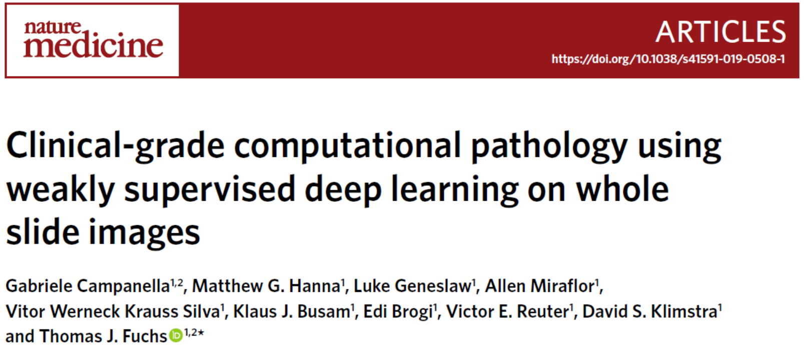


nature medicine

ARTICLES

https://doi.org/10.1038/s41591-019-0508-1

Clinical-grade computational pathology using weakly supervised deep learning on whole slide images

Gabriele Campanella[1,2], Matthew G. Hanna[1], Luke Geneslaw[1], Allen Miraflor[1], Vitor Werneck Krauss Silva[1], Klaus J. Busam[1], Edi Brogi[1], Victor E. Reuter[1], David S. Klimstra[1] and Thomas J. Fuchs[1,2]*

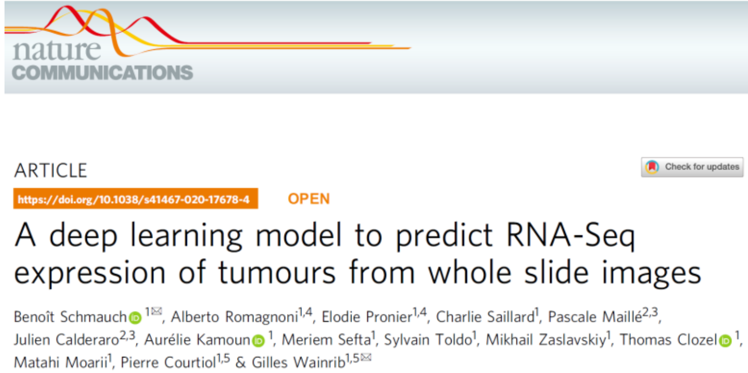
LUXEMBOURG
INSTITUTE
OF **HEALTH**

Multiomics Data Science

Journal Club:
Methods for Deep Analysis in Digital Pathology

Petr Nazarov

petr.nazarov@lih.lu



nature COMMUNICATIONS

ARTICLE

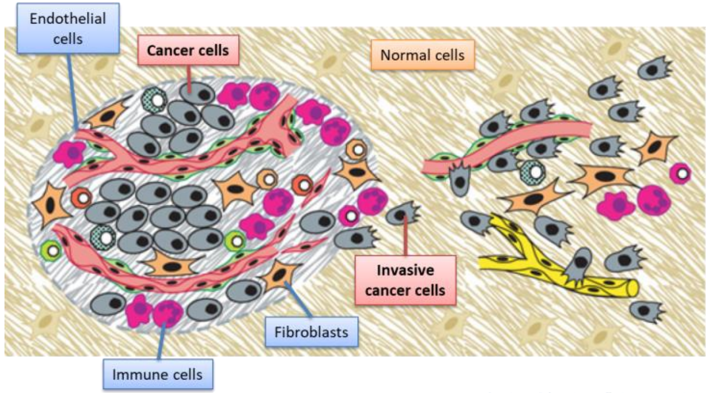
Check for updates

https://doi.org/10.1038/s41467-020-17678-4 OPEN

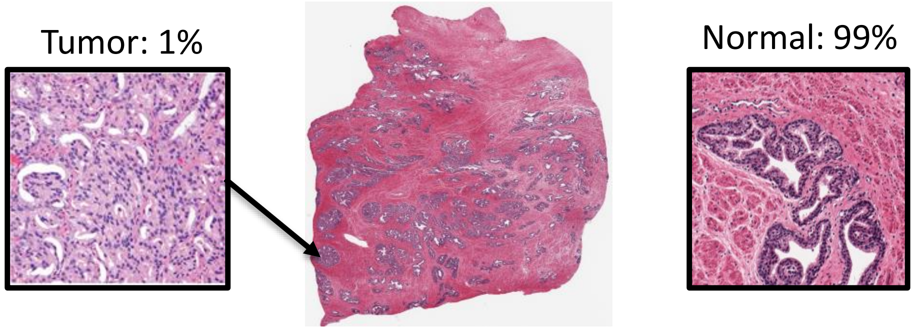
A deep learning model to predict RNA-Seq expression of tumours from whole slide images

Benoît Schmauch[1]✉, Alberto Romagnoni[1,4], Elodie Pronier[1,4], Charlie Saillard[1], Pascale Maillé[2,3], Julien Calderaro[2,3], Aurélie Kamoun[1], Meriem Sefta[1], Sylvain Toldo[1], Mikhail Zaslavskiy[1], Thomas Clozel[1], Matahi Moarii[1], Pierre Courtiol[1,5] & Gilles Wainrib[1,5]✉

Background



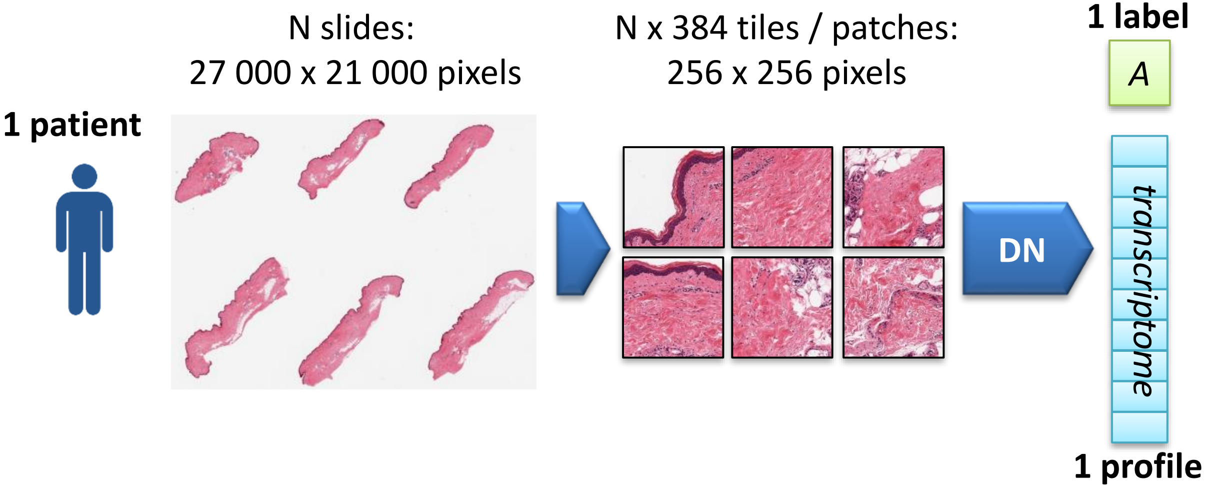
Hanahan D, Weinberg RA. *Cell* 2011, 144, 646-74



- ➢ Native heterogeneity of tissues
- ➢ Inter/intra tumor heterogeneity

Issues in histopathological image analysis:

- ➢ Tedious analysis
- ➢ In some cancers (e.g. prostate) < 1% of the image is cancer-related
- ➢ Standard approaches require supervised "pixel-wise" labelling - unrealistic



1 patient

N slides:
27 000 x 21 000 pixels

N x 384 tiles / patches:
256 x 256 pixels

DN

1 label

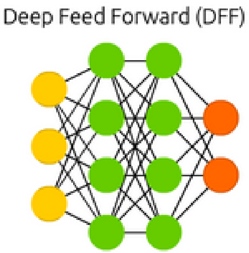
A

transcriptome

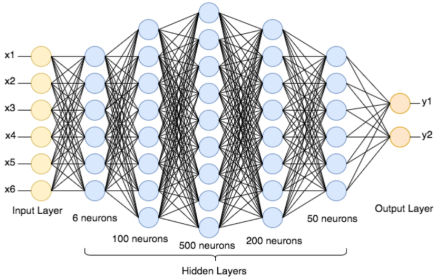
1 profile

Artificial Neural Networks

MLP



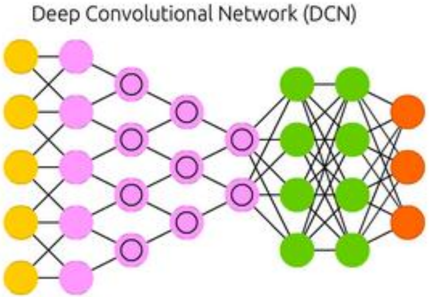
Multilayer perceptron, a.k.a.
(Deep) feed-forward network,
back-propagation network
fully-connected layers,
etc…



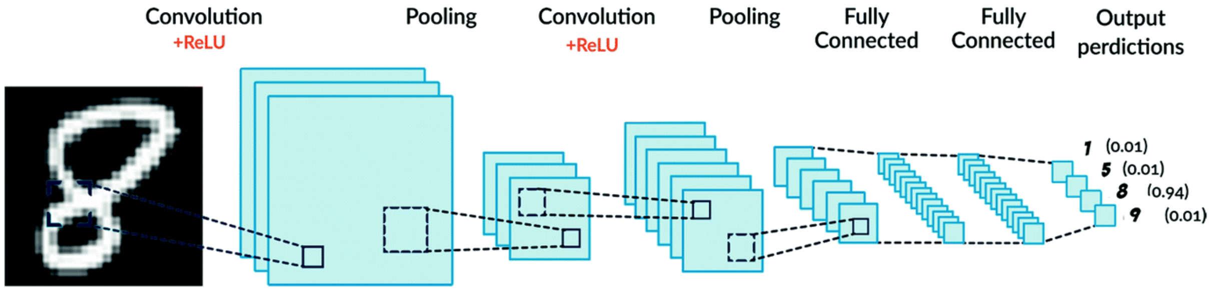
My first "love"… ☺

Nazarov et al (2004)
J Chem Inf Comput Sci

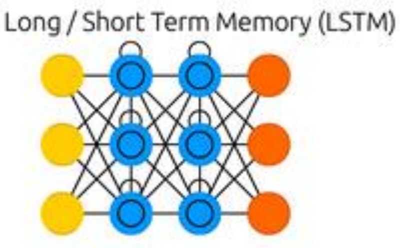
CNN



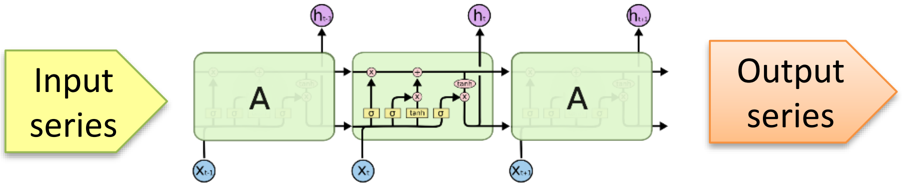
Convolutional networks



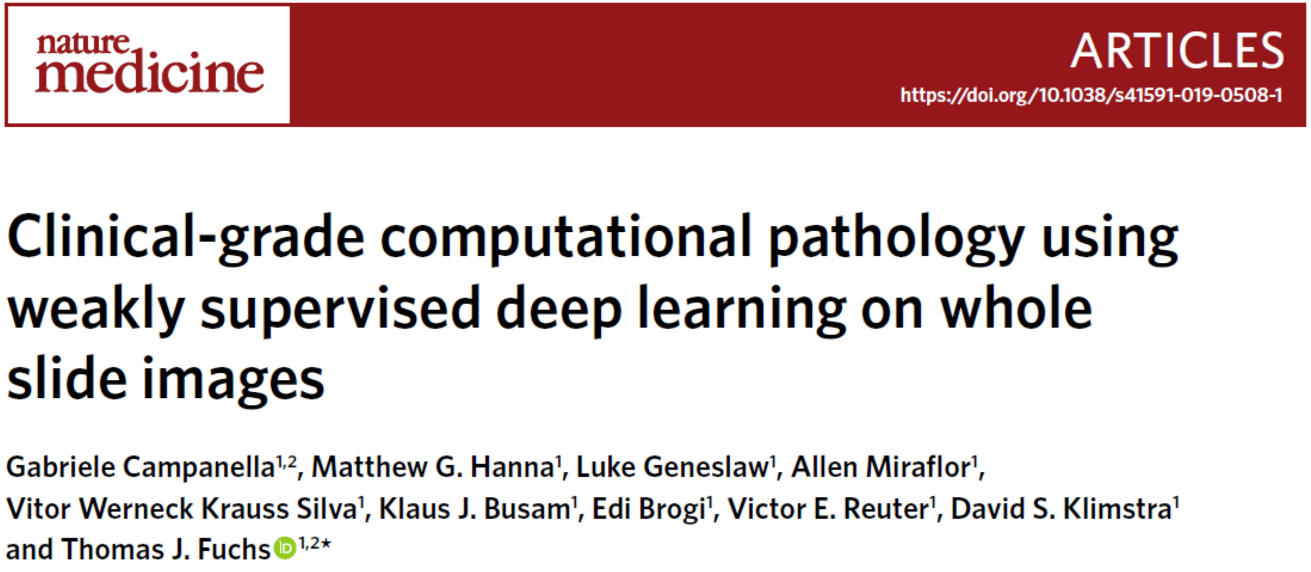
RNN



Recurrent networks



Paper 1



nature medicine

ARTICLES

https://doi.org/10.1038/s41591-019-0508-1

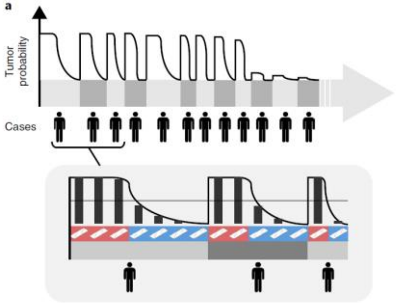
Clinical-grade computational pathology using weakly supervised deep learning on whole slide images

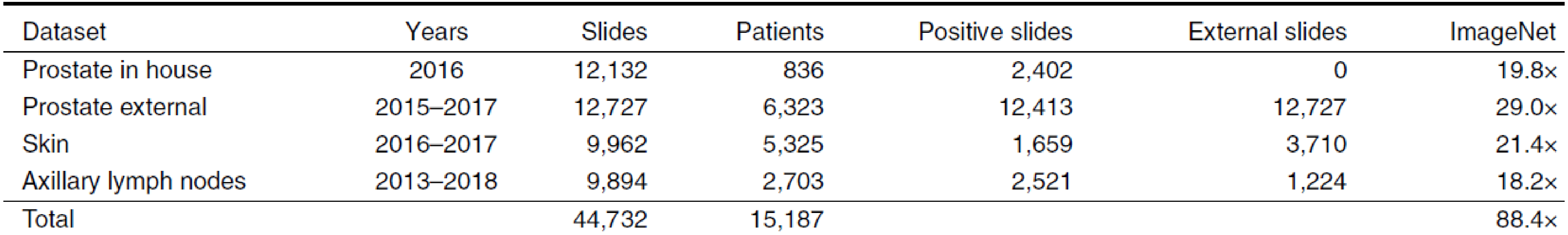
Gabriele Campanella[1,2], Matthew G. Hanna[1], Luke Geneslaw[1], Allen Miraflor[1],
Vitor Werneck Krauss Silva[1], Klaus J. Busam[1], Edi Brogi[1], Victor E. Reuter[1], David S. Klimstra[1]
and Thomas J. Fuchs[1,2]*

Task: classification positive/negative

- Prostatic carcinoma classification
- Skin basal cell carcinoma
- Brest cancer metastasis in axillary lymph nodes

Specifically addresses:



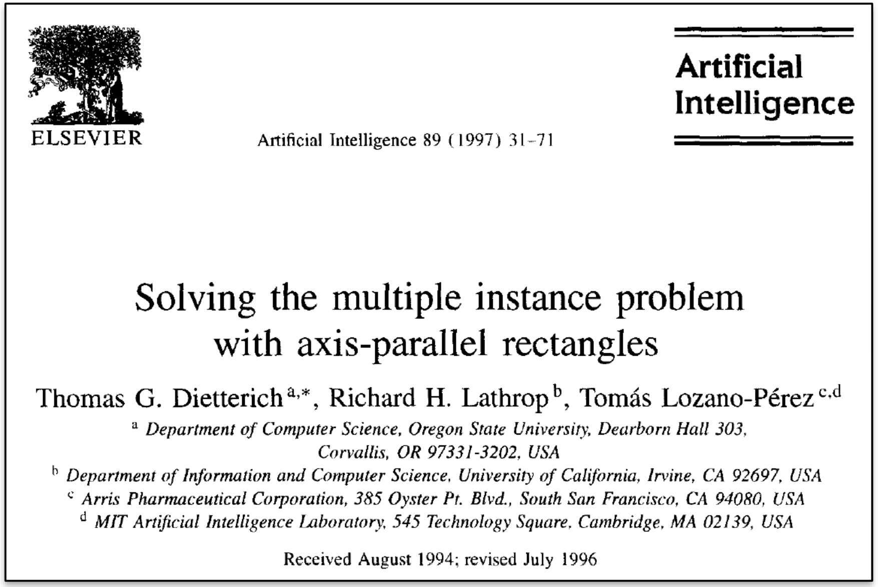


Dataset	Years	Slides	Patients	Positive slides	External slides	ImageNet
Prostate in house	2016	12,132	836	2,402	0	19.8×
Prostate external	2015–2017	12,727	6,323	12,413	12,727	29.0×
Skin	2016–2017	9,962	5,325	1,659	3,710	21.4×
Axillary lymph nodes	2013–2018	9,894	2,703	2,521	1,224	18.2×
Total		44,732	15,187			88.4×

Methods 1: Multiple Instance Learning (MIL)

MIL: multiple instance learning

Originates from this paper and was related to drug activity predictions



ELSEVIER

Artificial Intelligence 89 (1997) 31–71

Artificial Intelligence

Solving the multiple instance problem with axis-parallel rectangles

Thomas G. Dietterich[a,*], Richard H. Lathrop[b], Tomás Lozano-Pérez[c,d]

[a] *Department of Computer Science, Oregon State University, Dearborn Hall 303, Corvallis, OR 97331-3202, USA*

[b] *Department of Information and Computer Science, University of California, Irvine, CA 92697, USA*

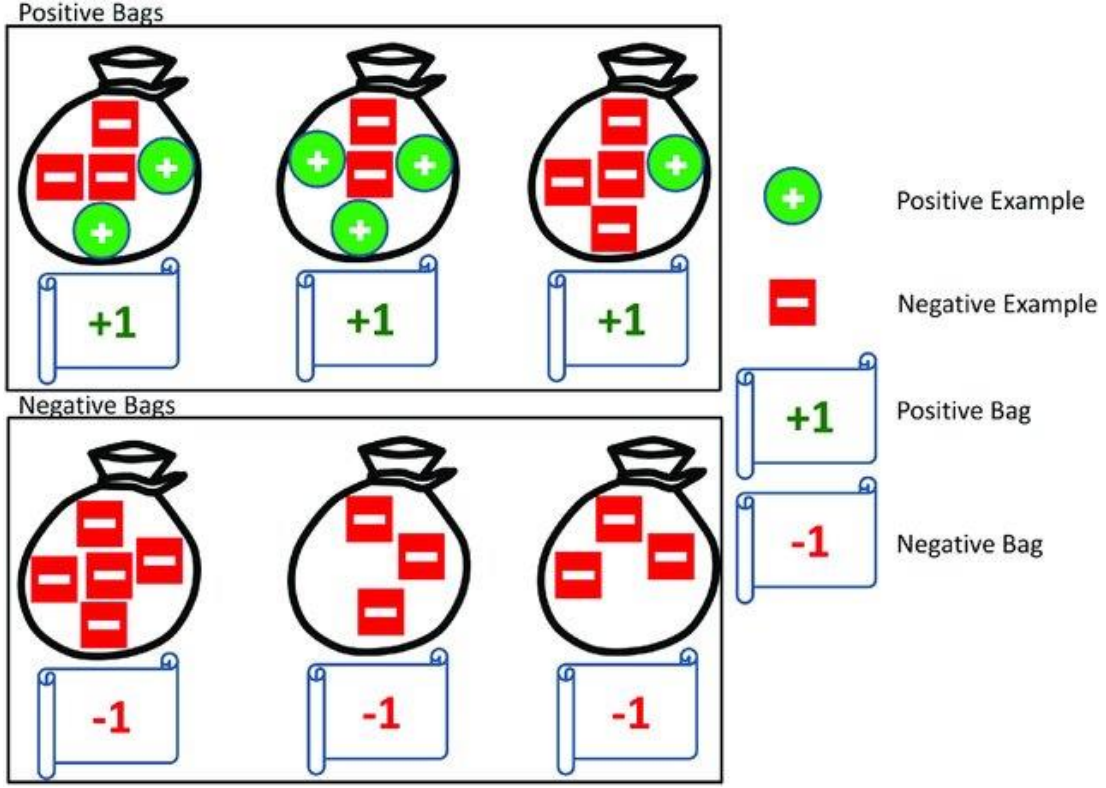
[c] *Arris Pharmaceutical Corporation, 385 Oyster Pt. Blvd., South San Francisco, CA 94080, USA*

[d] *MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139, USA*

Received August 1994; revised July 1996

Several algorithms are presented – need to dig into it ☺

The main idea:



DOI: 10.1371/journal.pcbi.1005465

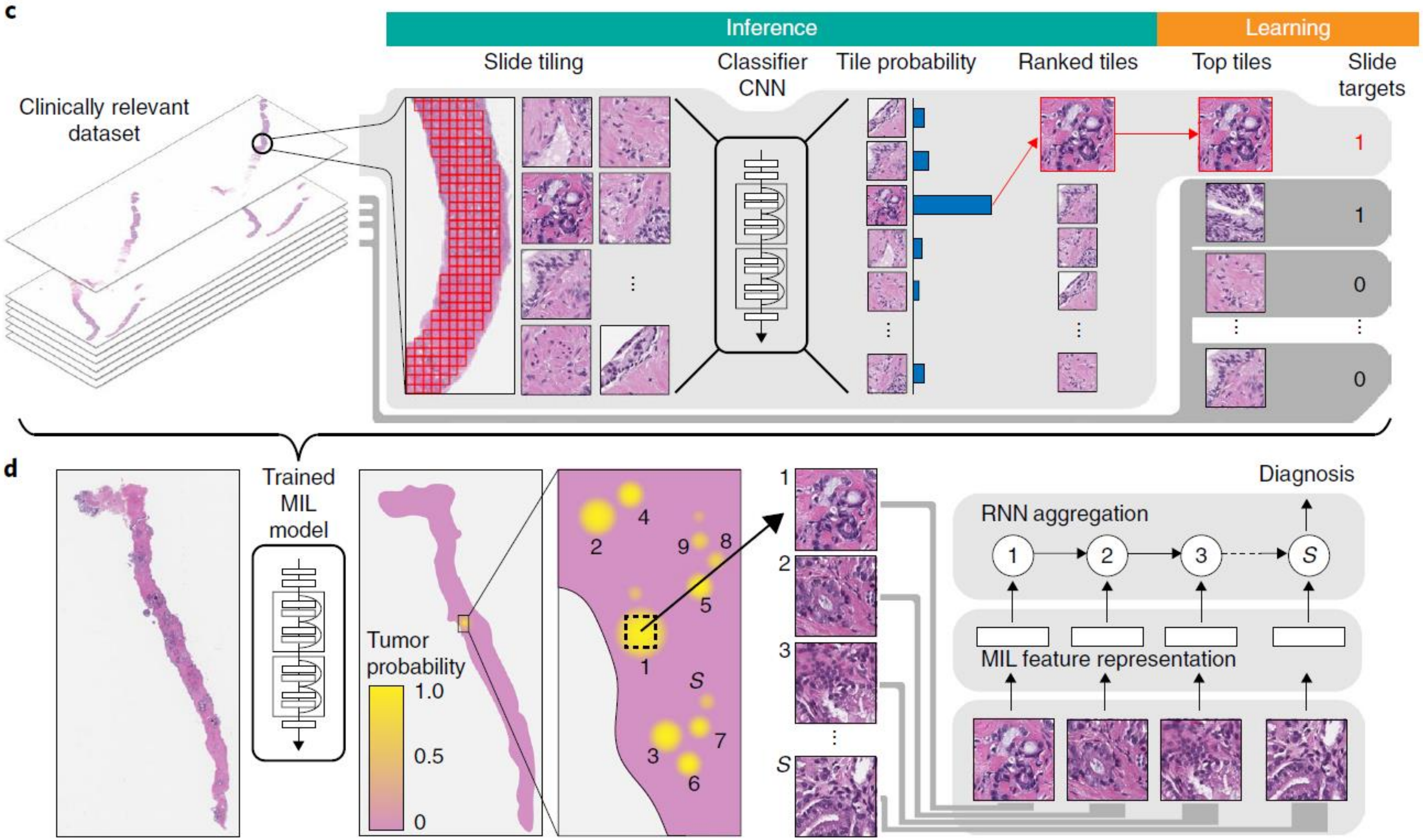
Methods 1: Training

1) **Tiling**: Otsu's method to discard b/g tiles. 3 magnitudes were investigated

2) Use **CNN** with min balanced error. *Tested:* ResNet34,18,101, AlexNet,VGG11BN, DenseNet201

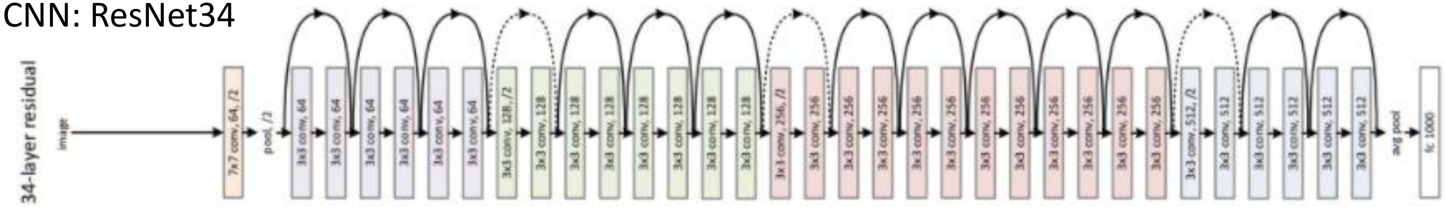
CNN is (1) trained, (2) used to refine classes of tiles (a.f.a.i.u ☺)

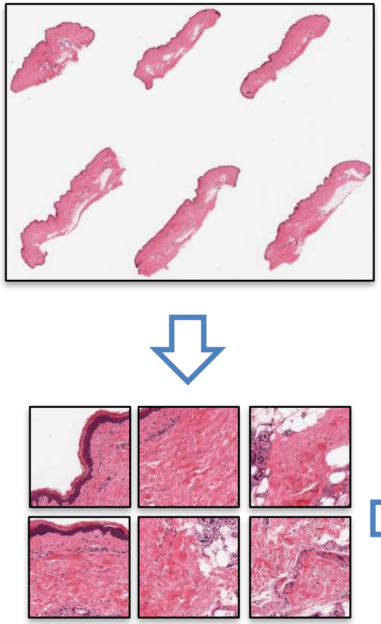
3) Use **RNN** for aggregate CNN feature (512) representation into a single class

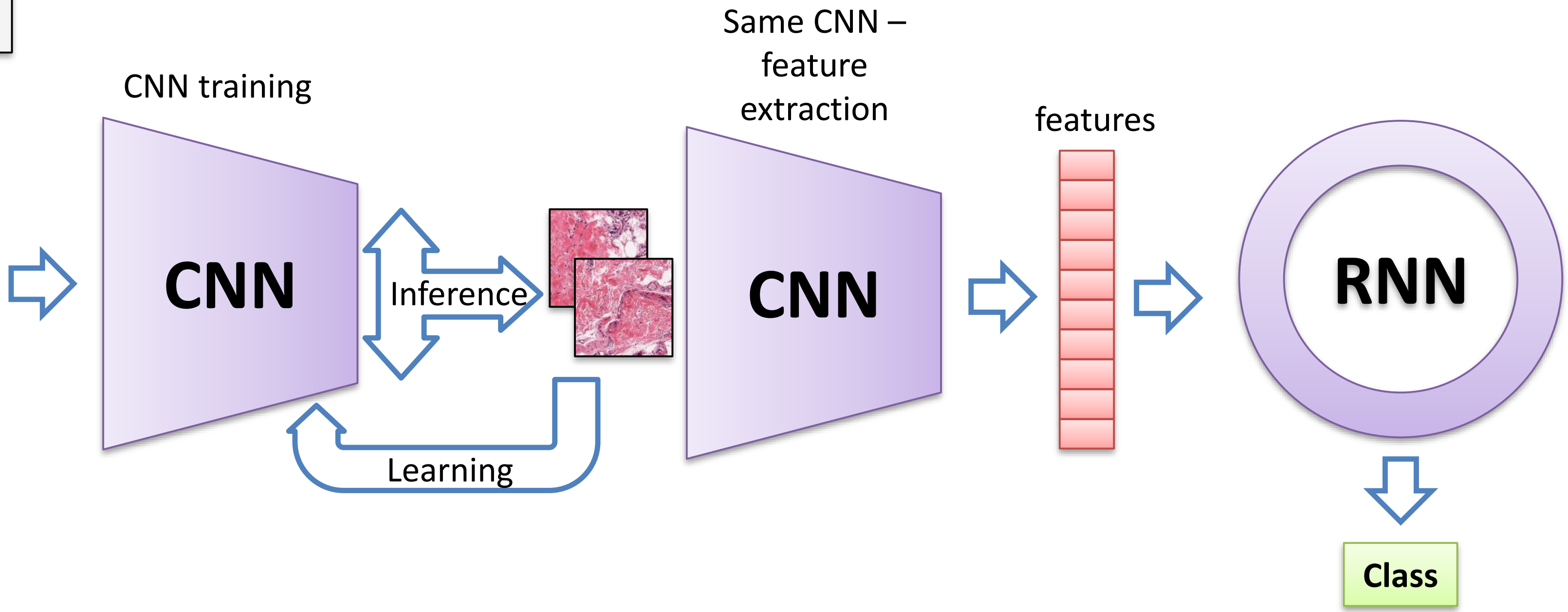


Methods 1: Simpler View on Training

CNN: ResNet34







Results 1: Accuracy

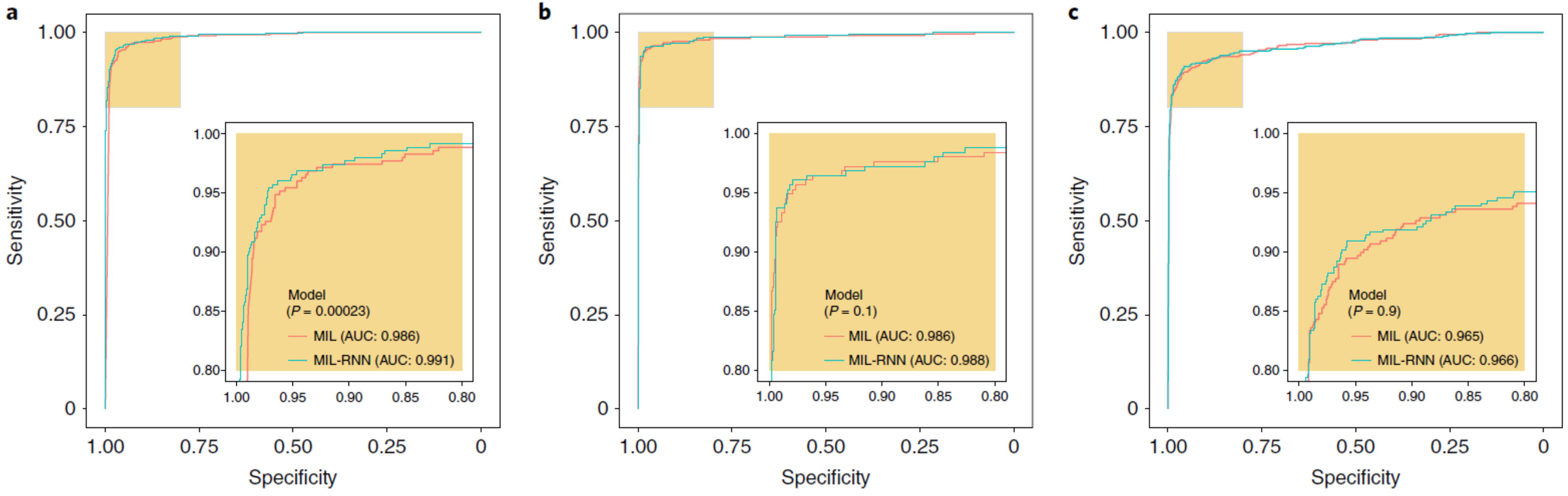


Fig. 3 | Weakly supervised models achieve high performance across all tissue types. The performances of the models trained at 20× magnification on the respective test datasets were measured in terms of AUC for each tumor type. **a**, For prostate cancer ($n = 1{,}784$) the MIL-RNN model significantly ($P < 0.001$) outperformed the model trained with MIL alone, resulting in an AUC of 0.991. **b**,**c**, The BCC model ($n = 1{,}575$) performed at 0.988 (**b**), while breast metastases detection ($n = 1{,}473$) achieved an AUC of 0.966 (**c**). For these latter datasets, adding an RNN did not significantly improve performance. Statistical significance was assessed using DeLong's test for two correlated ROC curves.

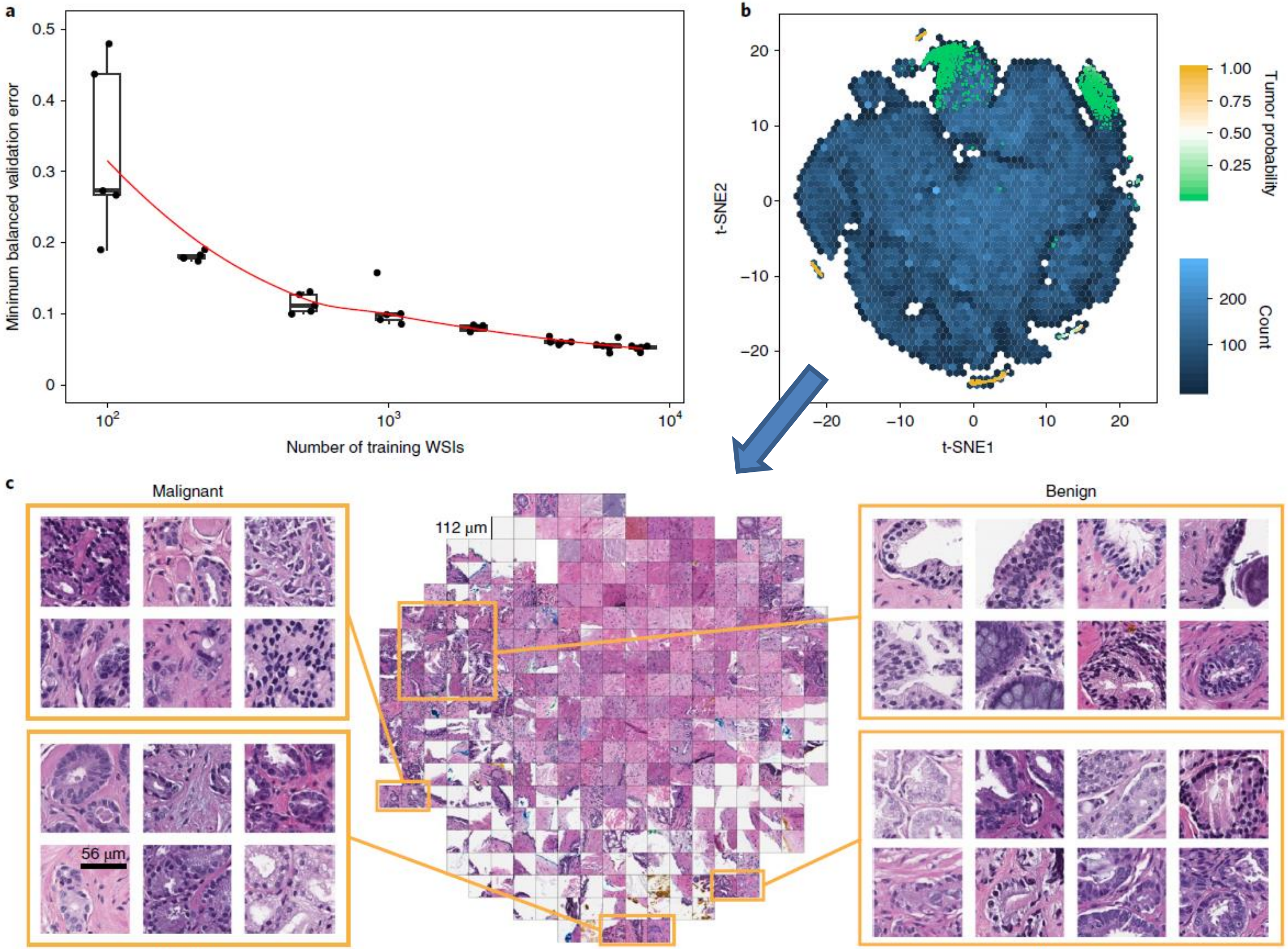
- ➢ MIL results can be used directly (not robust) or aggregated by logistic regression or RF. But RNN outperformed…

Results 1: Visualization of the feature space

a) The error depends strongly on the training set

b) CNN-based features (512) can be used for t-SNE representation.

c) Example representation with malignant, benign and suspicious tiles presented



Results 1:

Scanner & group effect

Compare to other datasets: dataset size effect ☺

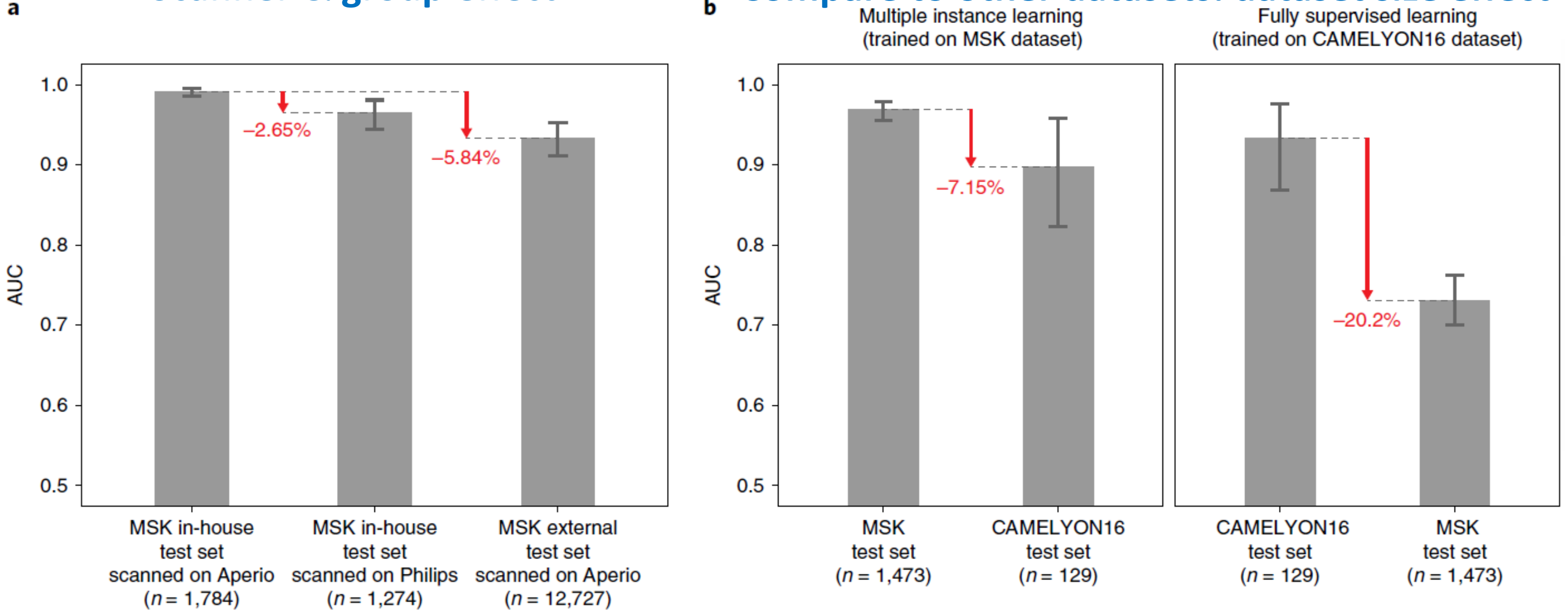


Fig. 5 | Weak supervision on large datasets leads to higher generalization performance than fully supervised learning on small curated datasets. The generalization performance of the proposed prostate and breast models were evaluated on different external test sets. **a**, Results of the prostate model trained with MIL on MSK in-house slides and tested on: (1) the in-house test set (*n* = 1,784) digitized on Leica Aperio AT2 scanners; (2) the in-house test set digitized on a Philips Ultra Fast Scanner (*n* = 1,274); and (3) external slides submitted to MSK for consultation (*n* = 12,727). Performance in terms of AUC decreased by 3 and 6% for the Philips scanner and external slides, respectively. **b**, Comparison of the proposed MIL approach with state-of-the-art fully supervised learning for breast metastasis detection in lymph nodes. Left, the model was trained on MSK data with our proposed method (MIL-RNN) and tested on the MSK breast data test set (*n* = 1,473) and on the test set of the CAMELYON16 challenge (*n* = 129), showing a decrease in AUC of 7%. Right, a fully supervised model was trained following ref. [18] on CAMELYON16 training data. While the resulting model would have won the CAMELYON16 challenge (*n* = 129), its performance drops by over 20% when tested on a larger test set representing real-world clinical cases (*n* = 1,473). Error bars represent 95% confidence intervals for the true AUC calculated by bootstrapping each test set.

CAMELYON16

CAMELYON17

Data link

One of the largest supervised dataset: breast cancer metastases in whole-slide images of histological lymph node sections.

Paper 2

nature COMMUNICATIONS

ARTICLE

Check for updates

https://doi.org/10.1038/s41467-020-17678-4 OPEN

A deep learning model to predict RNA-Seq expression of tumours from whole slide images

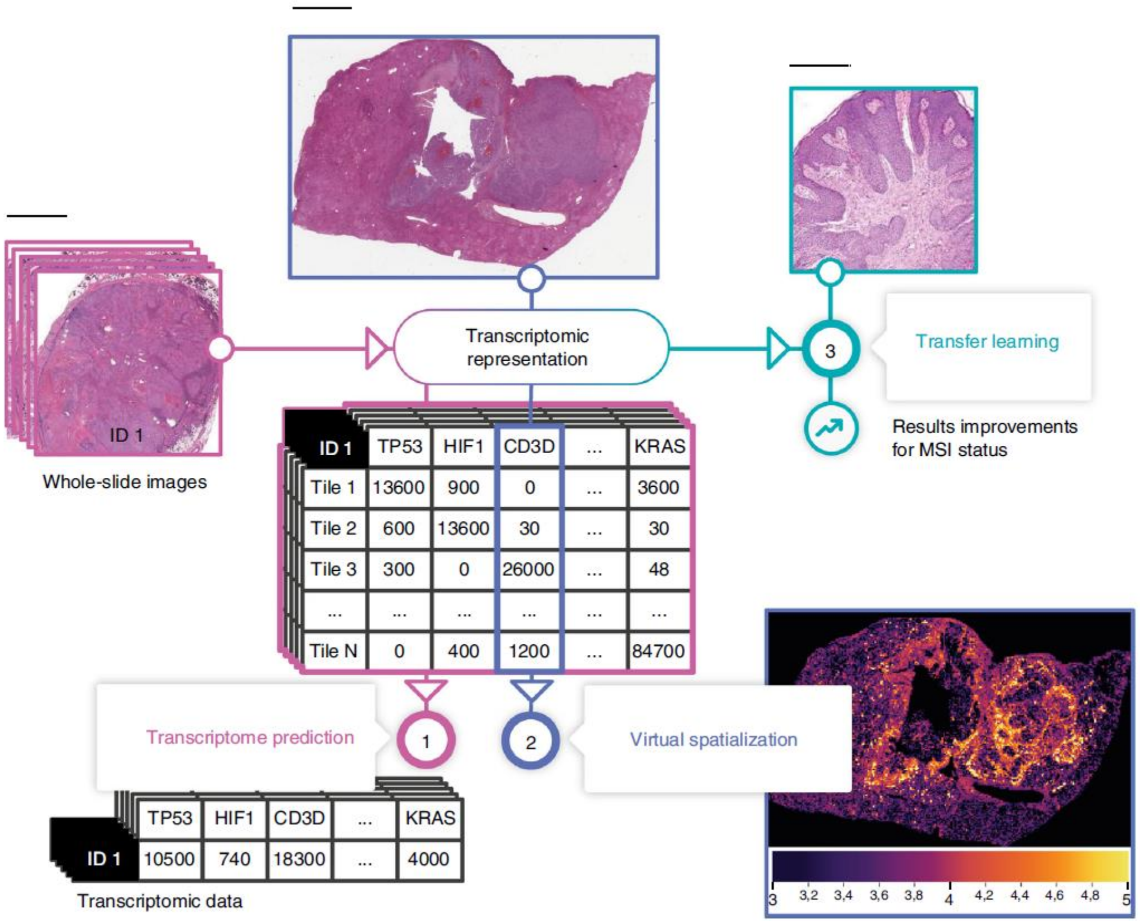
Benoît Schmauch[1]✉, Alberto Romagnoni[1,4], Elodie Pronier[1,4], Charlie Saillard[1], Pascale Maillé[2,3], Julien Calderaro[2,3], Aurélie Kamoun[1], Meriem Sefta[1], Sylvain Toldo[1], Mikhail Zaslavskiy[1], Thomas Clozel[1], Matahi Moarii[1], Pierre Courtiol[1,5] & Gilles Wainrib[1,5]✉

Task: multivariate regression

- ➢ Various cancers - input
- ➢ Gene expression - output

TCGA data:
8725 samples, 28 cancers,
30839 genes (med>0), normalized log FPKM-UQ
5-fold cross-validation

Methods 2: Graphical Abstract



ID 1	TP53	HIF1	CD3D	...	KRAS
Tile 1	13600	900	0	...	3600
Tile 2	600	13600	30	...	30
Tile 3	300	0	26000	...	48
...	...	...	...	...	...
Tile N	0	400	1200	...	84700

	TP53	HIF1	CD3D	...	KRAS
ID 1	10500	740	18300	...	4000

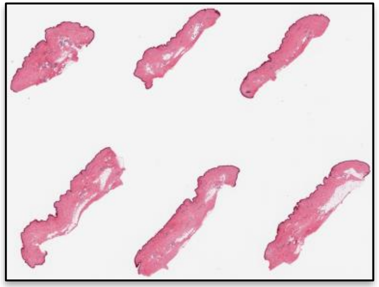
HE2RNA

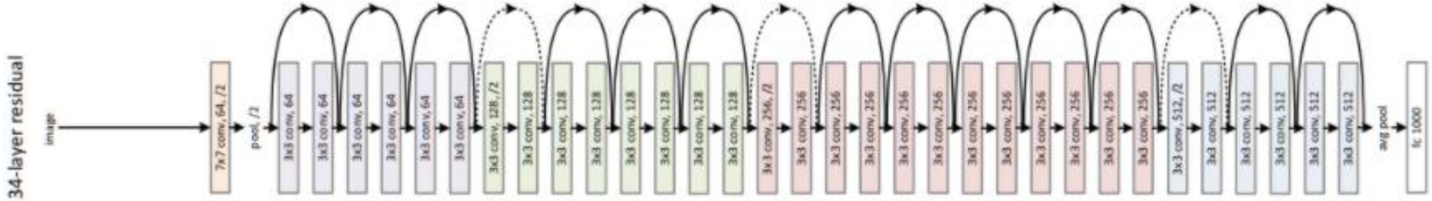
(1) Transctiptome prediction from images

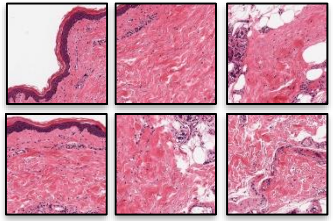
(2) Virtual spatialization of transcriptomic data (fro each gene over slide)

(3) Improving predictions by transfer leatning: e.g. microsatellite instability (MSI) from WSI

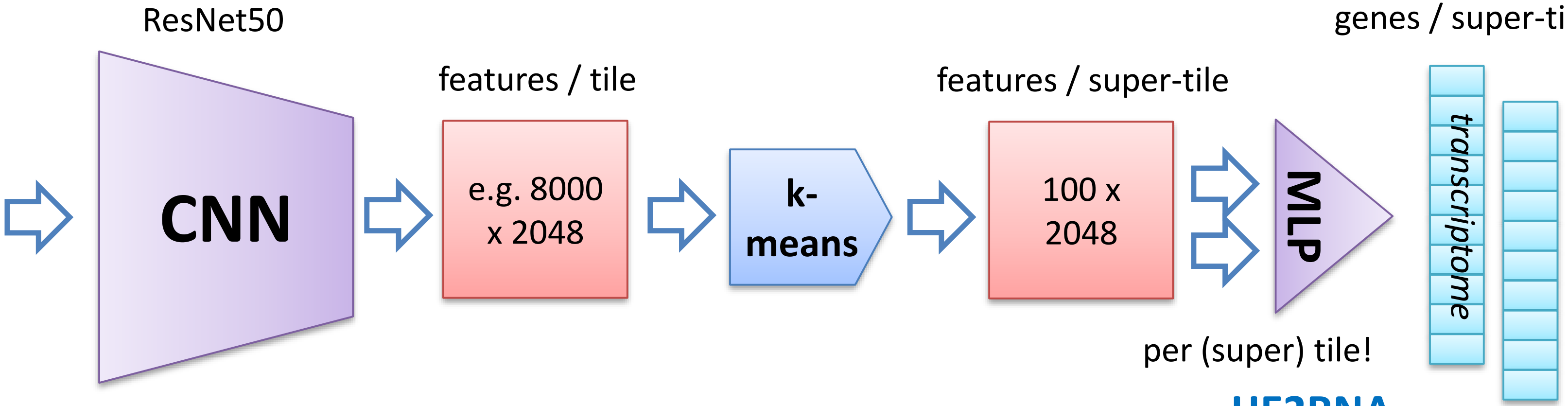
Methods 2: Training







Tiling: Otsu's method



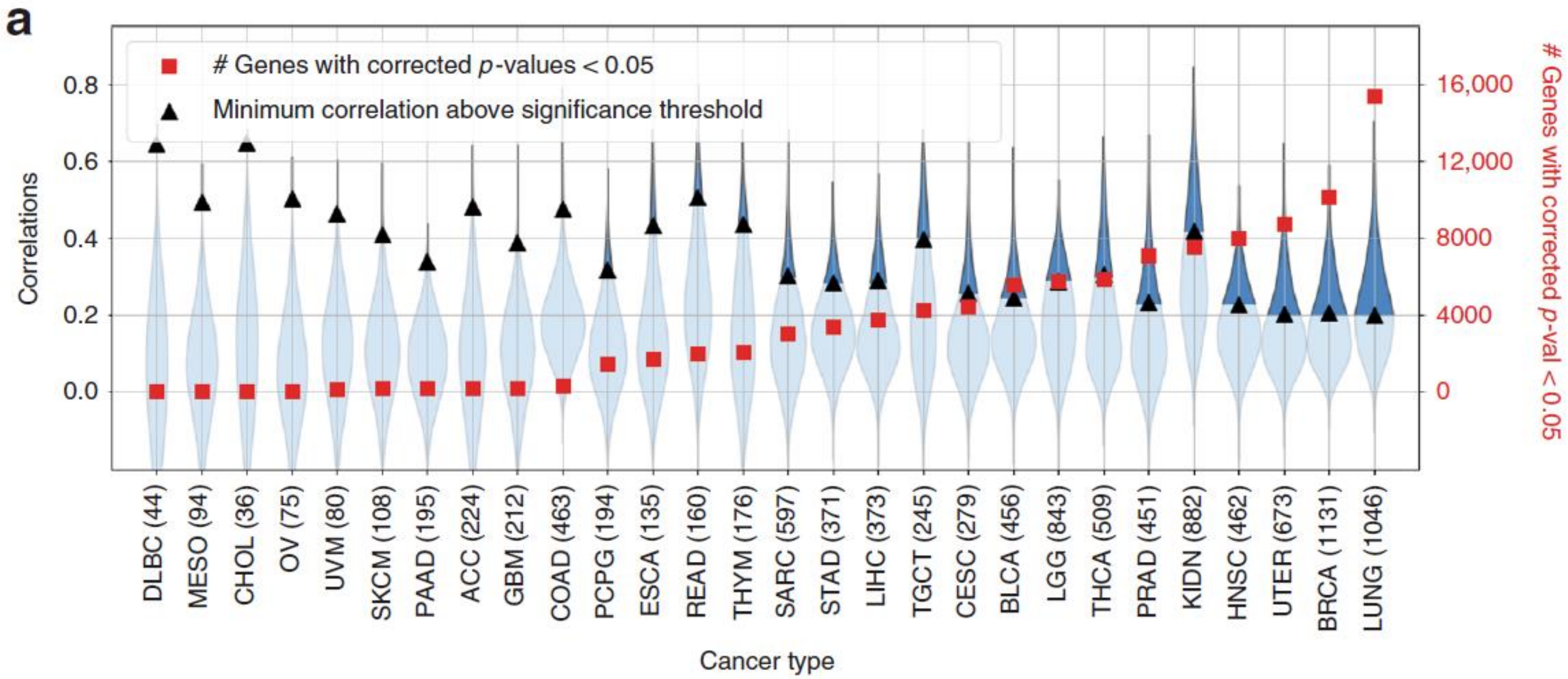
HE2RNA

1) **Tiling**: Otsu's method to discard b/g tiles.
2) Use a pre-trained **CNN**: ResNet50 to extract features
3) **Cluster** (k-means) to 100 super-tiles
4) Use a multi-layer perceptron (**MLP**) per (super-)slide

Aggregation: sampling k slides and averaging several the top predicted expression!

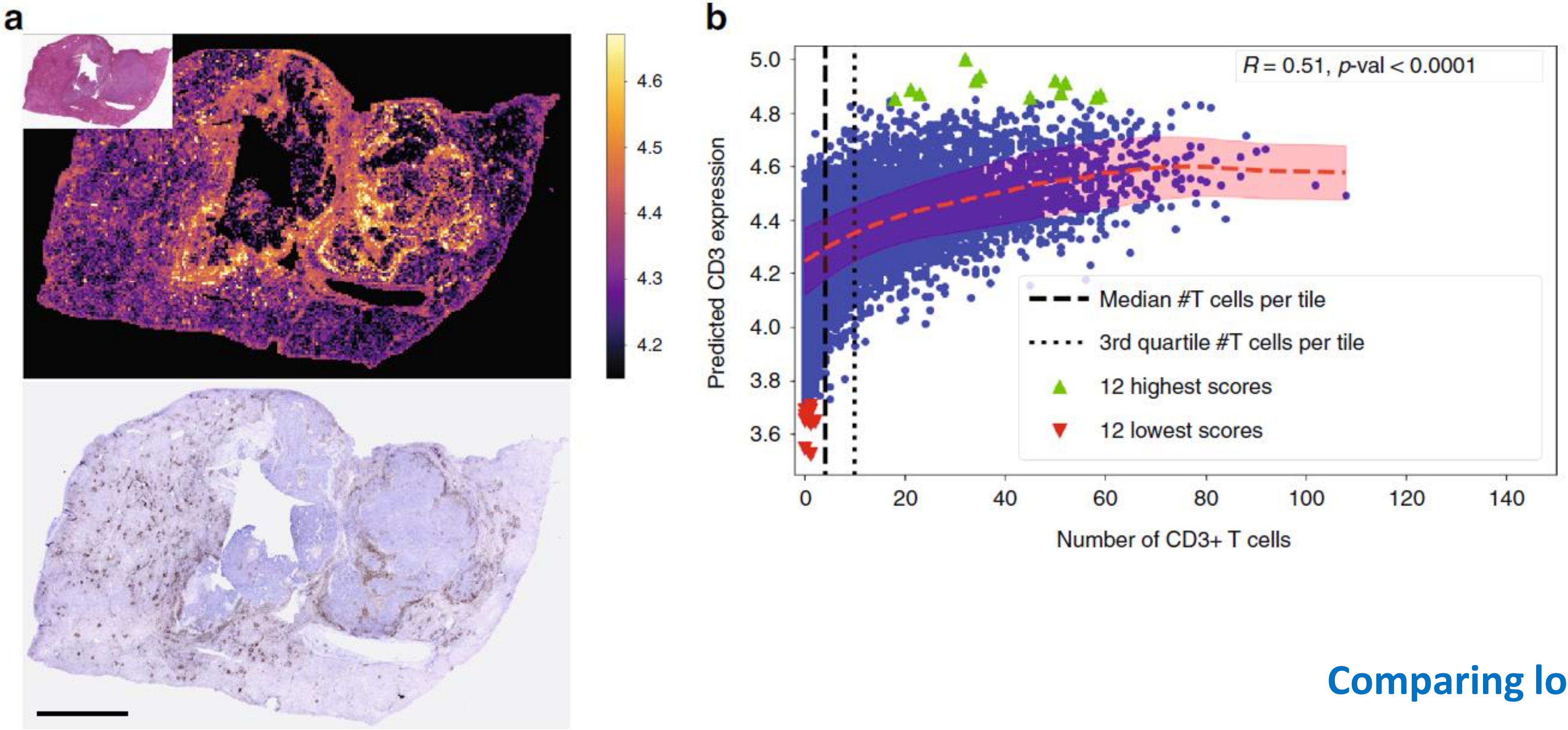
Results 2: Predicting Expression

A gene is predicted "correctly" if its **correlation** over samples r > 0 with adj.p-value < 0.05



How good is this measure?..

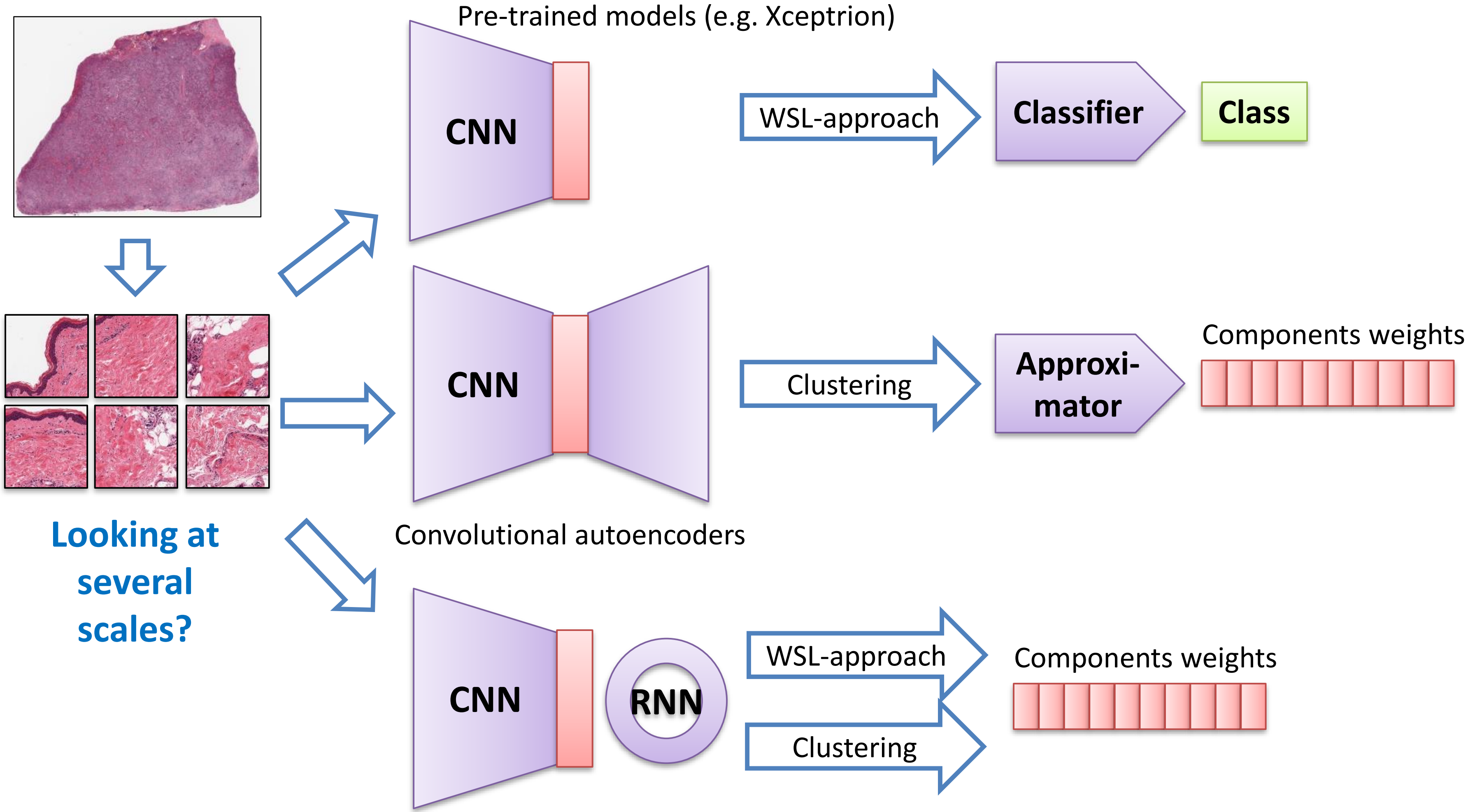
Results 2: Spatialization



Comparing log vs linear?

Fig. 4 Virtual spatialization of *CD3* and *CD20* expression, confirmed by immunohistochemistry. a Top left inset: H&E-stained slides were obtained from a LIHC patient. Main top image: The corresponding heatmap of the CD3-encoding genes expression predicted by our model. Main bottom image: CD3 immunohistochemistry (IHC) results obtained by washing out H&E stain and staining the same slide for IHC. **b** Pearson's coefficient ($R = 0.51$, p-value < 10^{-4}, two-tailed Student's t test) for the correlation between the CD3 expression predicted by our model and the percentage of CD3^{+} cells actually

Our Work ☺



Looking at several scales?