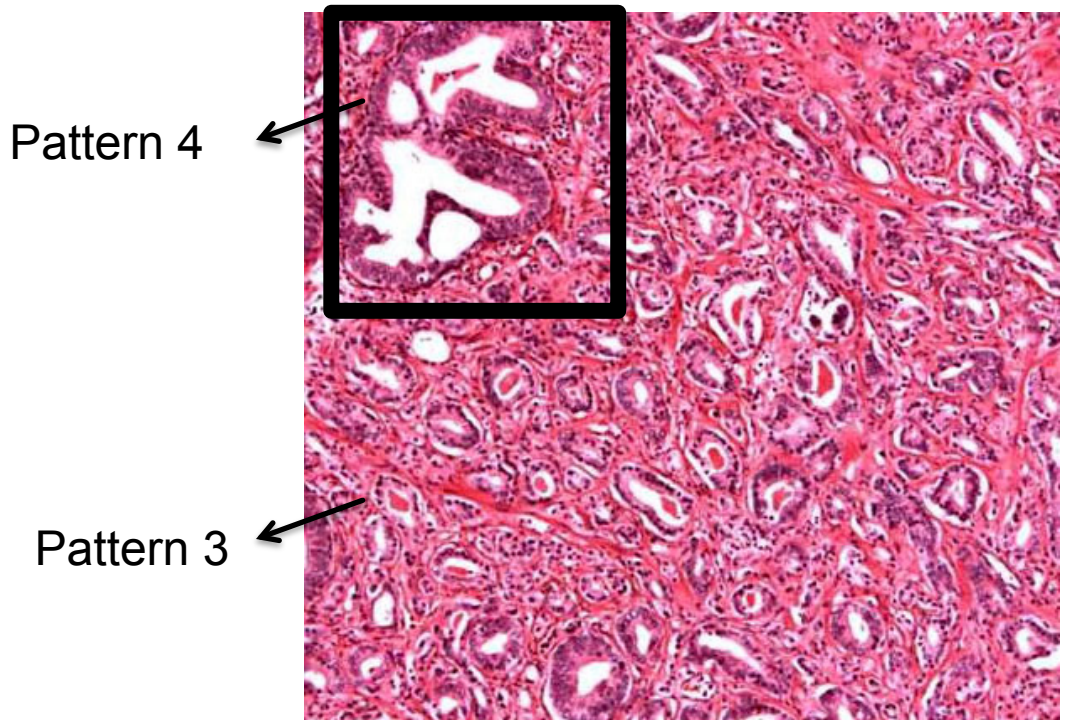
Moderately differentiated

↓

Poorly differentiated/ Anaplastic

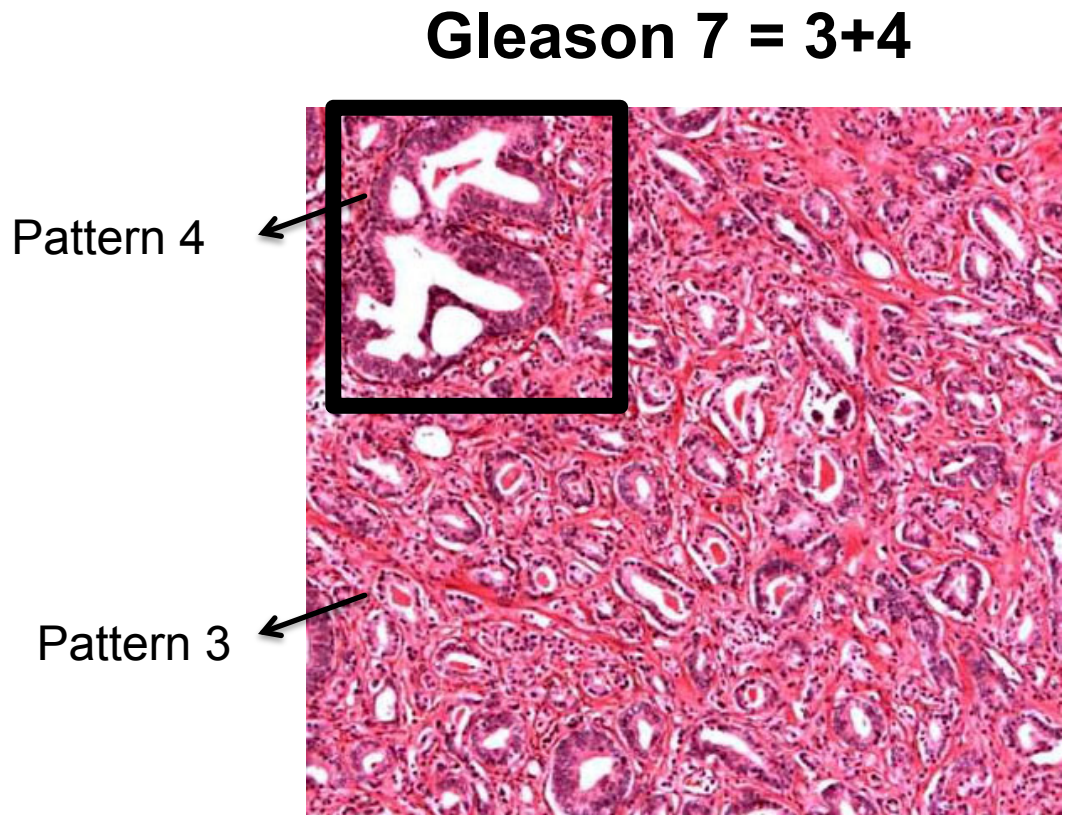
Gleason score assignment by pathologists

Gleason 7 = 3+4



Gleason score assignment by pathologists

- Gleason score predictive of patient prognosis

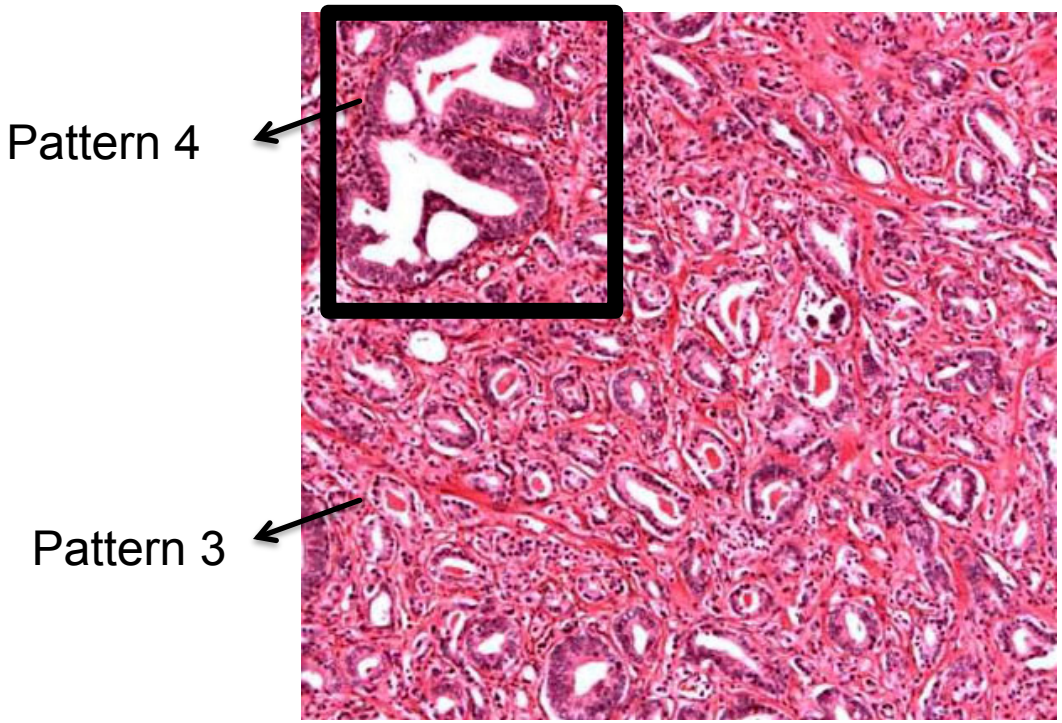


Gleason score assignment by pathologists

- Gleason score predictive of patient prognosis

- Tedious work
- High inter-pathologist variability

Gleason 7 = 3+4



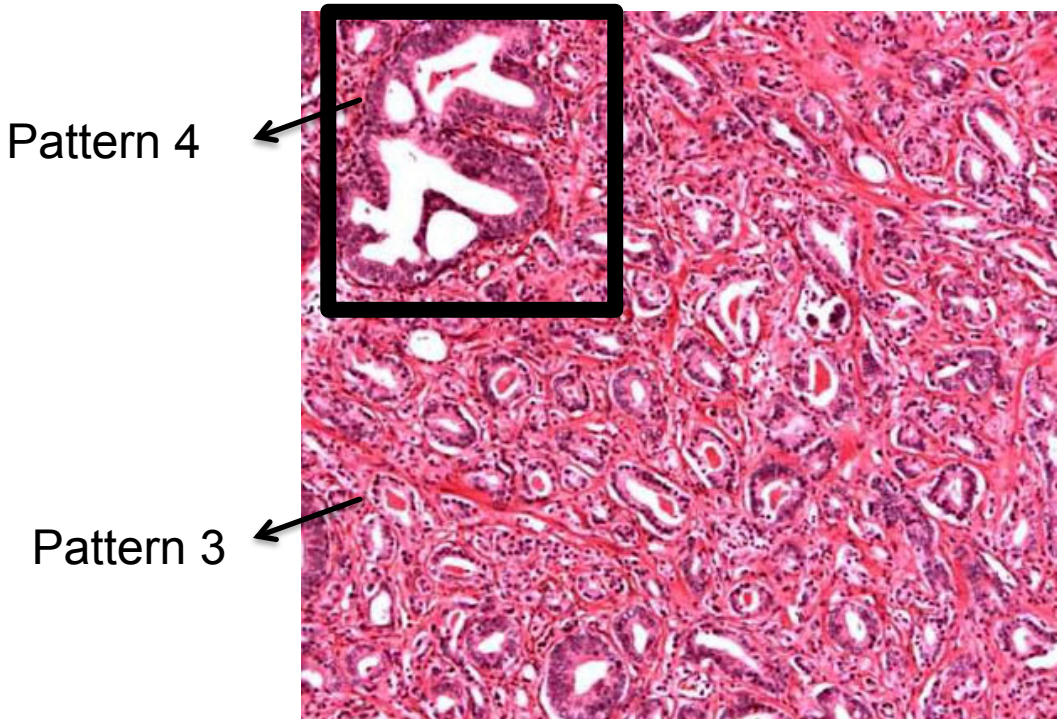
Gleason score assignment by pathologists

- Gleason score predictive of patient prognosis

- Tedious work
- High inter-pathologist variability

- **Project goal:** Gleason score assignment by machine learning
  - ➢ fast and reproducible annotations

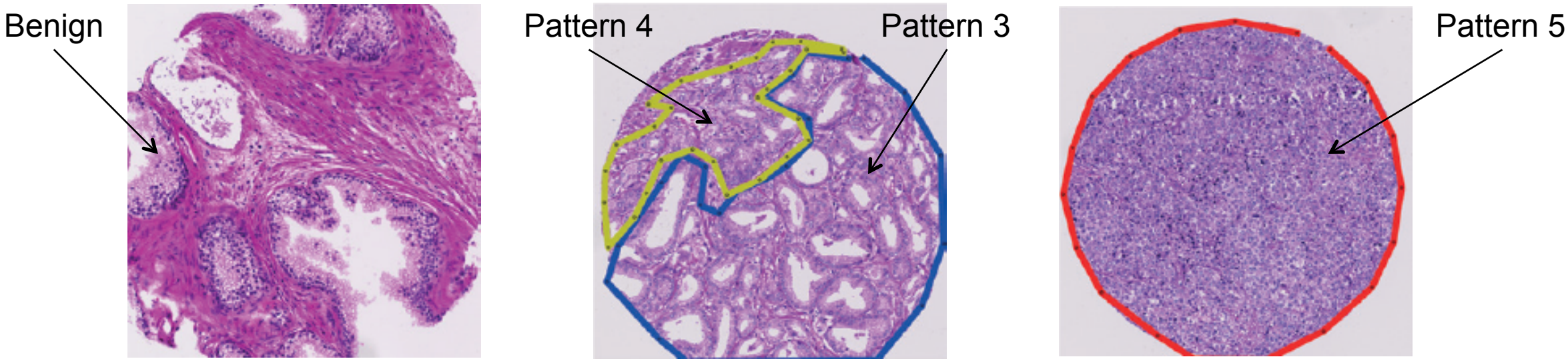
Gleason 7 = 3+4



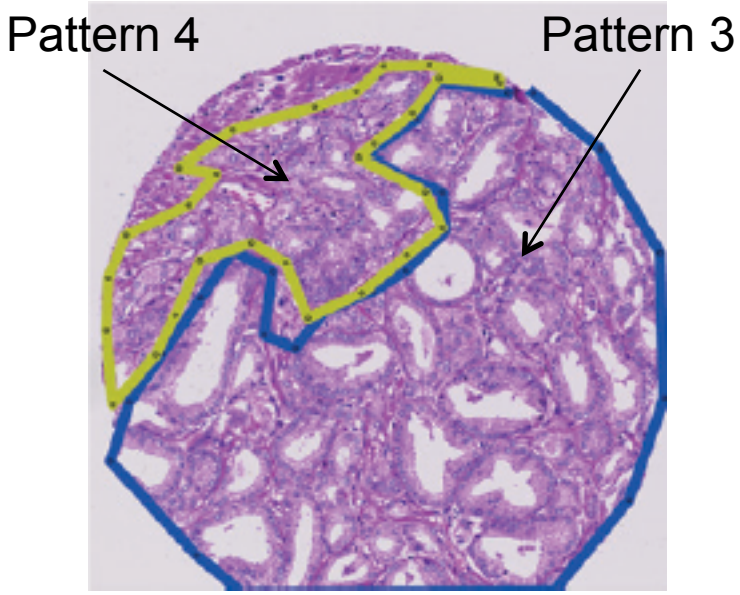
Obtaining ground-truth annotations

K. Fricker

- Domain-expert Gleason annotations on ~ 900 Tissue MicroArray (TMA) images

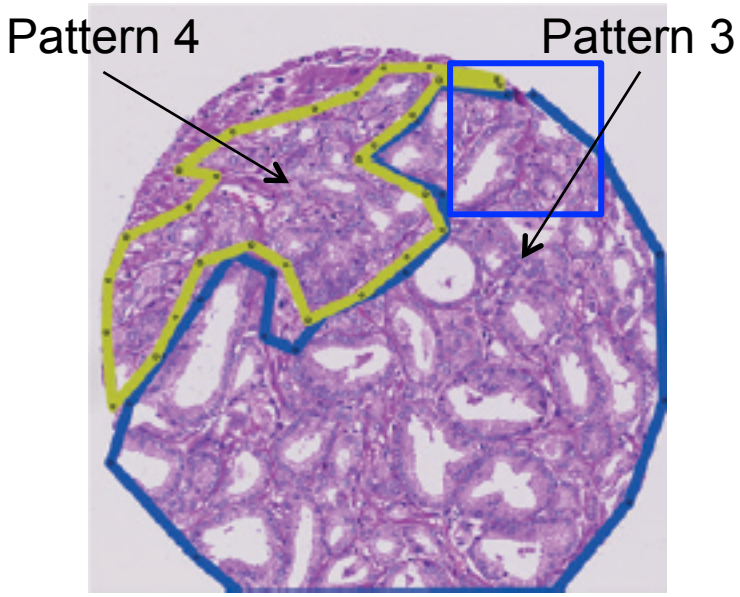


Model training on small image patches



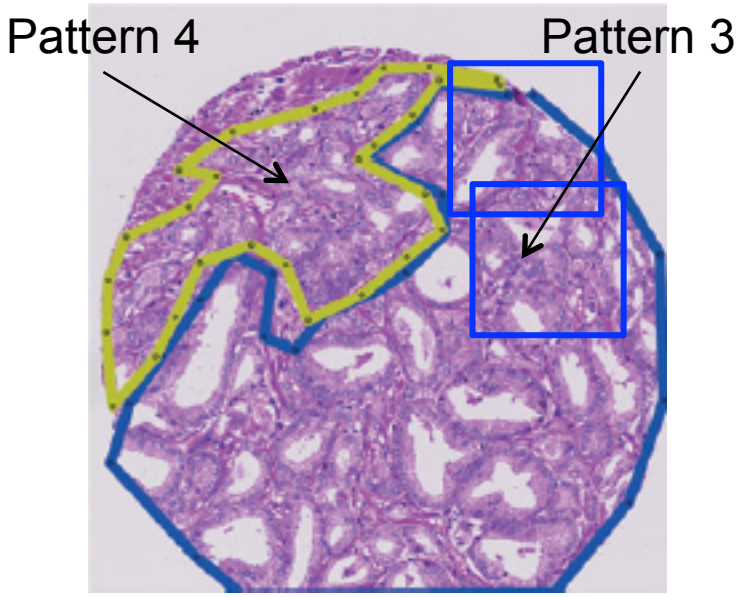
- **Training cohort:** 641 TMA images

Model training on small image patches



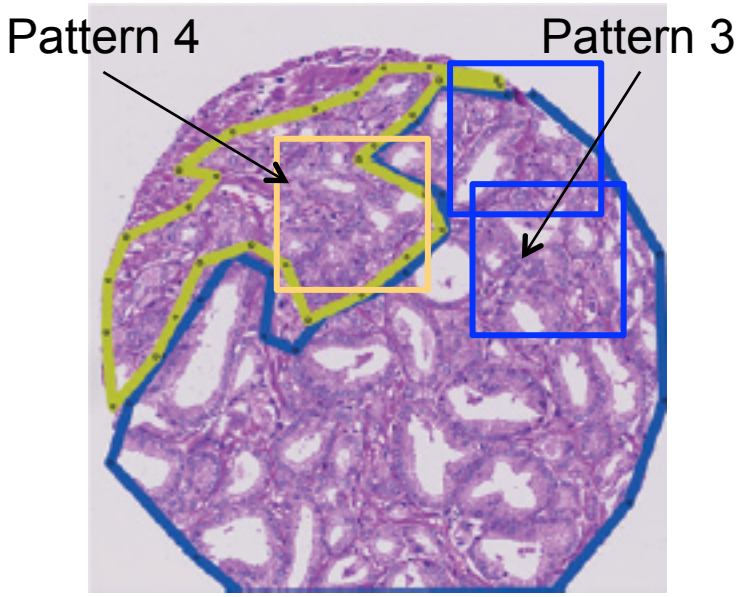
- **Training cohort:** 641 TMA images
- Extract training examples with corresponding Gleason labels

Model training on small image patches



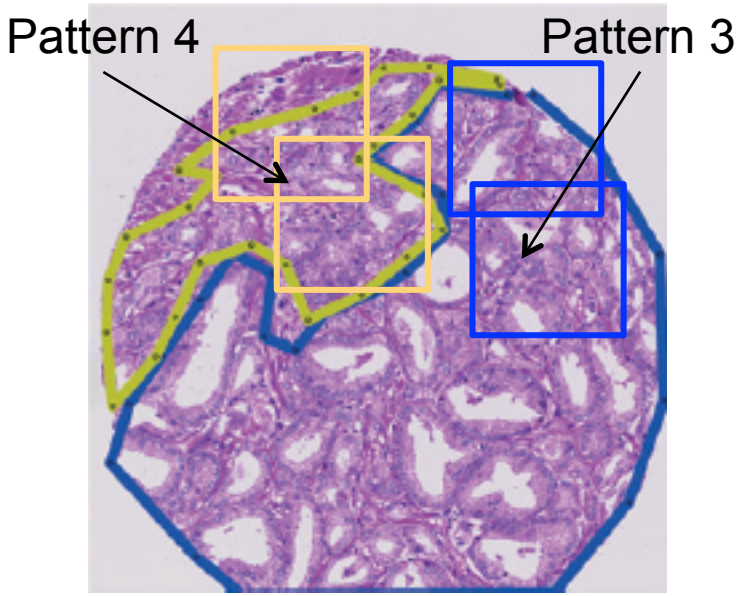
- **Training cohort:** 641 TMA images
- Extract training examples with corresponding Gleason labels

Model training on small image patches



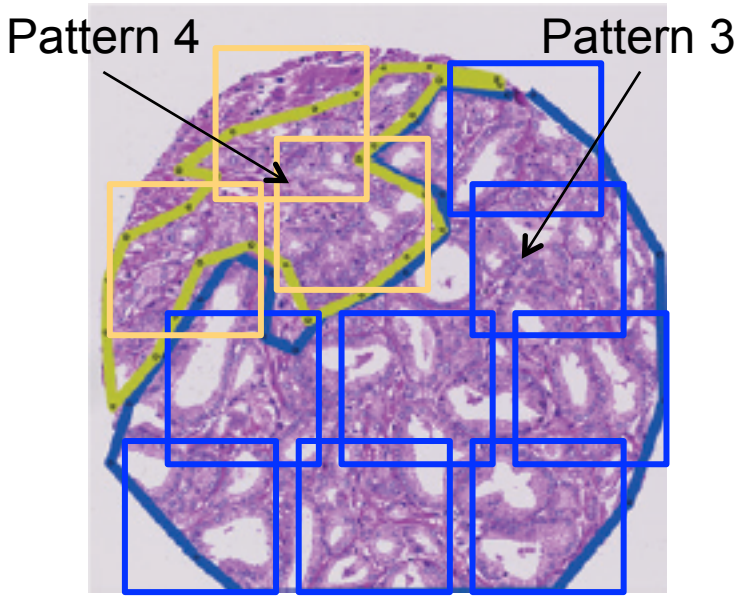
- **Training cohort:** 641 TMA images
- Extract training examples with corresponding Gleason labels

Model training on small image patches



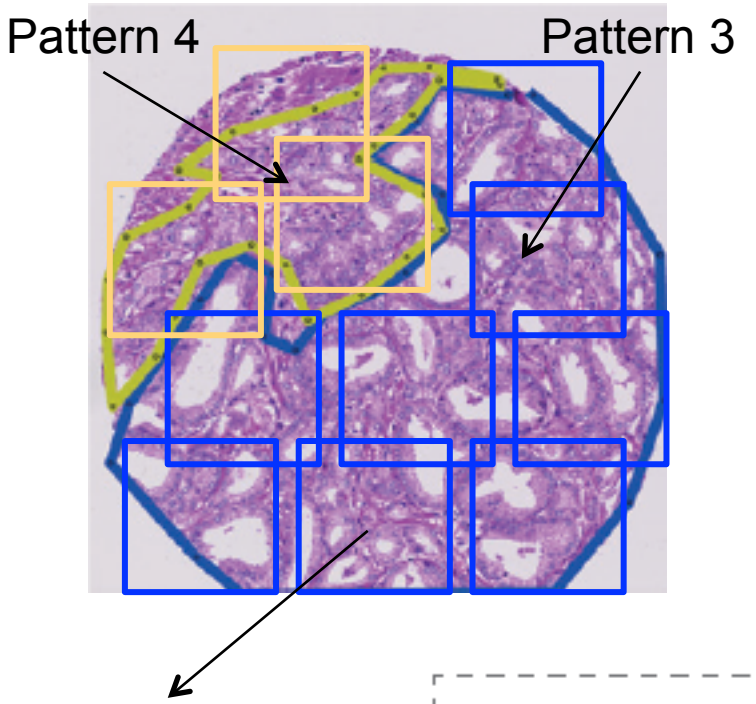
- **Training cohort:** 641 TMA images
- Extract training examples with corresponding Gleason labels

Model training on small image patches

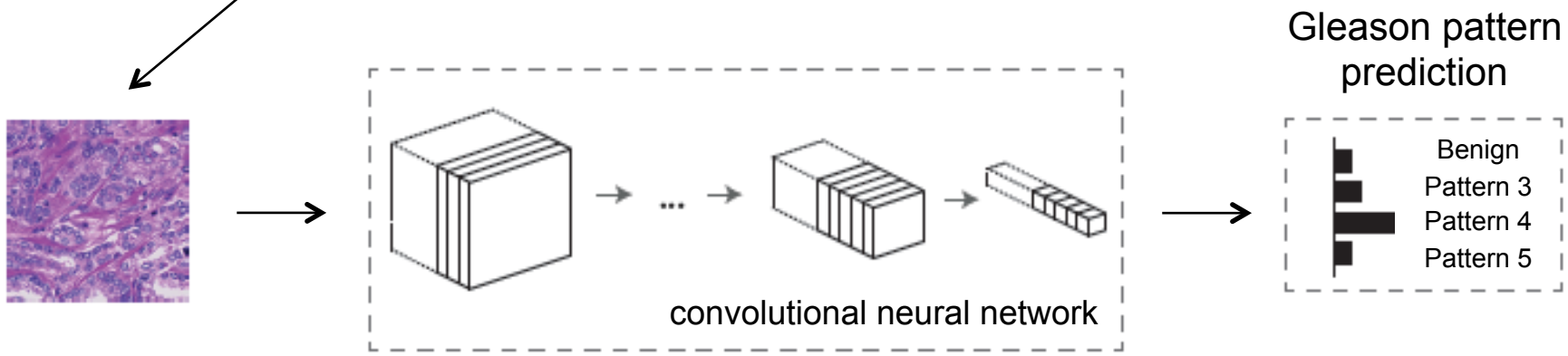


- **Training cohort:** 641 TMA images
- Extract training examples with corresponding Gleason labels

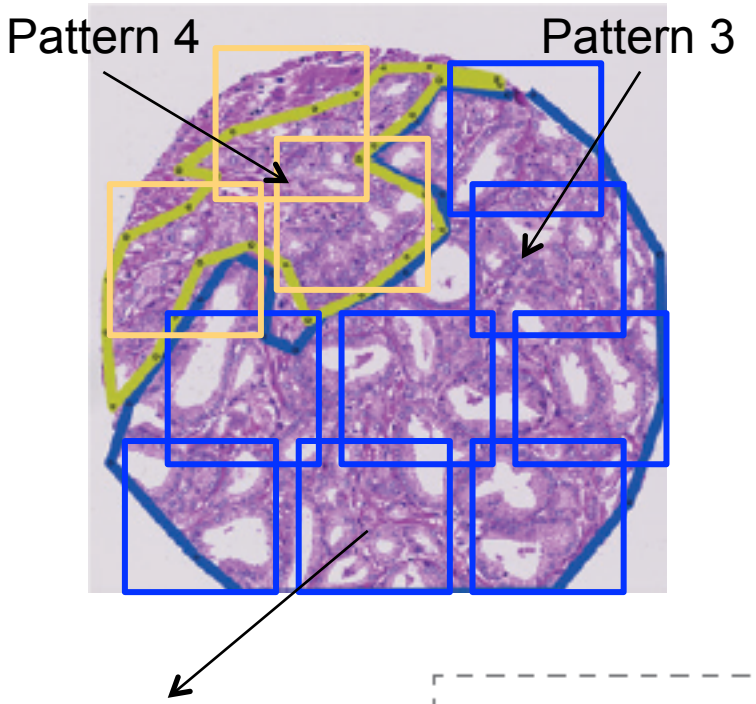
Model training on small image patches



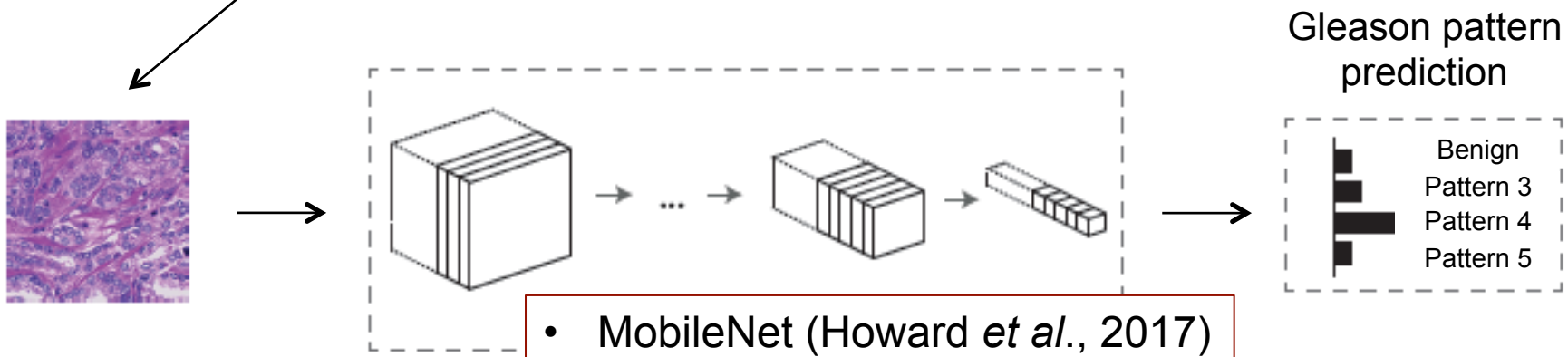
- **Training cohort:** 641 TMA images
- Extract training examples with corresponding Gleason labels
- Train a Gleason pattern classifier



Model training on small image patches

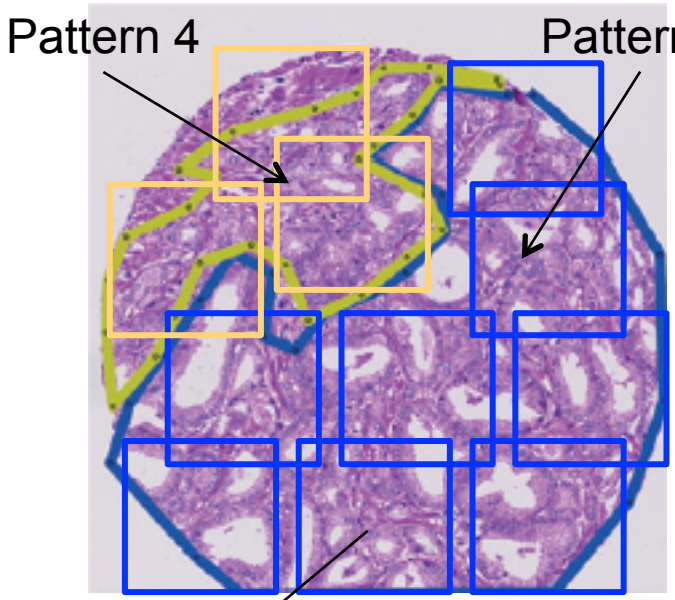


- **Training cohort:** 641 TMA images
- Extract training examples with corresponding Gleason labels
- Train a Gleason pattern classifier



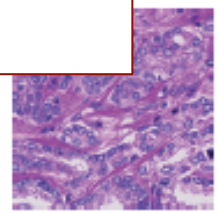
- MobileNet (Howard *et al.*, 2017)
- Fine-tuning from ImageNet

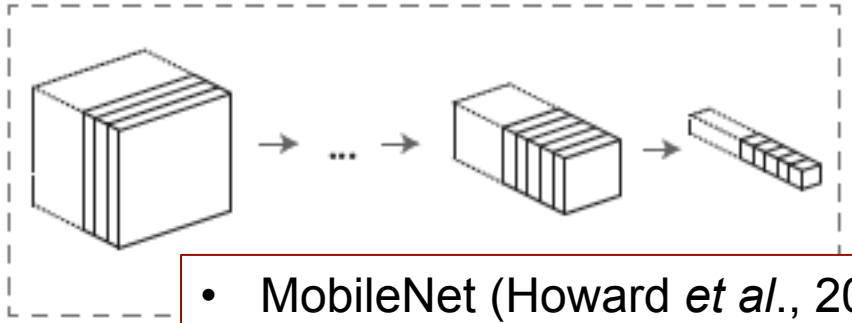
Model training on small image patches



- **Training cohort:** 641 TMA images
- Extract training examples with corresponding Gleason labels
- Train a Gleason pattern classifier

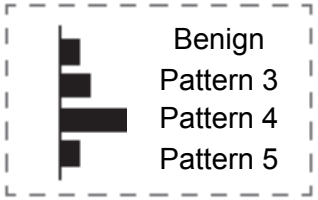
- Data augmentation
- Class-balanced mini-batches



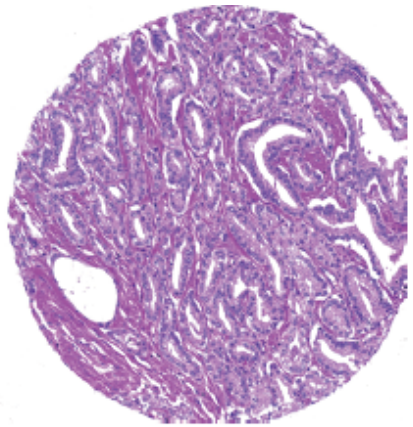


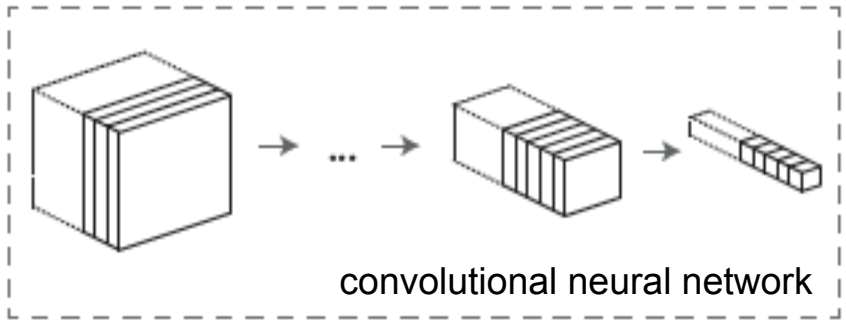
- MobileNet (Howard *et al.*, 2017)
- Fine-tuning from ImageNet

Gleason pattern prediction

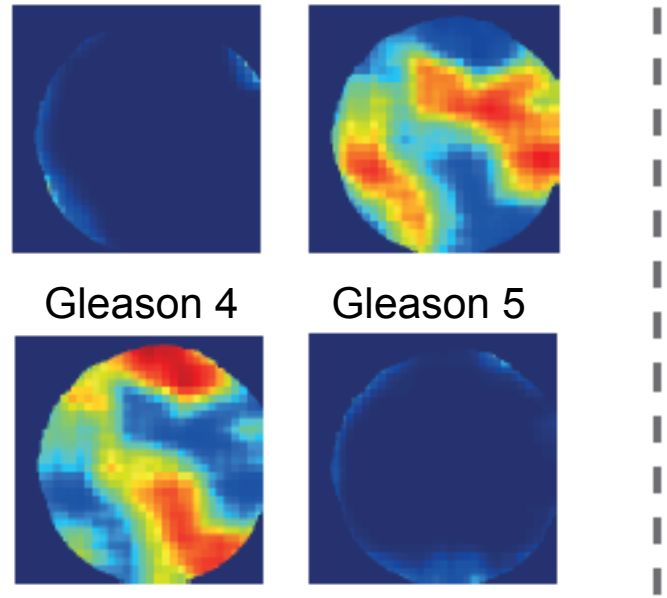


Trained model used to imitate pathologist workflow





Pixel-level annotation



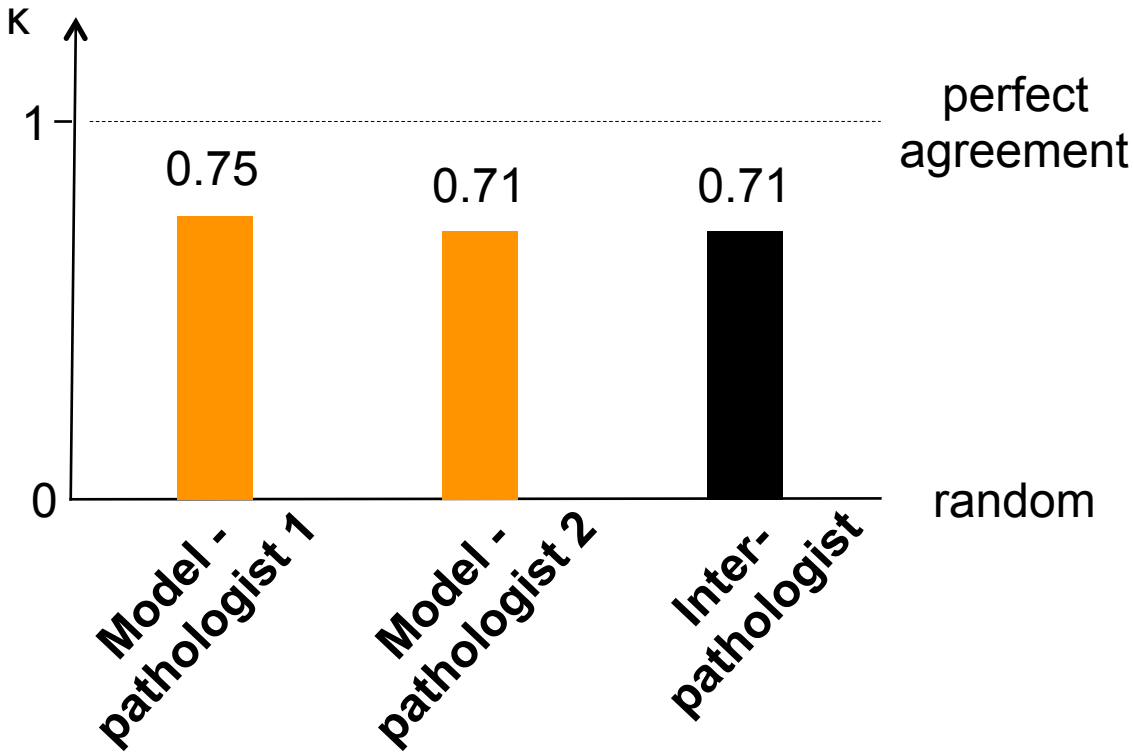
Arvaniti E. *et al*., *Scientific Reports* (2018)

Test cohort results: image-level agreement

K. Fricker J. Rüschoff

Test cohort: 245 TMA images

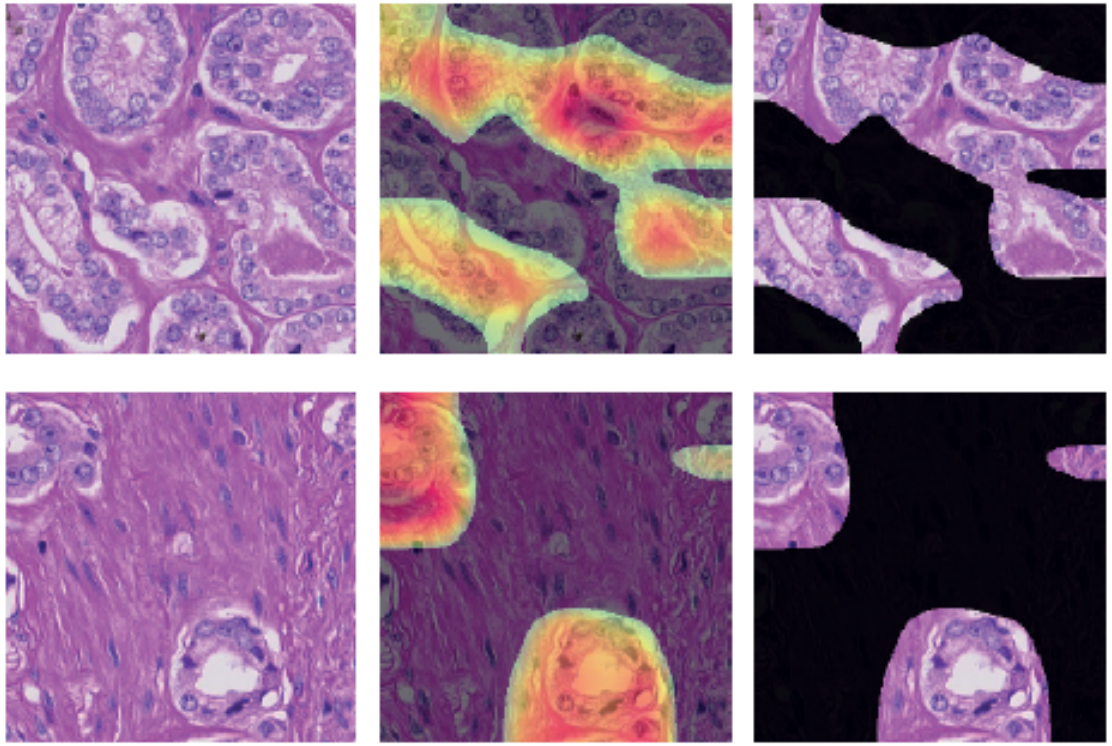
Cohen's κ: agreement rate



Arvaniti E. *et al*., *Scientific Reports* (2018)

Highlighting Gleason-pattern-discriminative regions

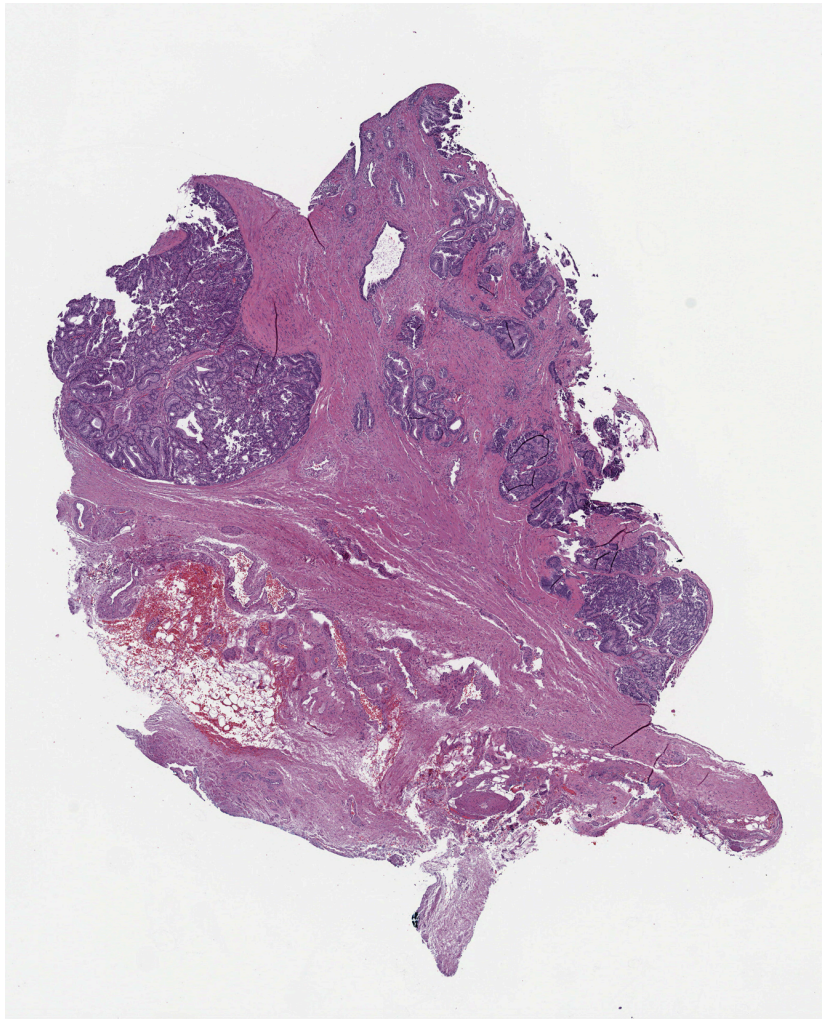
Focus regions for Gleason pattern 3 predictions



input patch | model focus | focus regions

- Model focus obtained by *Class Activation Mapping* (Zhou et al., *CVPR*, 2016)

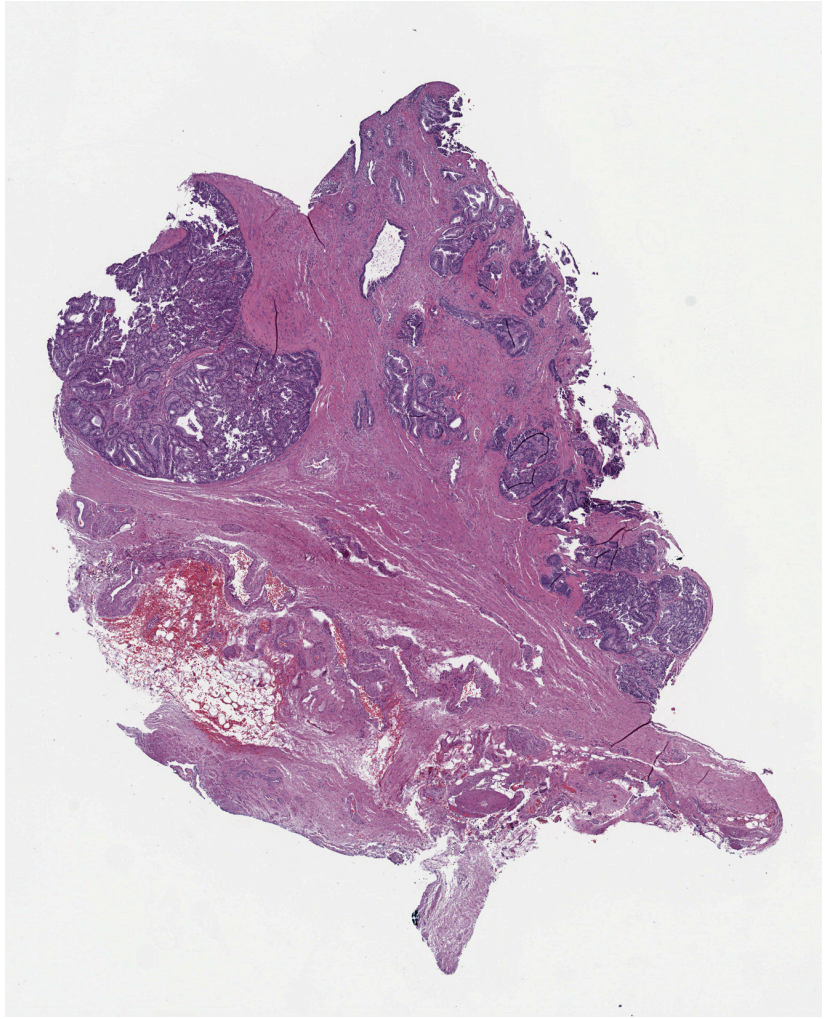
Can we reduce the need for detailed pathologist annotations?



Can we reduce the need for detailed pathologist annotations?

- Global-level (*weak*) annotations often available

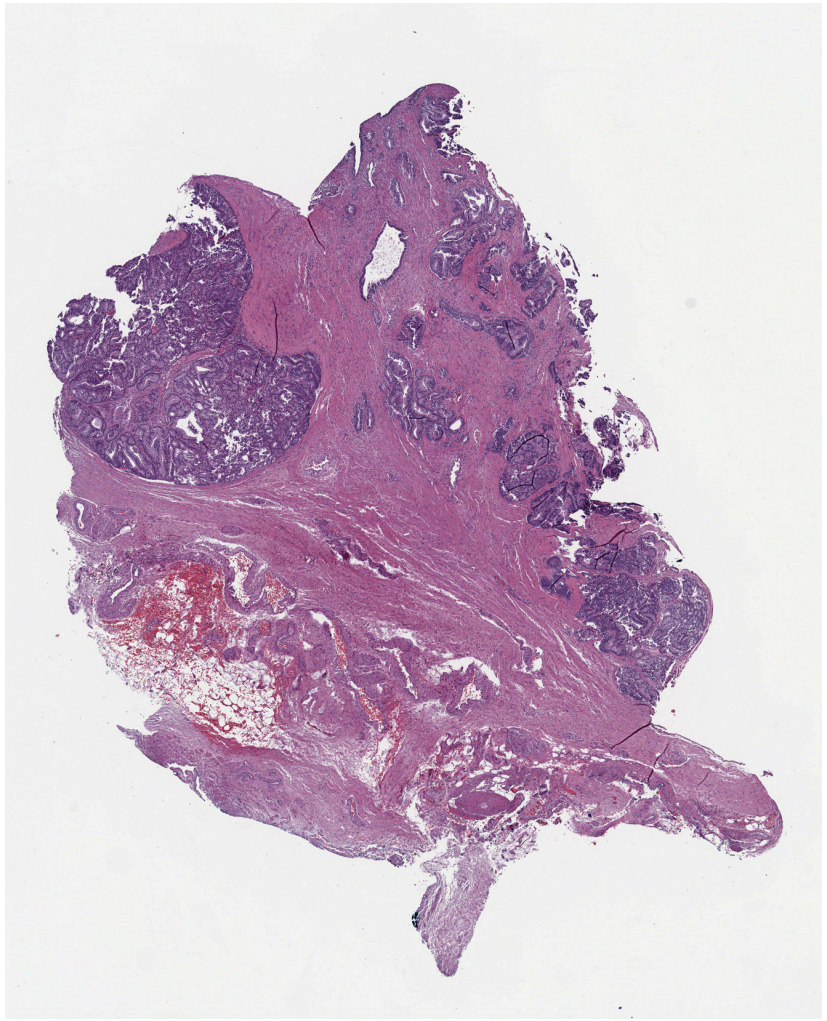
Global-level label: e.g. Gleason 7



Can we reduce the need for detailed pathologist annotations?

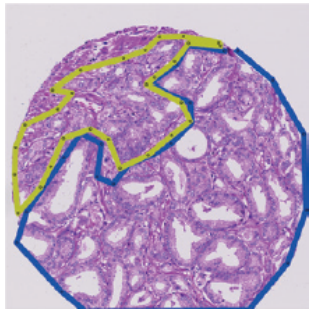
- Global-level (*weak*) annotations often available
- We can *learn* from such *weak annotations*

Global-level label: e.g. Gleason 7

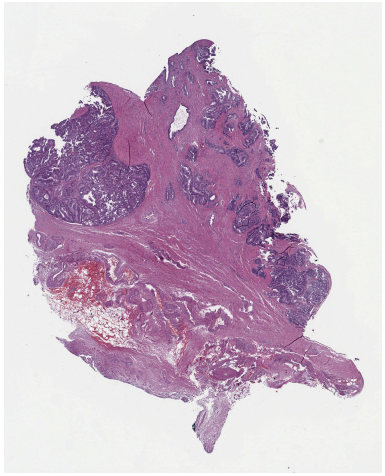


Training with both local- and global-level annotations

Local annotations



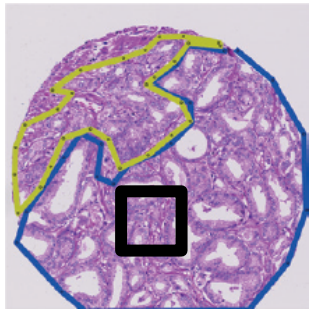
Global (weak) annotation



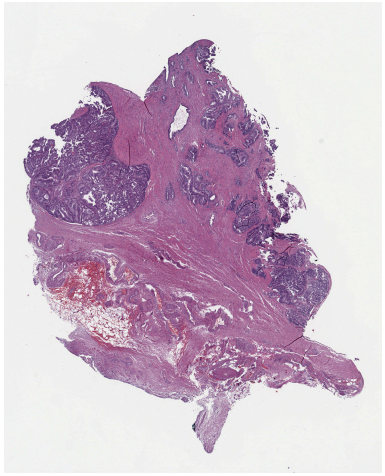
Arvaniti E. *et al.*, *Medical Imaging meets NeurIPS* 2018

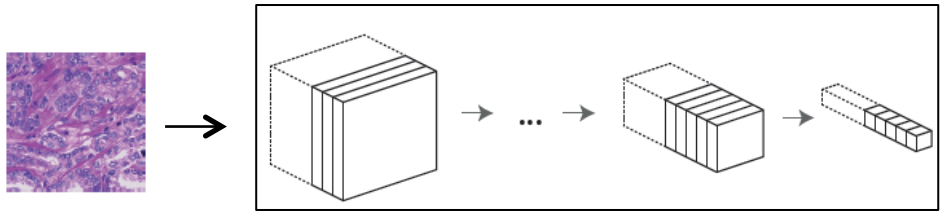
Training with both local- and global-level annotations

Local annotations



Global (weak) annotation





y_{pred}

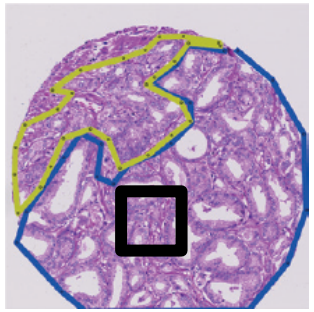
y_{local}

minimize: Loss(y_{pred}, y_{local})

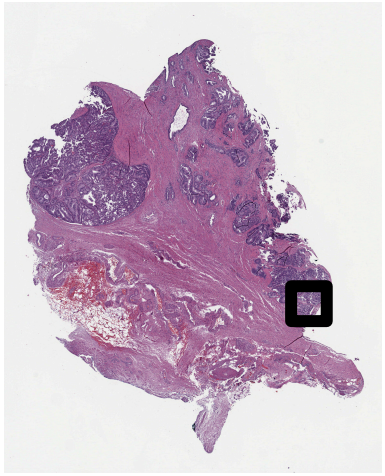
Arvaniti E. *et al.*, *Medical Imaging meets NeurIPS* 2018

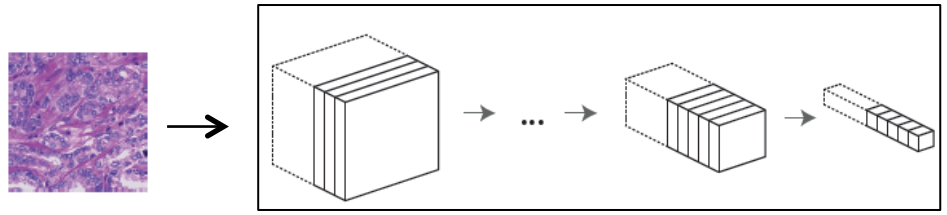
Training with both local- and global-level annotations

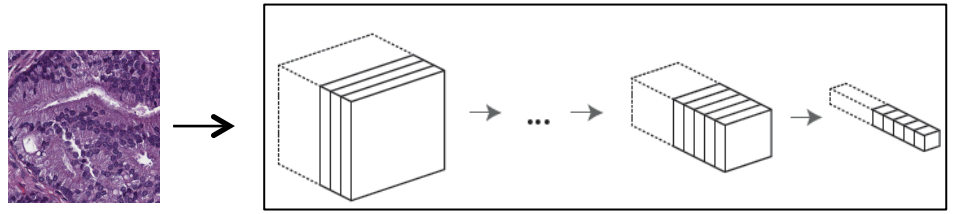
Local annotations



Global (weak) annotation







y_{pred} y_{local}

minimize: Loss(y_{pred}, y_{local})

y_{pred} y_{global}

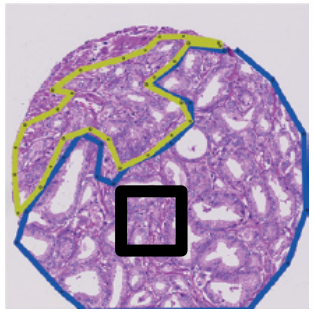
minimize: $\mathbf{w}$ * Loss(y_{pred}, y_{global})

$$0 \leq \mathbf{w} \leq 1$$

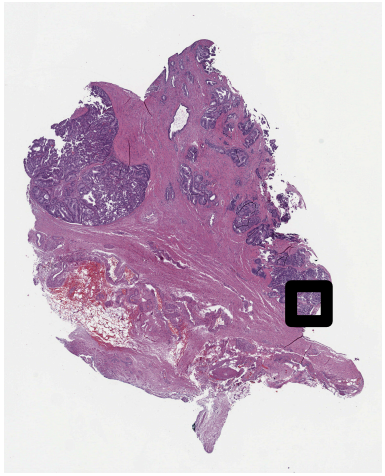
Arvaniti E. *et al.*, *Medical Imaging meets NeurIPS* 2018

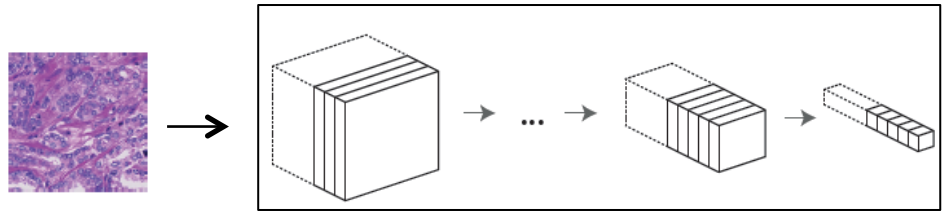
Training with both local- and global-level annotations

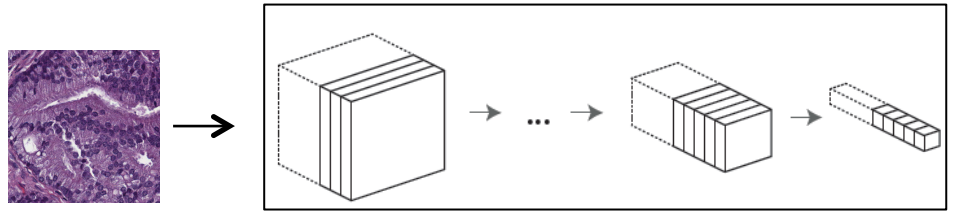
Local annotations



Global (weak) annotation







y_{pred} y_{local}

minimize: Loss(y_{pred}, y_{local})

y_{pred} y_{global}

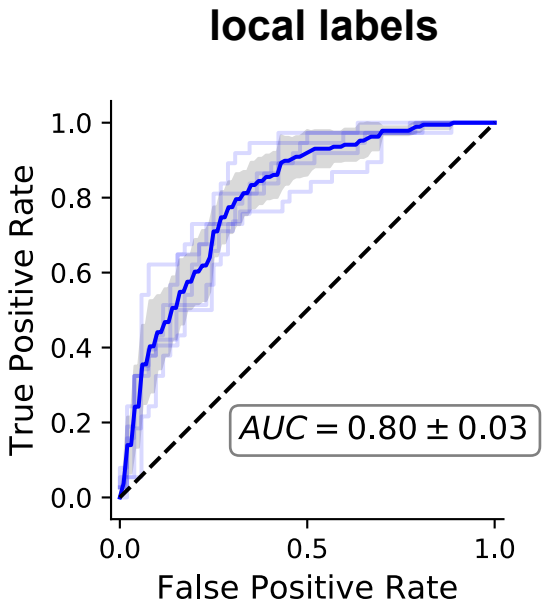
minimize: **w** * Loss(y_{pred}, y_{global})

$$0 \leq \mathbf{w} \leq 1$$

w: predicted probability for the given (weak) label.

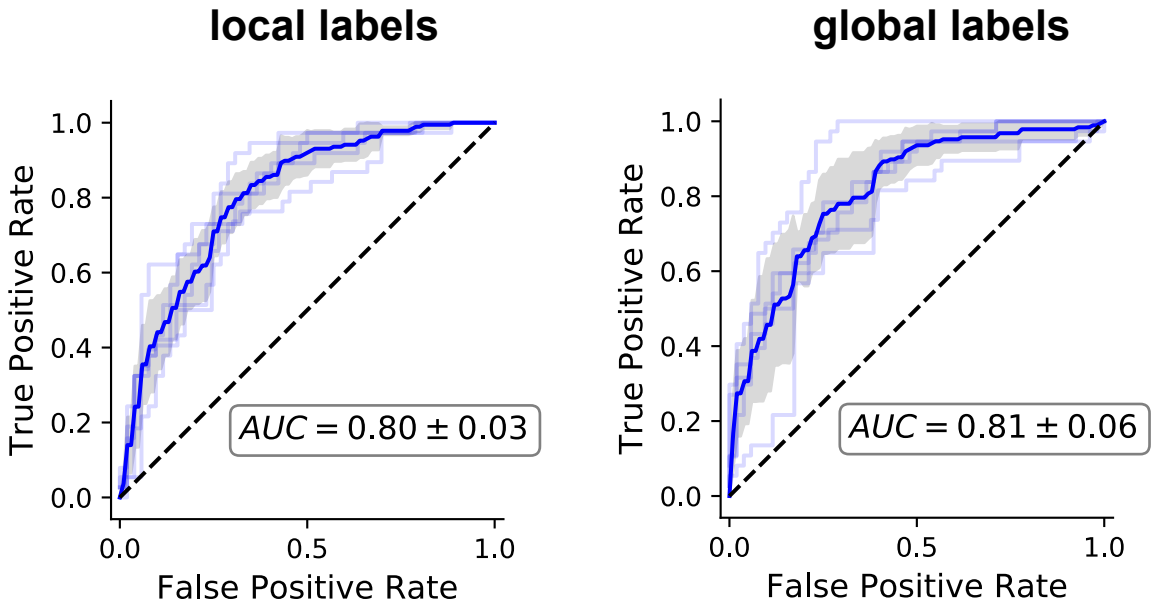
Arvaniti E. *et al.*, *Medical Imaging meets NeurIPS* 2018

Weighted weak supervision outperforms simpler approaches



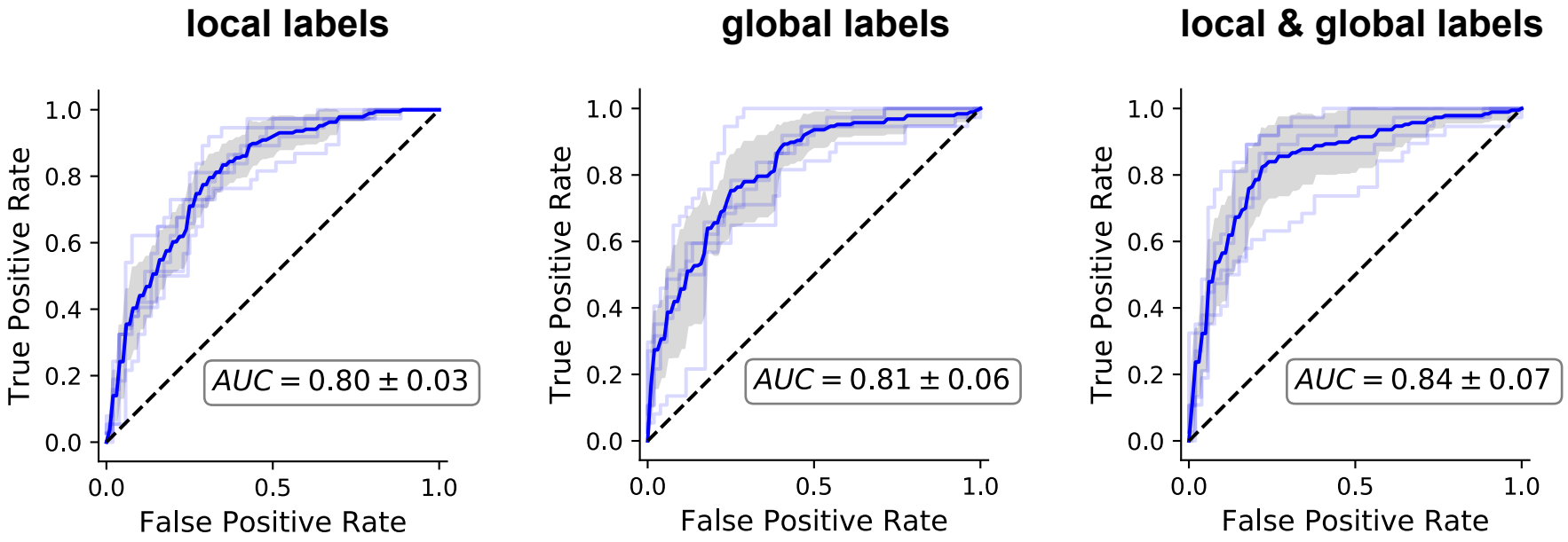
- **Task:** classify *low vs high Gleason score* cases from The Cancer Genome Atlas (TCGA) whole-slide images.

Weighted weak supervision outperforms simpler approaches



- **Task:** classify *low vs high Gleason score* cases from The Cancer Genome Atlas (TCGA) whole-slide images.

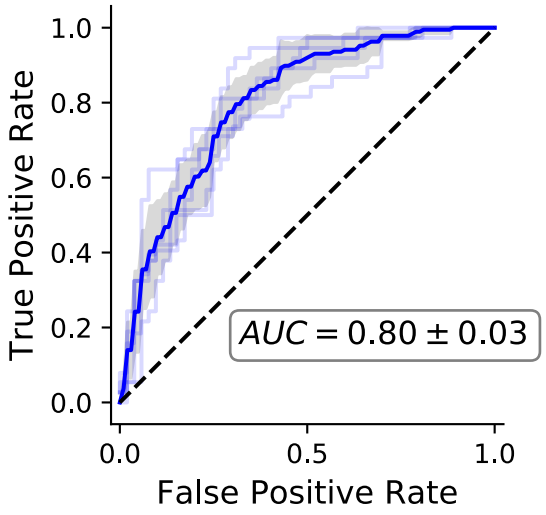
Weighted weak supervision outperforms simpler approaches



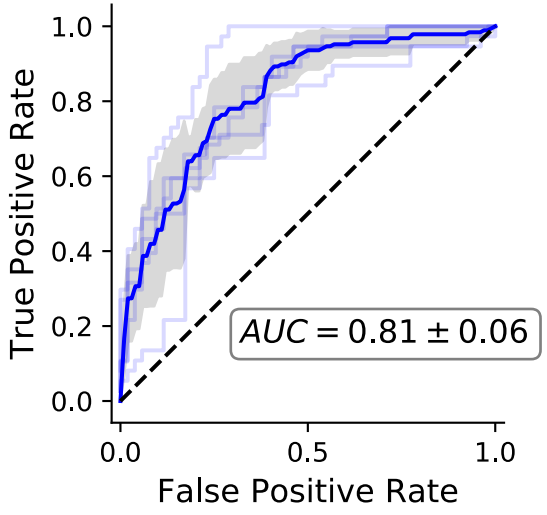
- **Task:** classify *low vs high Gleason score* cases from The Cancer Genome Atlas (TCGA) whole-slide images.

Weighted weak supervision outperforms simpler approaches

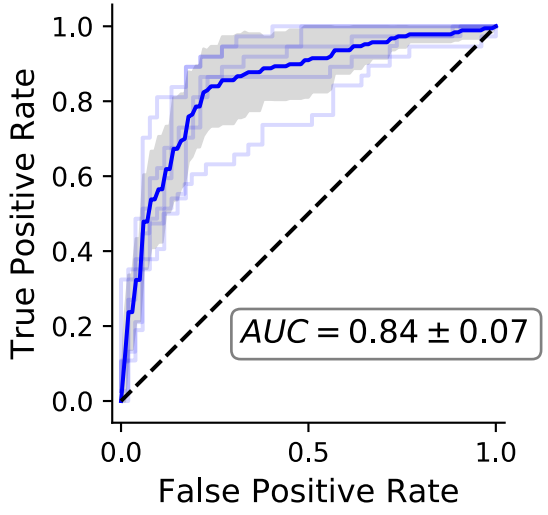
local labels



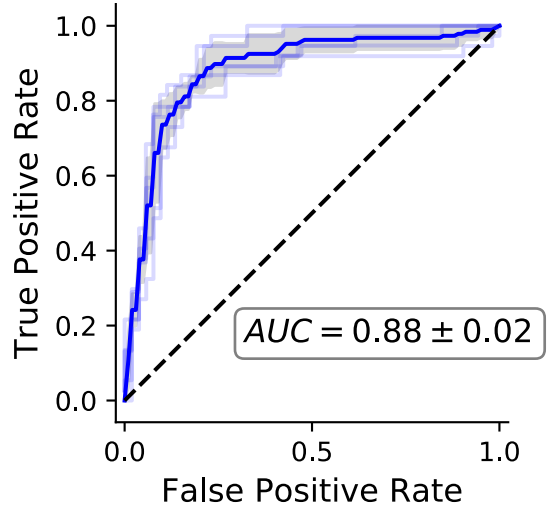
global labels



local & global labels



local & weighted global labels



- **Task:** classify *low vs high Gleason score* cases from The Cancer Genome Atlas (TCGA) whole-slide images.

Summary & Outlook

- Automated Gleason grading with machine learning

Summary & Outlook

- Automated Gleason grading with machine learning
- Weakly-supervised training:
  - ➢ reduced need for expert annotations

Summary & Outlook

- Automated Gleason grading with machine learning
- Weakly-supervised training:
  - ➢ reduced need for expert annotations

Next steps

- Validate weakly-supervised training locally

Summary & Outlook

- Automated Gleason grading with machine learning
- Weakly-supervised training:
  - ➢ reduced need for expert annotations

Next steps

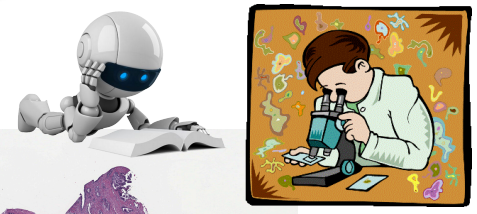
- Validate weakly-supervised training locally

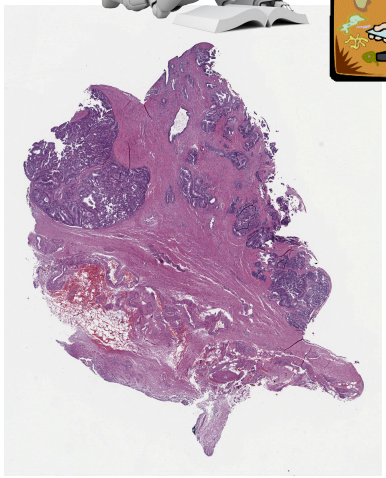
- Beyond predicting human annotations:
  - ➢ associate tissue images with molecular features, cancer recurrence, …

Summary & Outlook

- Automated
- Weakly-sup
  - ➢ reduced ne ations

Thank you for your attention!





Next steps

- Validate weakly-supervised training locally

- Beyond predicting human annotations:
  - ➢ associate tissue images with molecular features, cancer recurrence, …