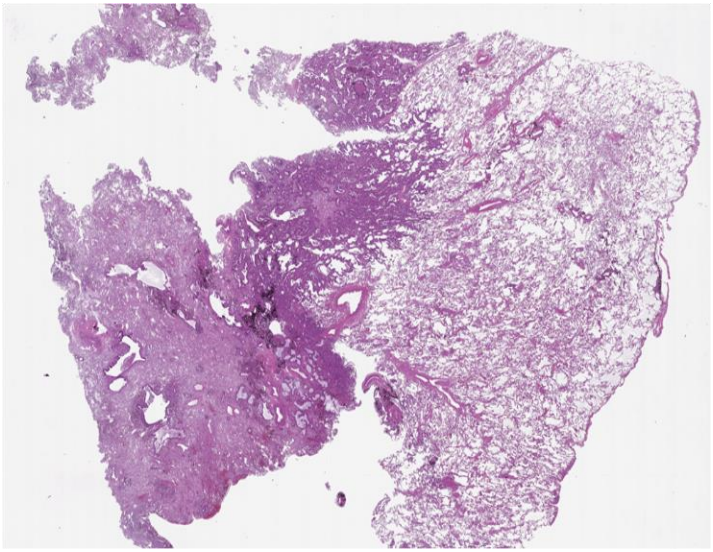
HEREDITARY

A full pipeline to analyse lung histopathology images

Lluis Borràs Ferrís

@SPIE, San Diego
19th February 2024



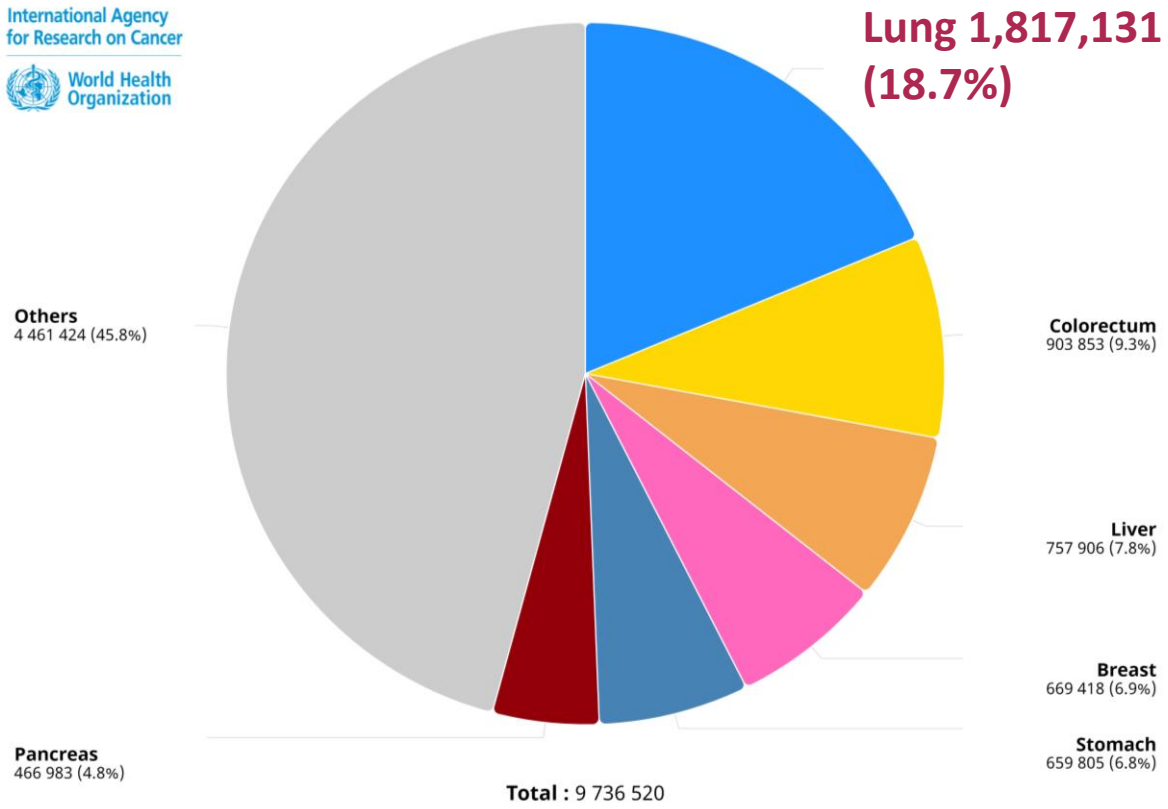
This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

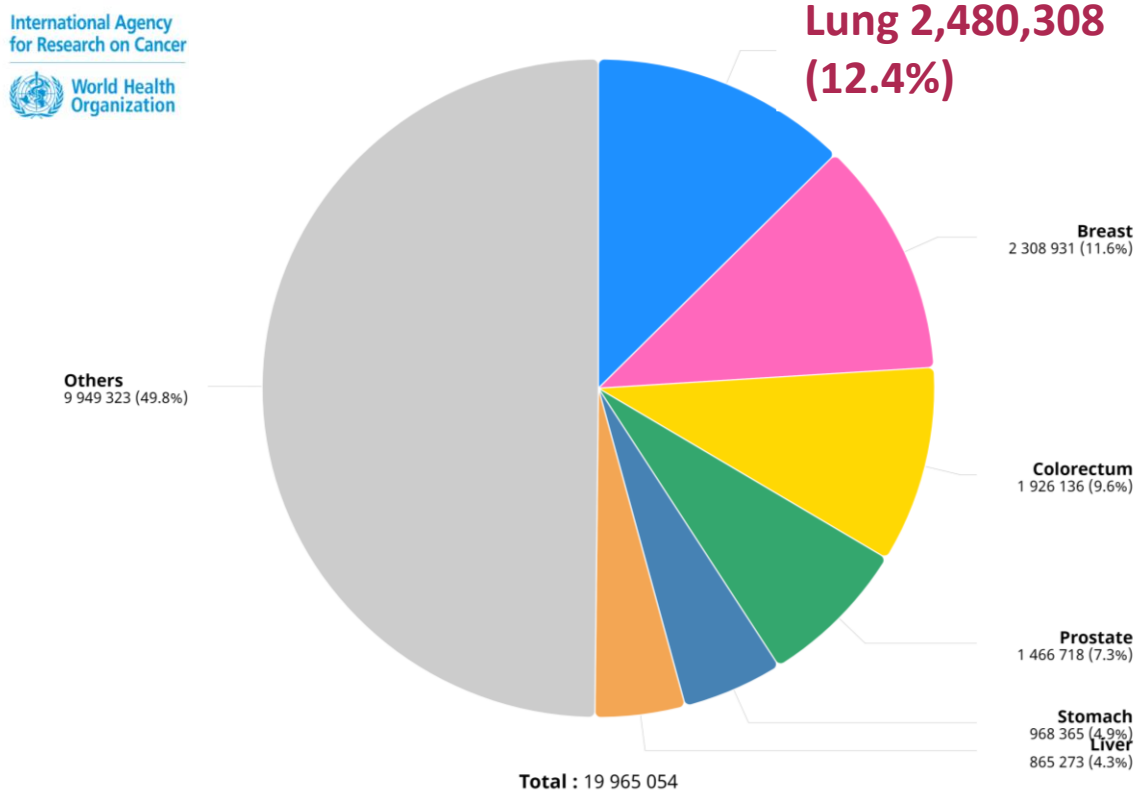
Background

- **Lung cancer** has **high mortality** and incidence

Absolute numbers, **Mortality**, Both sexes, in 2022.

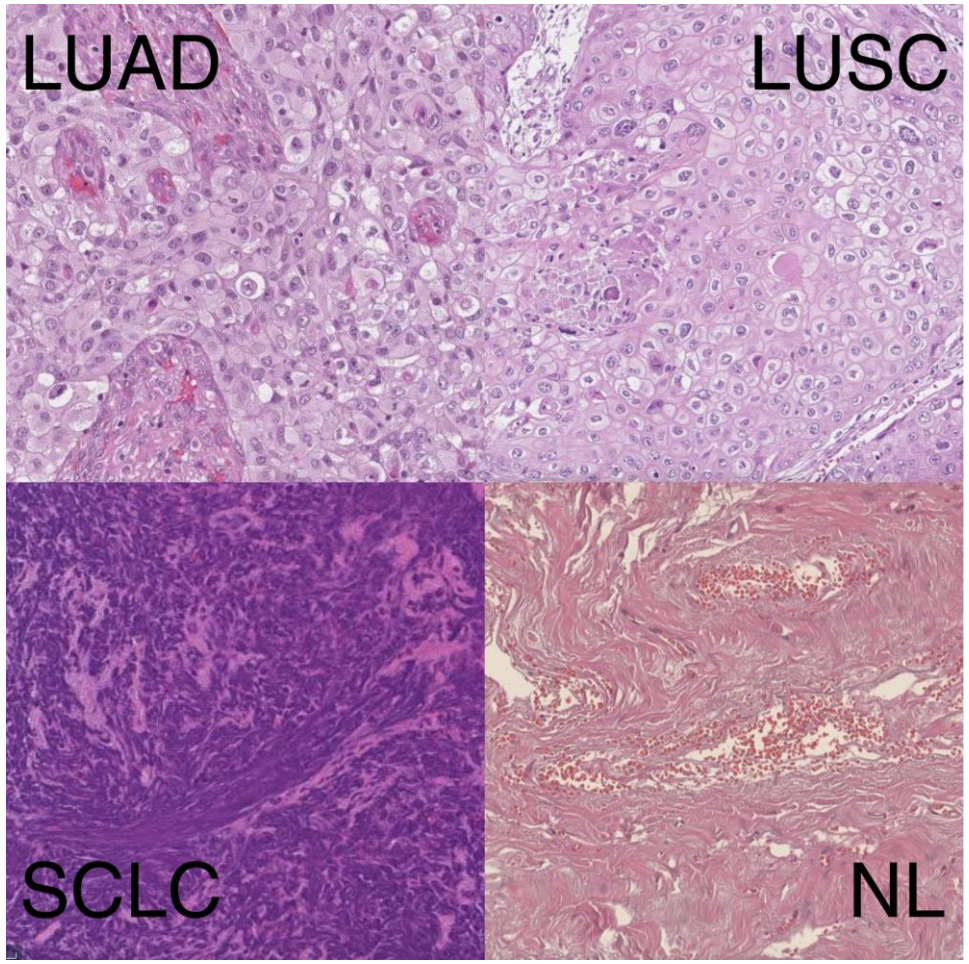


Absolute numbers, **Incidence**, Both sexes, in 2022.



Lung cancer subtypes

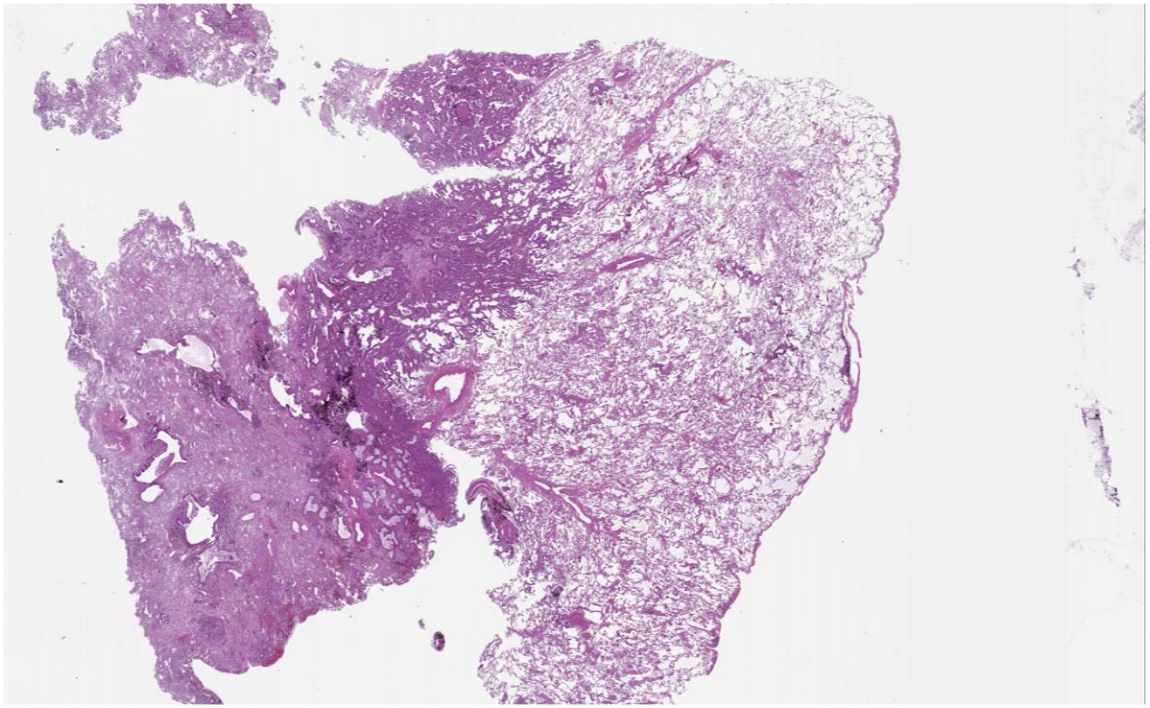
- Small-cell lung cancer (**SCLC**), 20%
- Non-small-cell lung cancer, 80%
  - Adenocarcinoma (**LUAD**), 50%
  - Squamous cell carcinoma (**LUSC**), 30%
  - Others, 20%
- Normal tissue (**NL**)
- Data are **heterogeneous**



Goldstraw, P., Ball, D., Jett, J.R., Le Chevalier, T., Lim, E., Nicholson, A.G., Shepherd, F.A., 2011. Non-small-cell lung cancer. The Lancet 378, 1727–1740.
Van Meerbeeck, J.P., Fennell, D.A., De Ruysscher, D.K., 2011. Small-cell lung cancer. The Lancet 378, 1741–1755.

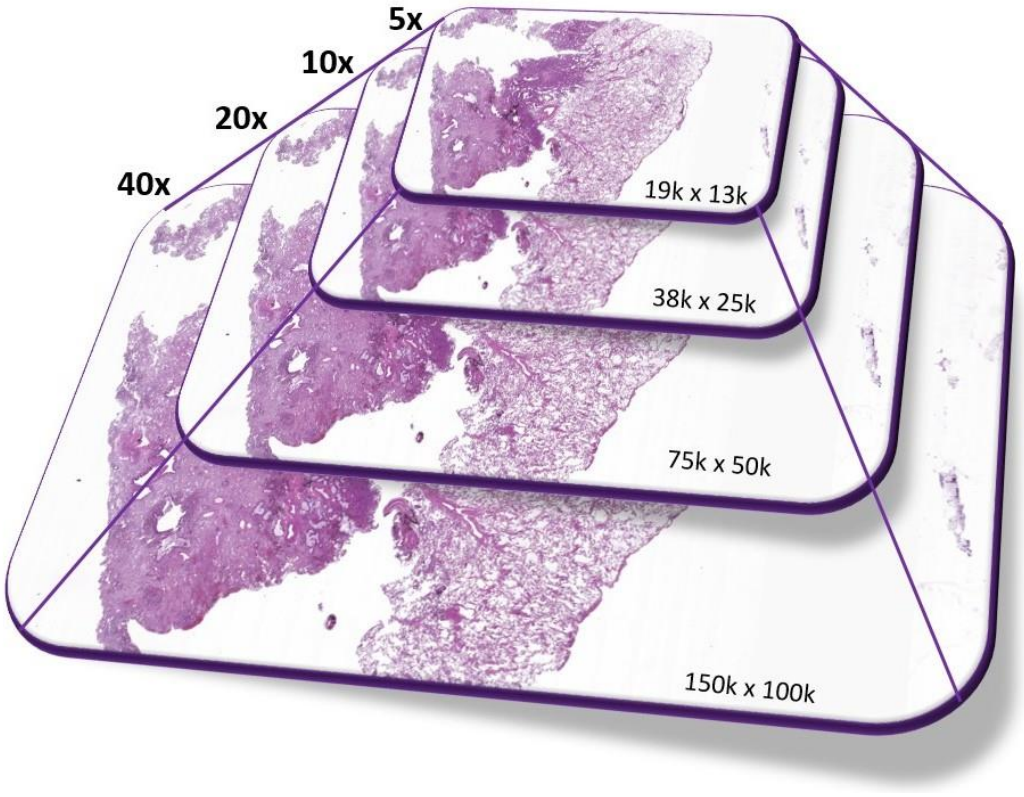
Histopathology

- **Microscopic images** of the tissues
- **Gold standard** to **diagnose** cancer
- **Identify** which **cancer subtype**
- **Experts** are needed



Digital pathology

- **Whole Slide Image** (WSI) is a **digitized slide** scanned at high-resolution
- **Multi-scale** format
- **Gigapixel** images
- **SOTA** based on **Deep Learning**



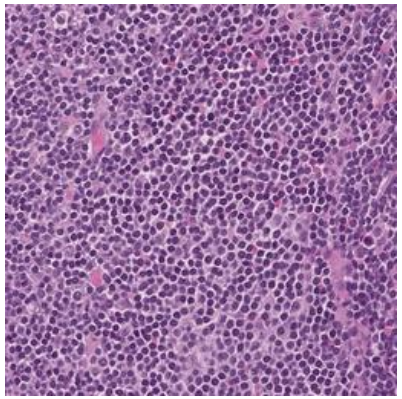
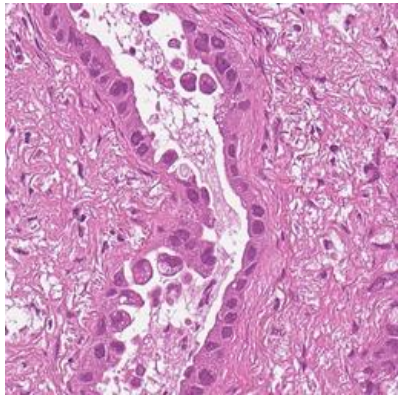
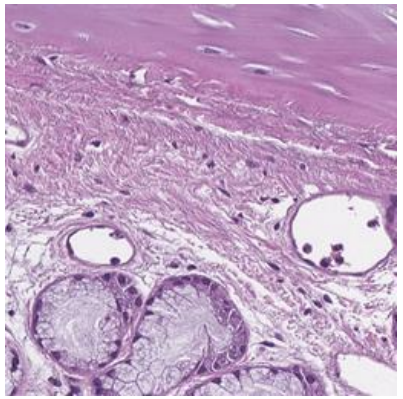
Marini, N., Otalora, S., Podareanu, D., van Rijthoven, M., van der Laak, J., Ciompi, F., Muller, H., Atzori, M., 2021a. Multi scale tools: a python library to exploit multi-scale whole slide images. Frontiers in Computer Science 3, 684521.

Data representation

- **Pre-training**
  - Transfer learning

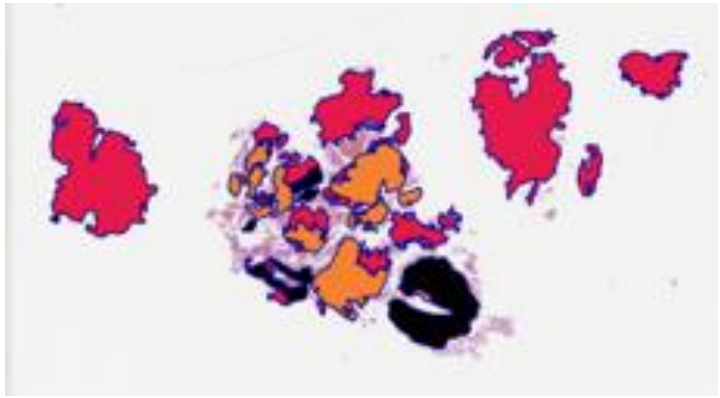
IMAGENET

  - Self-supervised learning

Annotations

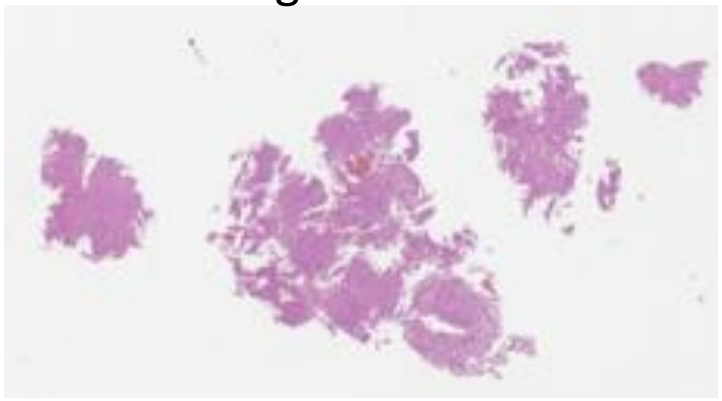
- **Training**
  - **Local Annotations**
    - Fully-supervised learning



- Label at **pixel-level**
- **Costly** to collect
- Highest performance

Label: Lung Adenocarcinoma

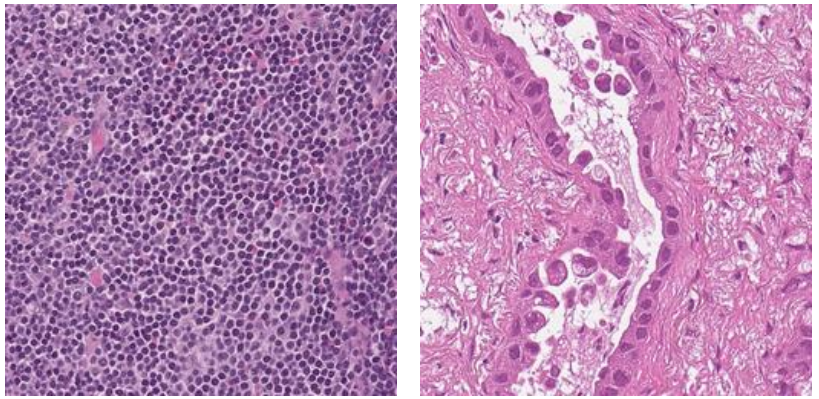
  - **Global Annotations**
    - Weakly-supervised learning



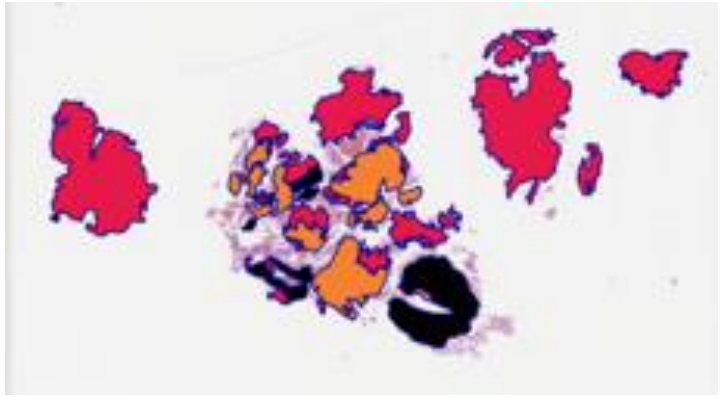
- Label at **WSI-level**
- **Cheaper** to collect

Motivation

- Representation from natural images may not be that effective for medical data



- Hard to collect large **locally-annotated datasets**



Objective

Develop a **fully automatic pipeline** for **WSI cancer subtype classification** using **self-supervision** and **weakly-supervised learning**

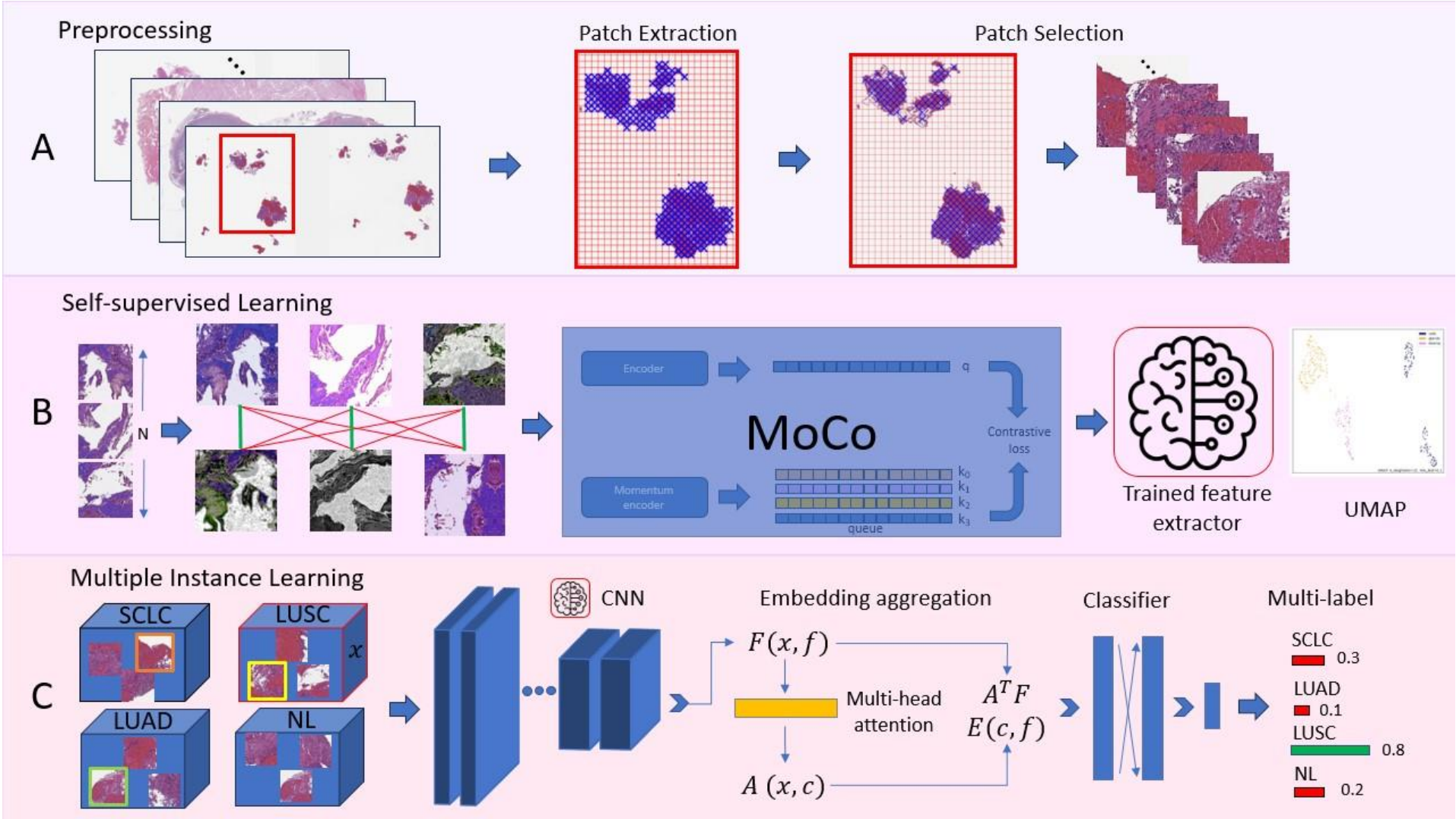
Datasets

- Data are **Multi-label**

Source	SCLC	LUAD	LUSC	Normal	Total labels	Total images
Training dataset from two different private datasets:						
AOEC	53	601	353	237	1,244	1,225
RUMC	0	297	205	499	1,001	1,001
Total	53	898	558	736	2,245	2,226
Testing private datastets:						
AOEC	17	16	9	14	46	46
RUMC	0	29	18	45	92	92
Total	17	45	27	59	138	138
Testing public dataset:						
TCGA	0	530	506	0	1,036	1,036

Albertina, B.,et al.. The cancer genome atlas lung adenocarcinoma collection (tcga-luad).
Kirk, S et al.. The cancer genome atlas lung squamous cell carcinoma collection (tcga-lusc).

Pipeline



He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9729–9738.

Chen, X., Fan, H., Girshick, R., He, K., 2020b. Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297.

Ilse, M., Tomczak, J., Welling, M., 2018. Attention-based deep multiple instance learning, in: International conference on machine learning, PMLR. pp. 2127–2136.

Results

- Test on the private AOEC and RUMC datasets

Pre-training	AUC SCLC	AUC LUAD	AUC LUSC	AUC Normal	micro-AUC	weighted f1-score
Train on AOEC and RUMC:						
Self-supervised	0.8825 ± 0.0712	0.7457 ± 0.0267	0.8428 ± 0.0171	0.8468 ± 0.0	**0.8558 ± 0.0051**	**0.6537 ± 0.0237**

Results

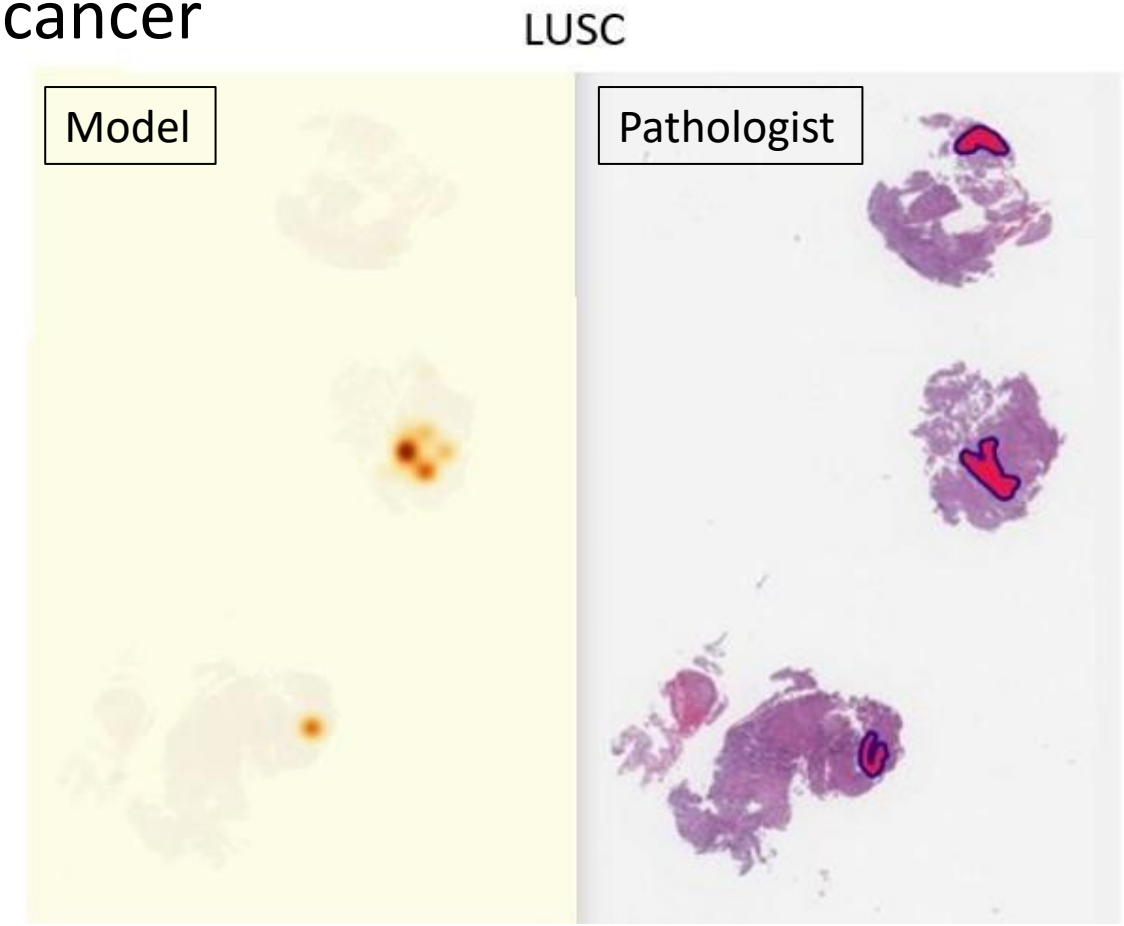
- Test on the public TCGA dataset
  - Capability of **generalize** on public data

Pre-training	micro-AUC	weighted f1-score
Test on TCGA:		
Train on AOEC and RUMC:		
Self-supervised	**0.9433 ± 0.0198**	**0.7726 ± 0.0438**

Albertina, B.,et al.. The cancer genome atlas lung adenocarcinoma collection (tcga-luad).
Kirk, S et al.. The cancer genome atlas lung squamous cell carcinoma collection (tcga-lusc).

Results

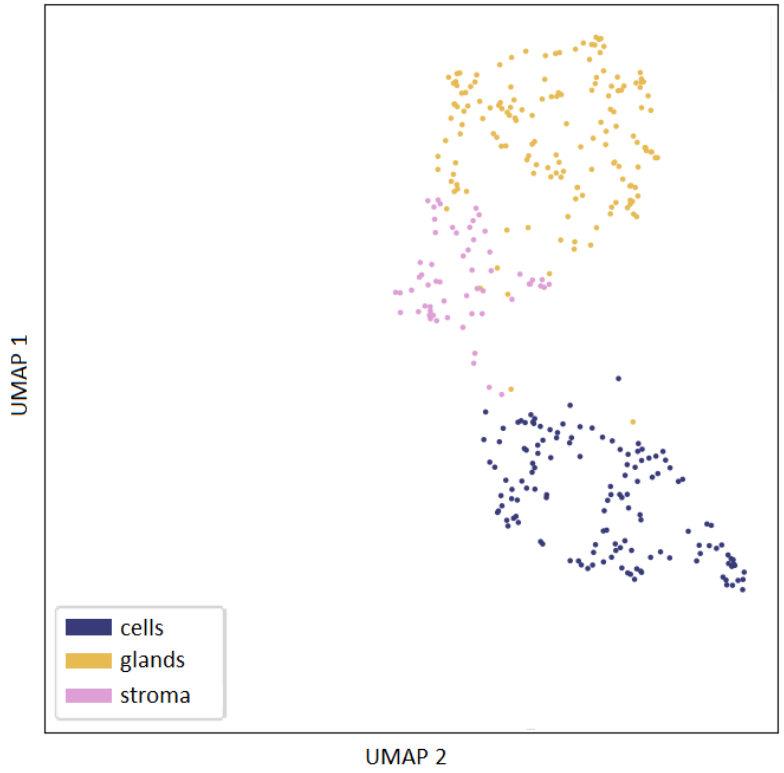
- **Heatmaps** can **help** pathologists **localizing** and **diagnosing** lung cancer



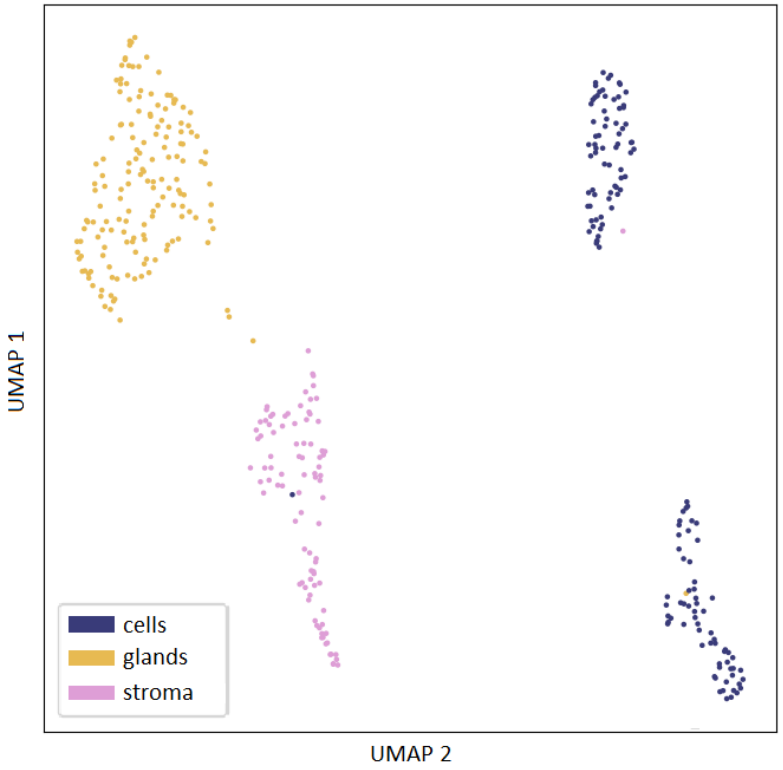
Discussion

- Differences on **Data representation**

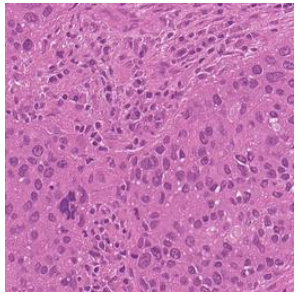
UMAP ImageNet



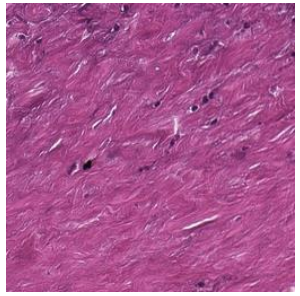
UMAP Self-supervised



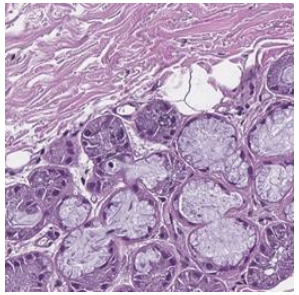
cells



stroma



glands



McInnes, L., Healy, J., Melville, J., 2018. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 .

Conclusions

- A **fully automatic pipeline** to classify **4-class** lung cancer WSIs
- **Pre-training** with **self-supervision** for a better **data representation**
- **Weakly-supervised learning** enables **training** the model using only **global annotations**
- The model generates **accurate predictions** on the **TCGA** dataset showing its **generalization** capabilities
- **Heatmaps** a potential **tool** to **help pathologists** in the **localization** and **diagnosis** of lung cancer on WSI

HEREDITARY

Twitter/X: @Hereditary_EU
YouTube: @Hereditary_EU
LinkedIn: @HEREDITARY

medGIFT

https://medgift.hevs.ch/wordpress/

Lluis Borràs Ferrís, PhD Student

Email: lluis.borrasferris@hevs.ch
LinkedIn: @lluis-borras-ferris

Supervisor: Henning Müller

@SPIE, San Diego 19th February 2024

UNIVERSITÉ DE GENÈVE

Hes·so VALAIS WALLIS

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union