Machine learning for integrating imaging, anatomo-pathological and omics data to predict outcome of patients treated with immunotherapy

Nicolas Captier

Orsay inter-unit seminar

10/05/2023

Outline

1. **Introduction:** predict the outcome of metastatic non-small cell lung cancer patients treated by immunotherapy

2. **Unimodal approaches:** extract relevant features from each modality

3. **Multimodal approach:** handle missing modalities with late fusion strategies

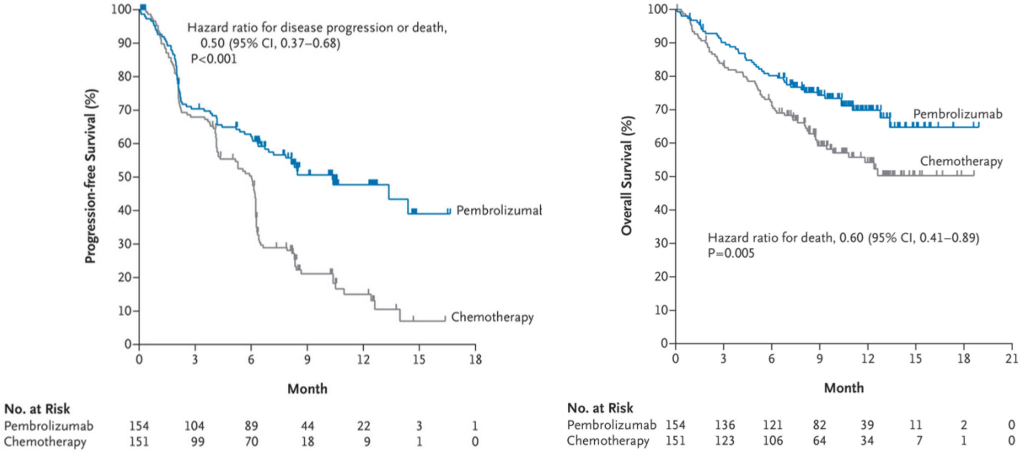
Lung cancer in France

- Lung cancer is the first cause of cancer-related death in France

- The most frequent type, non-small cell lung cancer (NSCLC) is diagnosed at a metastatic stage in about 70% of the patients

Lung cancer in France

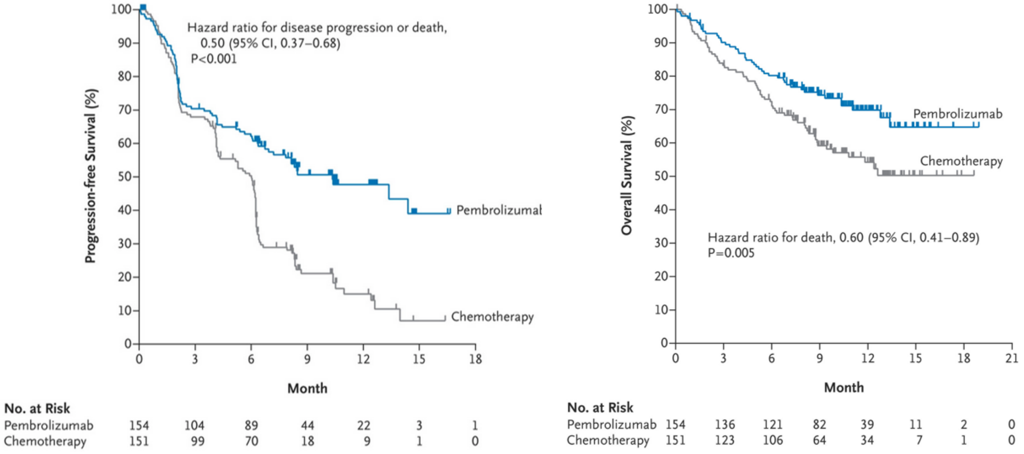
- Lung cancer is the first cause of cancer-related death in France
- The most frequent type, non-small cell lung cancer (NSCLC) is diagnosed at a metastatic stage in about 70% of the patients
- Checkpoint inhibitor-based immunotherapies have transformed the standard of care for NSCLC patients.



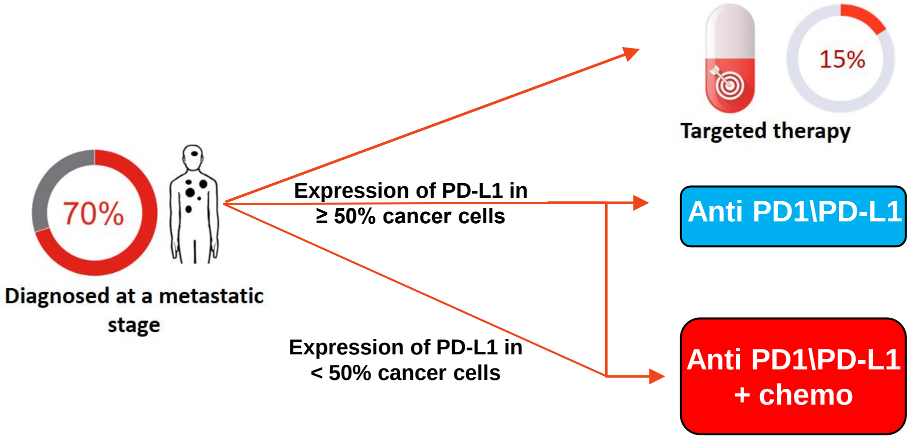
KEYNOTE-024 trial, chemotherapy vs pembrolizumab for ≥ 50% PD-L1 advanced NSCLC patients (Reck et al. 2016)

Lung cancer in France

- Lung cancer is the first cause of cancer-related death in France
- The most frequent type, non-small cell lung cancer (NSCLC) is diagnosed at a metastatic stage in about 70% of the patients
- Checkpoint inhibitor-based immunotherapies have transformed the standard of care for NSCLC patients.



KEYNOTE-024 trial, chemotherapy vs pembrolizumab for ≥ 50% PD-L1 advanced NSCLC patients (Reck et al. 2016)



Targeted therapy

Diagnosed at a metastatic stage

Expression of PD-L1 in ≥ 50% cancer cells

Anti PD1\PD-L1

Expression of PD-L1 in < 50% cancer cells

Anti PD1\PD-L1 + chemo

Standard-of-care for advanced NSCLC in Europe

Standard-of-care still needs to be optimized

- **Only 45-50% of patients present an objective response** to anti PD-1/L1 immunotherapy

- The duration of response remains highly variable and **only 40% of the patients are alive at 2 years**

Standard-of-care still needs to be optimized

- **Only 45-50% of patients present an objective response** to anti PD-1/L1 immunotherapy

- The duration of response remains highly variable and **only 40% of the patients are alive at 2 years**

- Some straightforward predictors have been proposed for the response to immunotherapy (e.g TMB, PD-L1 expression, lymphocytic infiltrates…)

- **They are highly suboptimal** and not always reliable.

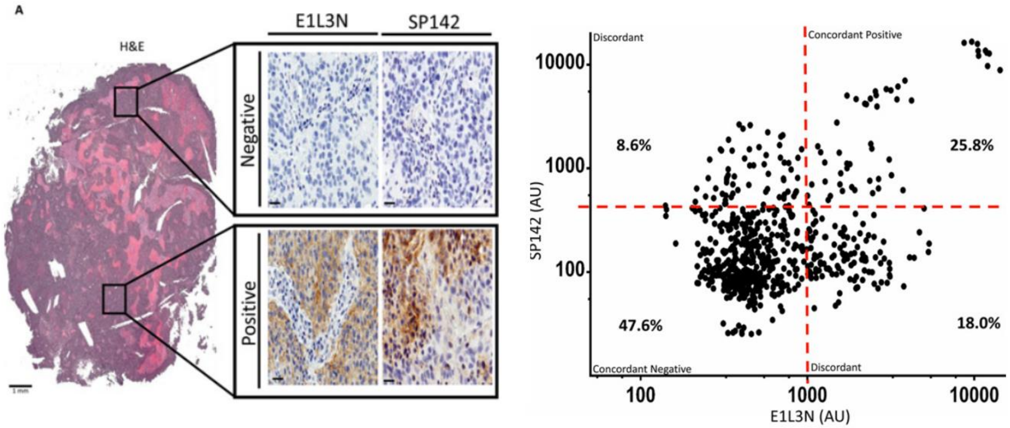
Standard-of-care still needs to be optimized

- **Only 45-50% of patients present an objective response** to anti PD-1/L1 immunotherapy

- The duration of response remains highly variable and **only 40% of the patients are alive at 2 years**

- Some straightforward predictors have been proposed for the response to immunotherapy (e.g TMB, PD-L1 expression, lymphocytic infiltrates…)

- **They are highly suboptimal** and not always reliable.



Heterogeneity of expression of PD-L1 and variation with antibody use (Maclaughlin et al. 2016).

Supervised Machine Learning for biomarker discovery

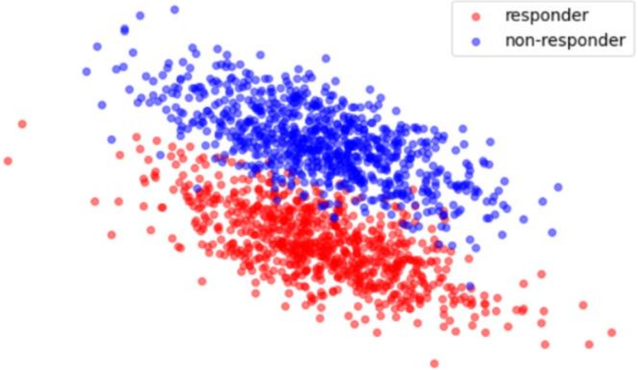
- Build a function f which takes as input baseline variables and predicts the patient's outcome

- We want f to minimize the error between the predicted outcome and the true observed one for all the patients of our cohort

Supervised Machine Learning for biomarker discovery

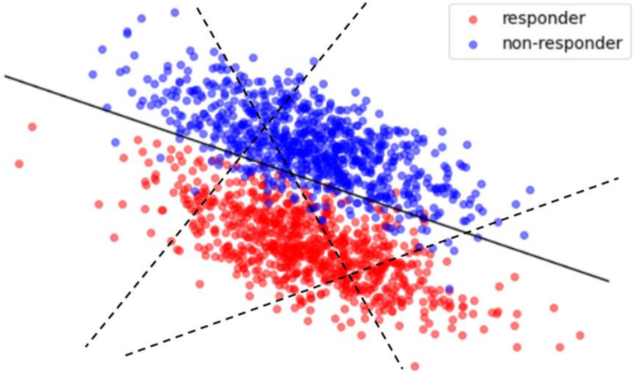
- Build a function f which takes as input baseline variables and predicts the patient's outcome

- We want f to minimize the error between the predicted outcome and the true observed one for all the patients of our cohort



Linear classifier

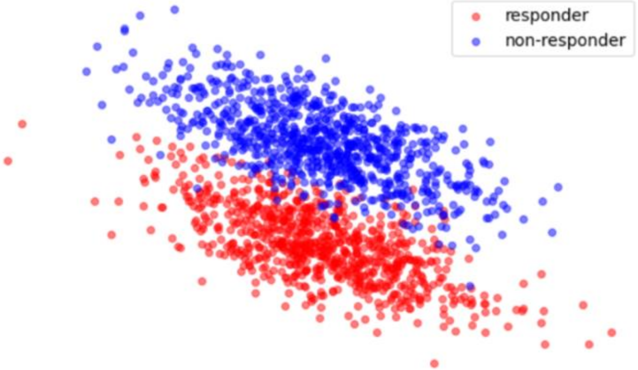
$$f(x_1, x_2) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = \hat{y}$$



Supervised Machine Learning for biomarker discovery

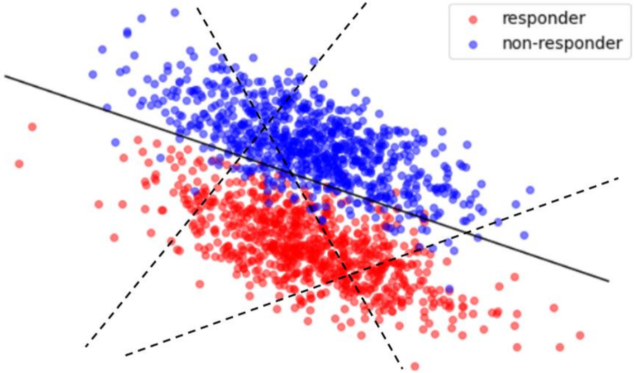
- Build a function f which takes as input baseline variables and predicts the patient's outcome

- We want f to minimize the error between the predicted outcome and the true observed one for all the patients of our cohort



Linear classifier

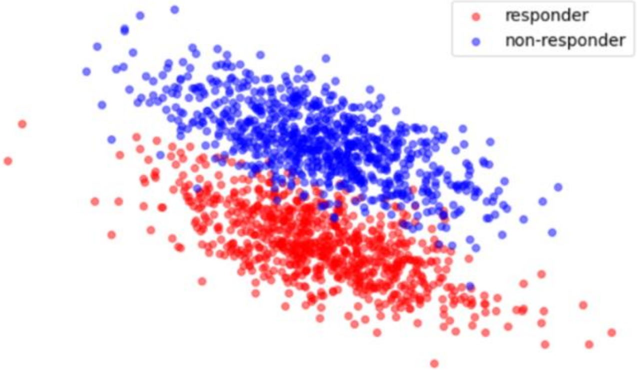
$$f(x_1, x_2) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = \hat{y}$$



❶ Learn how to optimally combine known biomarkers (e.g $f(PDL1, TMB)$)

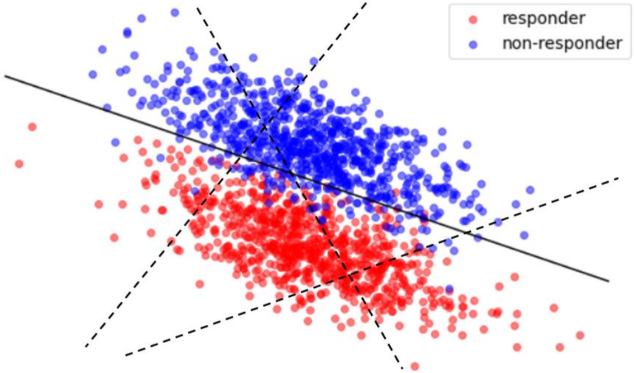
Supervised Machine Learning for biomarker discovery

- Build a function f which takes as input baseline variables and predicts the patient's outcome
- We want f to minimize the error between the predicted outcome and the true observed one for all the patients of our cohort



Linear classifier

$$f(x_1, x_2) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = \hat{y}$$

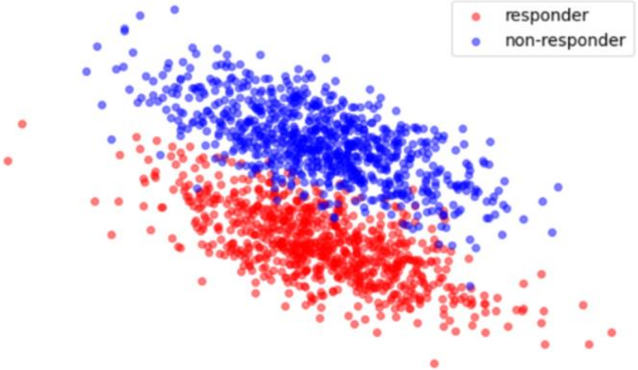


❶ Learn how to optimally combine known biomarkers (e.g $f(PDL1, TMB)$)

❷ Discover and integrate new informative features

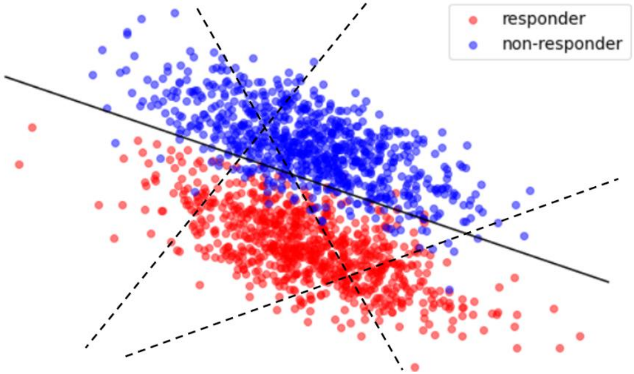
Supervised Machine Learning for biomarker discovery

- Build a function f which takes as input baseline variables and predicts the patient’s outcome
- We want f to minimize the error between the predicted outcome and the true observed one for all the patients of our cohort



Linear classifier

$$f(x_1, x_2) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = \hat{y}$$



❶ Learn how to optimally combine known biomarkers (e.g $f(PDL1, TMB)$)

❷ Discover and integrate new informative features

❸ Make new hypotheses on the mechanisms modulating treatment response and motivate further experiments

TIPIT: multimodal Machine Learning for powerful predictions

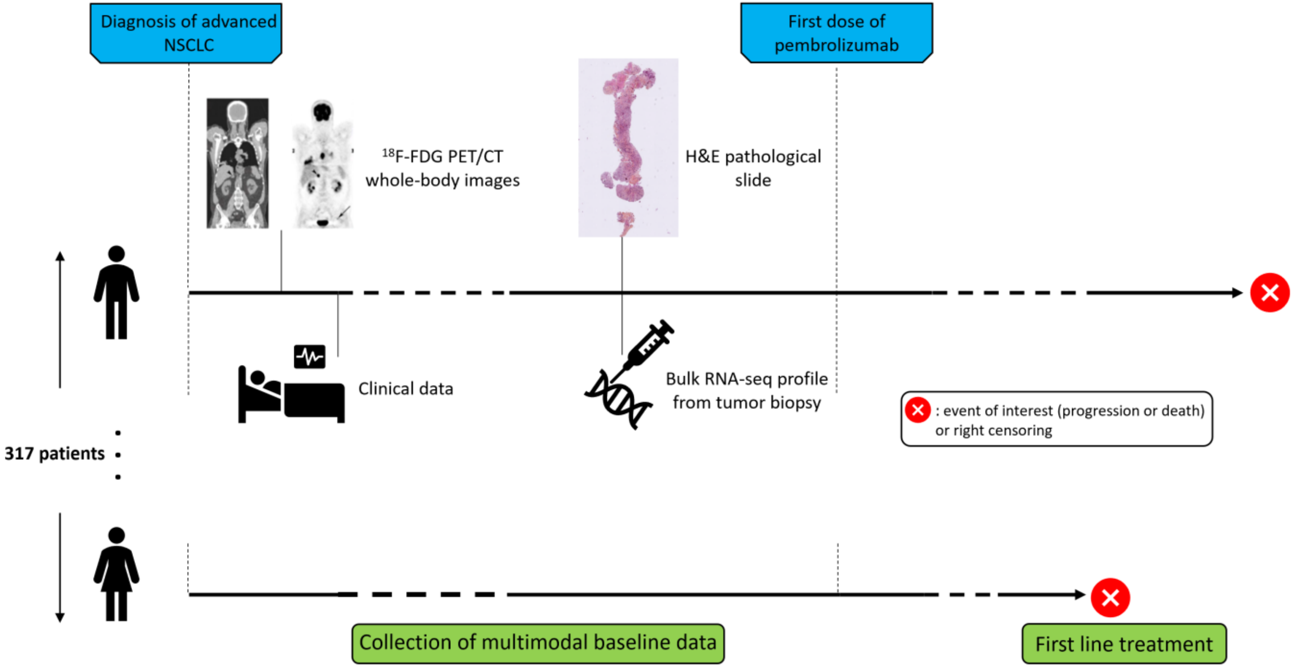
- A wide variety of data modalities can now be collected for many patients at baseline

- They offer complementary views of the disease at different scales (molecular, tissular, whole-body…)

- Integrating all these modalities together may lead to more powerful/accurate biomarkers

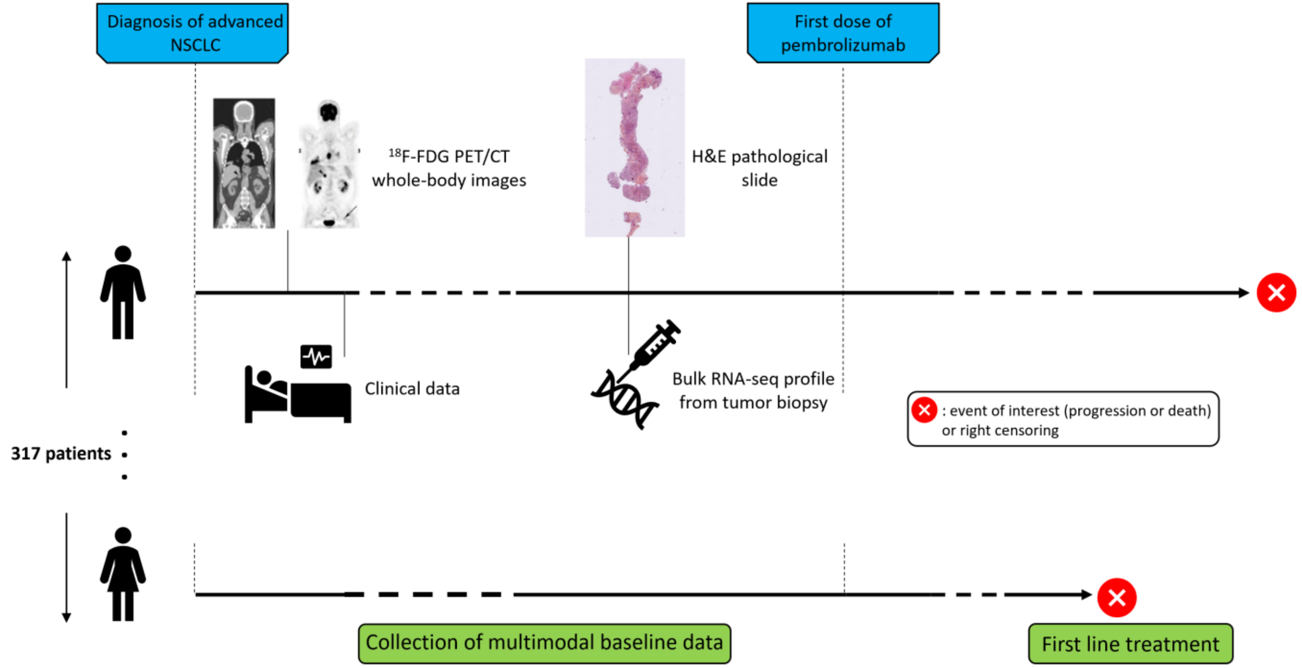
TIPIT: multimodal Machine Learning for powerful predictions

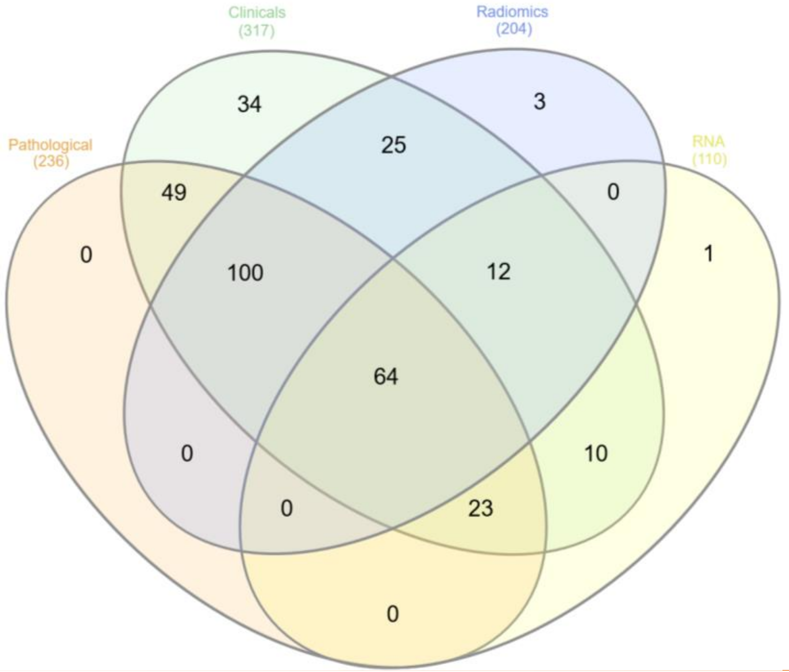
- A wide variety of data modalities can now be collected for many patients at baseline
- They offer complementary views of the disease at different scales (molecular, tissular, whole-body…)
- Integrating all these modalities together may lead to more powerful/accurate biomarkers



TIPIT: multimodal Machine Learning for powerful predictions

- A wide variety of data modalities can now be collected for many patients at baseline
- They offer complementary views of the disease at different scales (molecular, tissular, whole-body…)
- Integrating all these modalities together may lead to more powerful/accurate biomarkers



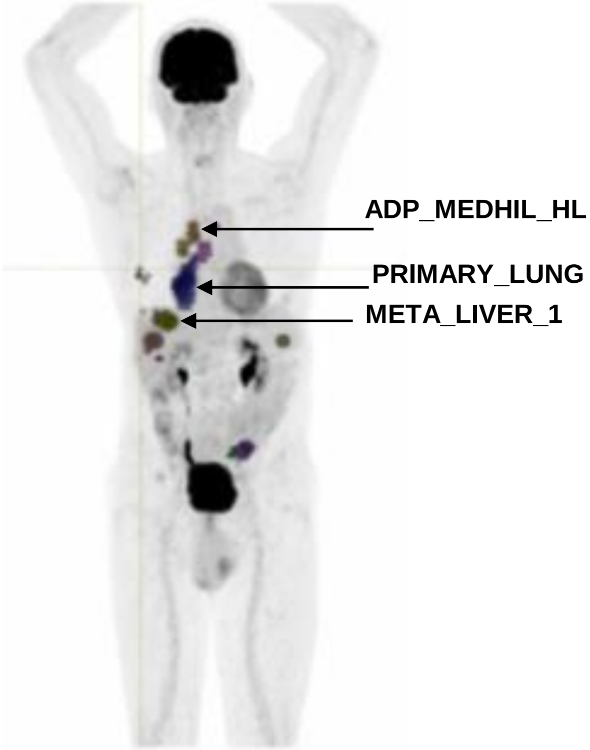


Outline

1. **Introduction:** predict the outcome of metastatic non-small cell lung cancer patients treated by immunotherapy
2. **Unimodal approaches:** extract relevant features from each modality
3. **Multimodal approach:** handle missing modalities with late fusion strategies

Medical imaging data & radiomics analyses

18F-FDG PET scans characterize the molecular heterogeneity of each metastasis as well as the spread of the disease in the whole-body

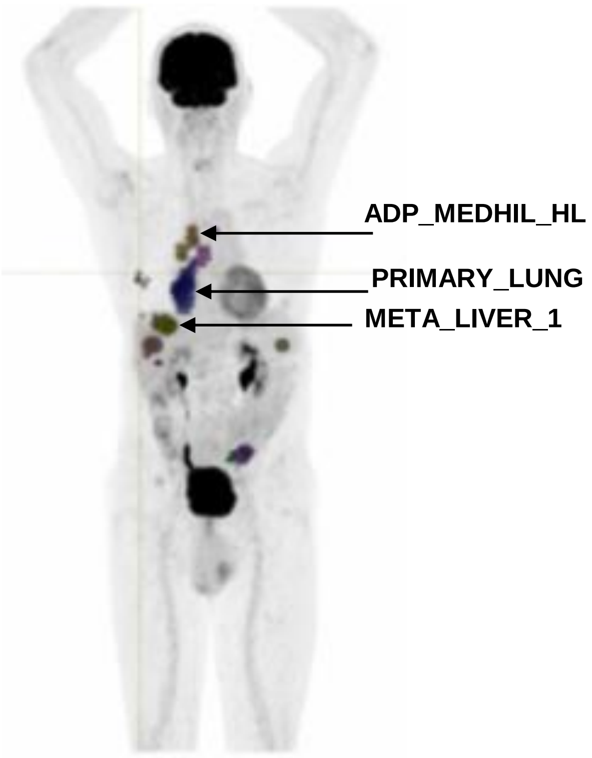


18F-FDG PET scan at baseline for a metastatic NSCLC patients. Segmentation and annotation of each lesion

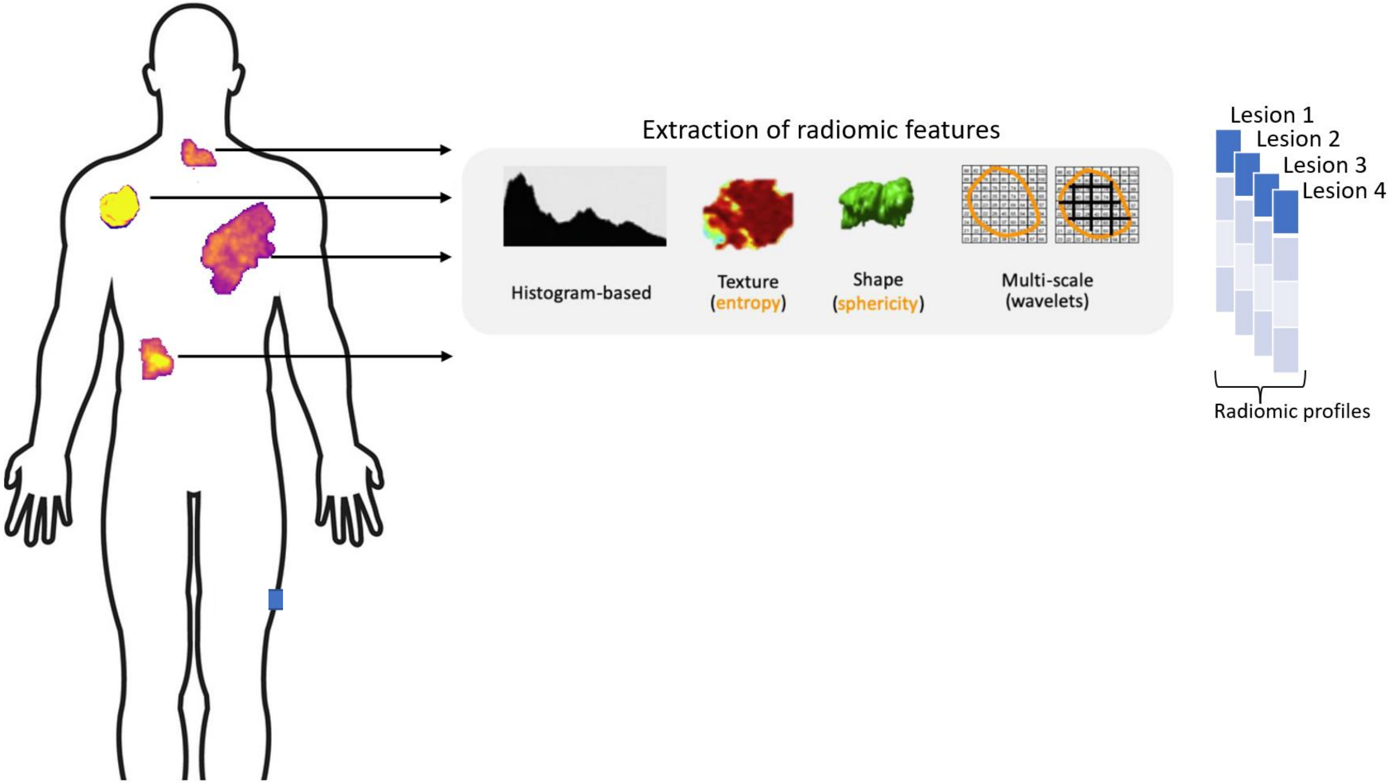
Extraction of radiomic features for machine learning (at the lesion and the whole-body levels)

Medical imaging data & radiomics analyses

18F-FDG PET scans characterize the molecular heterogeneity of each metastasis as well as the spread of the disease in the whole-body



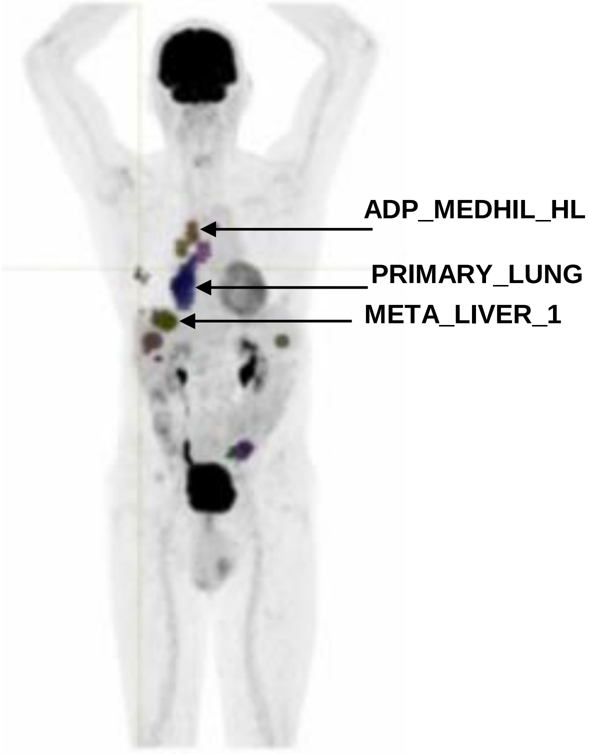
18F-FDG PET scan at baseline for a metastatic NSCLC patients. Segmentation and annotation of each lesion



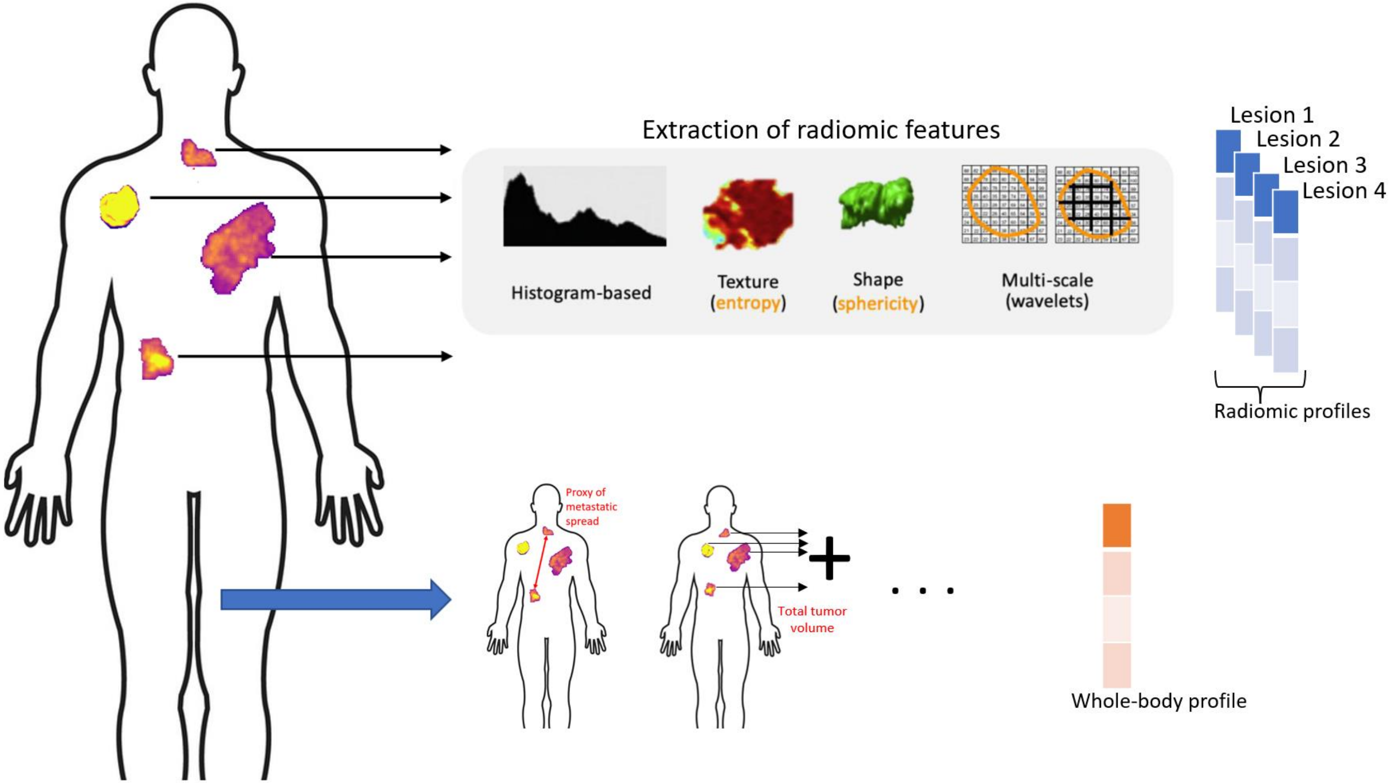
Extraction of radiomic features for machine learning (at the lesion and the whole-body levels)

Medical imaging data & radiomics analyses

18F-FDG PET scans characterize the molecular heterogeneity of each metastasis as well as the spread of the disease in the whole-body

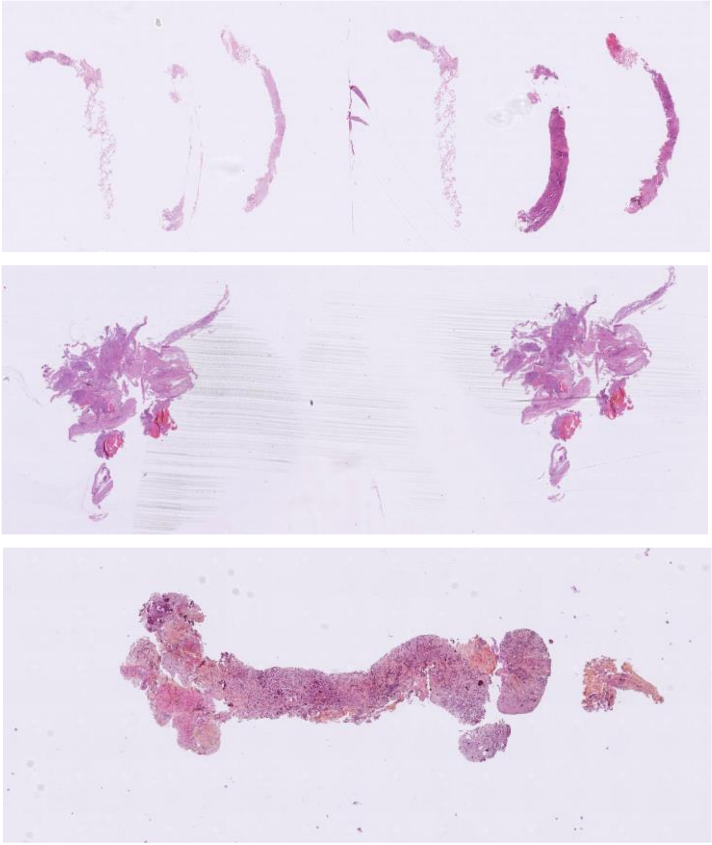


18F-FDG PET scan at baseline for a metastatic NSCLC patients. Segmentation and annotation of each lesion



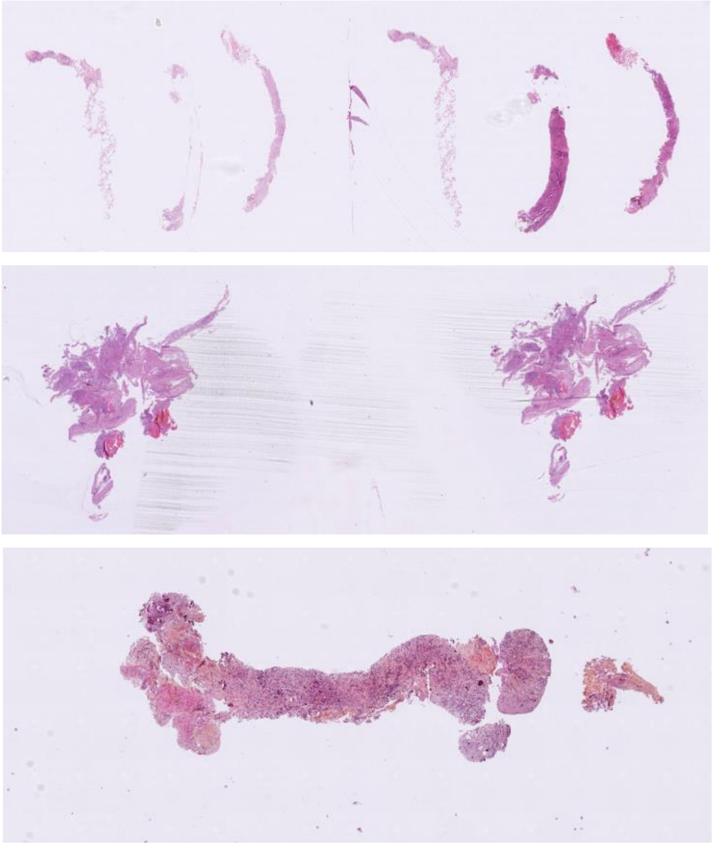
Extraction of radiomic features for machine learning (at the lesion and the whole-body levels)

H&E pathological slides & pathomics analyses

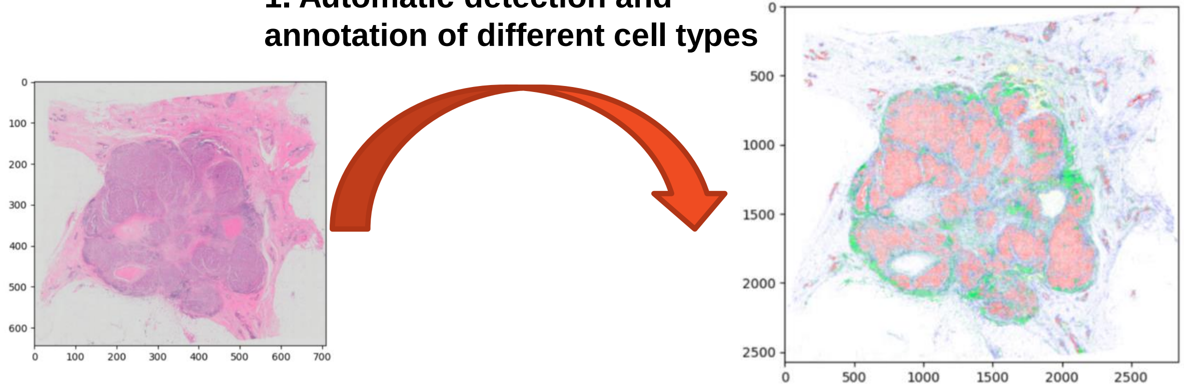


H&E slides from a diagnostic biopsy of a NSCLC patient

H&E pathological slides & pathomics analyses



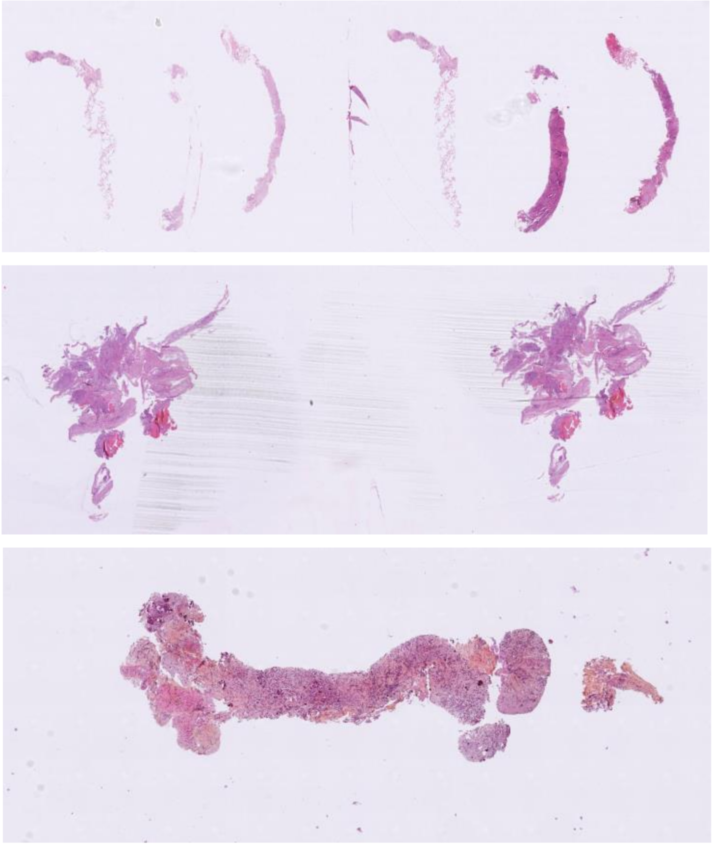
H&E slides from a diagnostic biopsy of a NSCLC patient



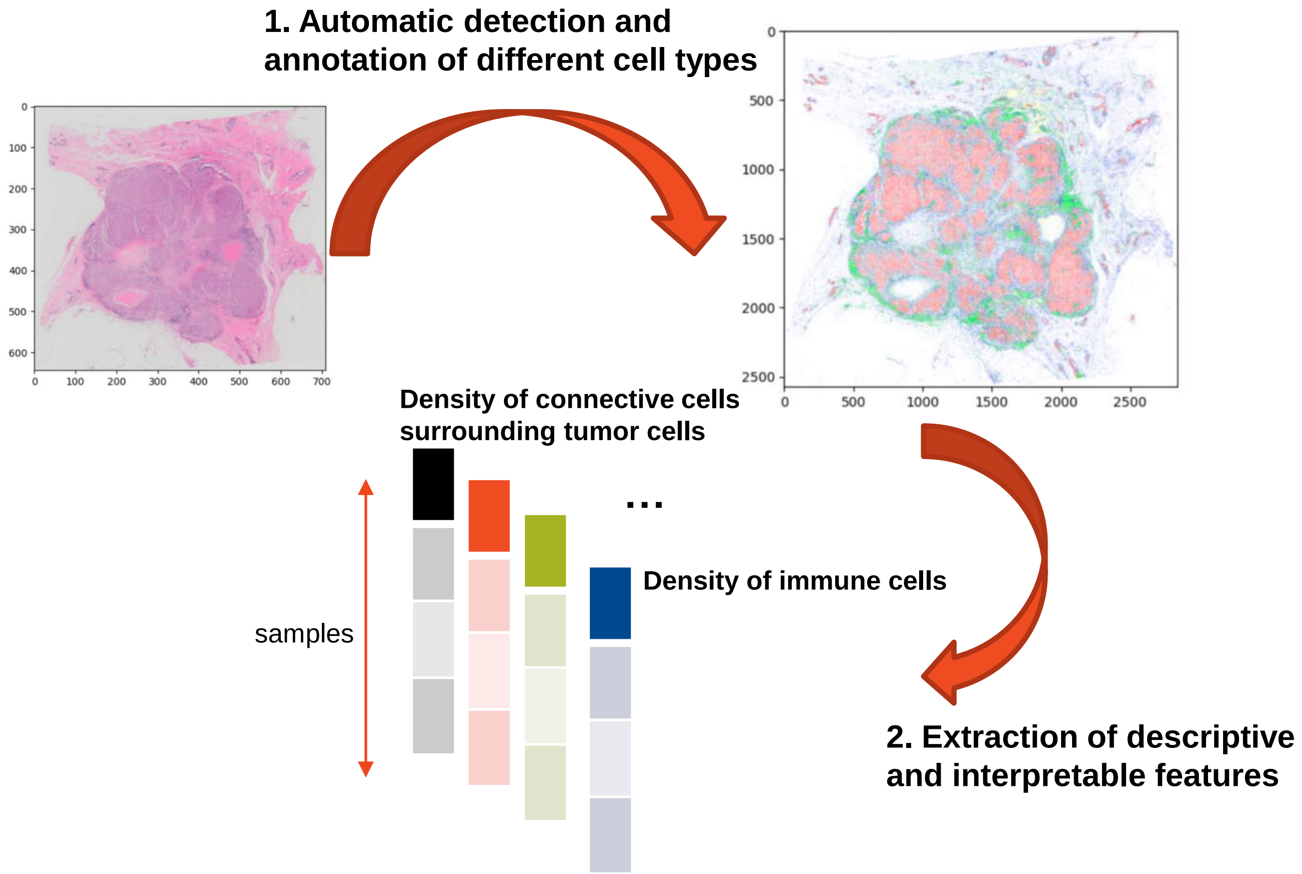
1. Automatic detection and annotation of different cell types

Extraction of pathomics features for machine learning. Illustration with a breast cancer slide with lymphocytes (green), connective cells (blue) and neoplastic cells (red) (Marvin Lerousseau)

H&E pathological slides & pathomics analyses



H&E slides from a diagnostic biopsy of a NSCLC patient



1. Automatic detection and annotation of different cell types

Density of connective cells surrounding tumor cells

Density of immune cells

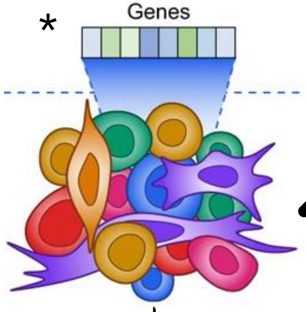
samples

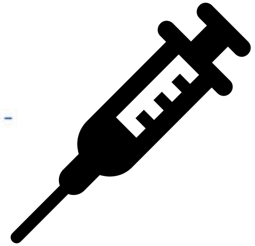
2. Extraction of descriptive and interpretable features

Extraction of pathomics features for machine learning. Illustration with a breast cancer slide with lymphocytes (green), connective cells (blue) and neoplastic cells (red) (Marvin Lerousseau)

Bulk RNAseq: handle highly dimensional data

*

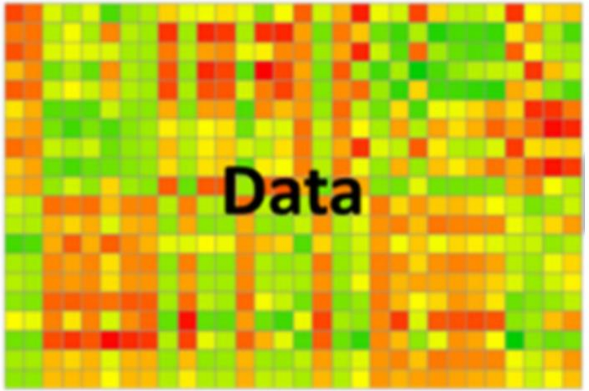




Solid biopsy of one patient

$\sim 2 \times 10^4$ protein-coding genes

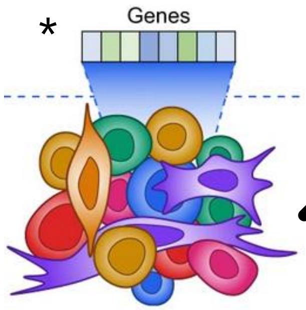
110 samples



* Liao *et al.* Nature Communications 2022

Bulk RNAseq: handle highly dimensional data

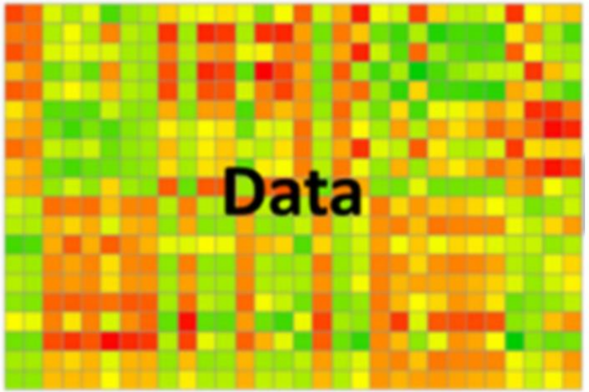
*



Solid biopsy of one patient

1. Expression of specific oncogenes

$\sim 2 \times 10^4$ protein-coding genes

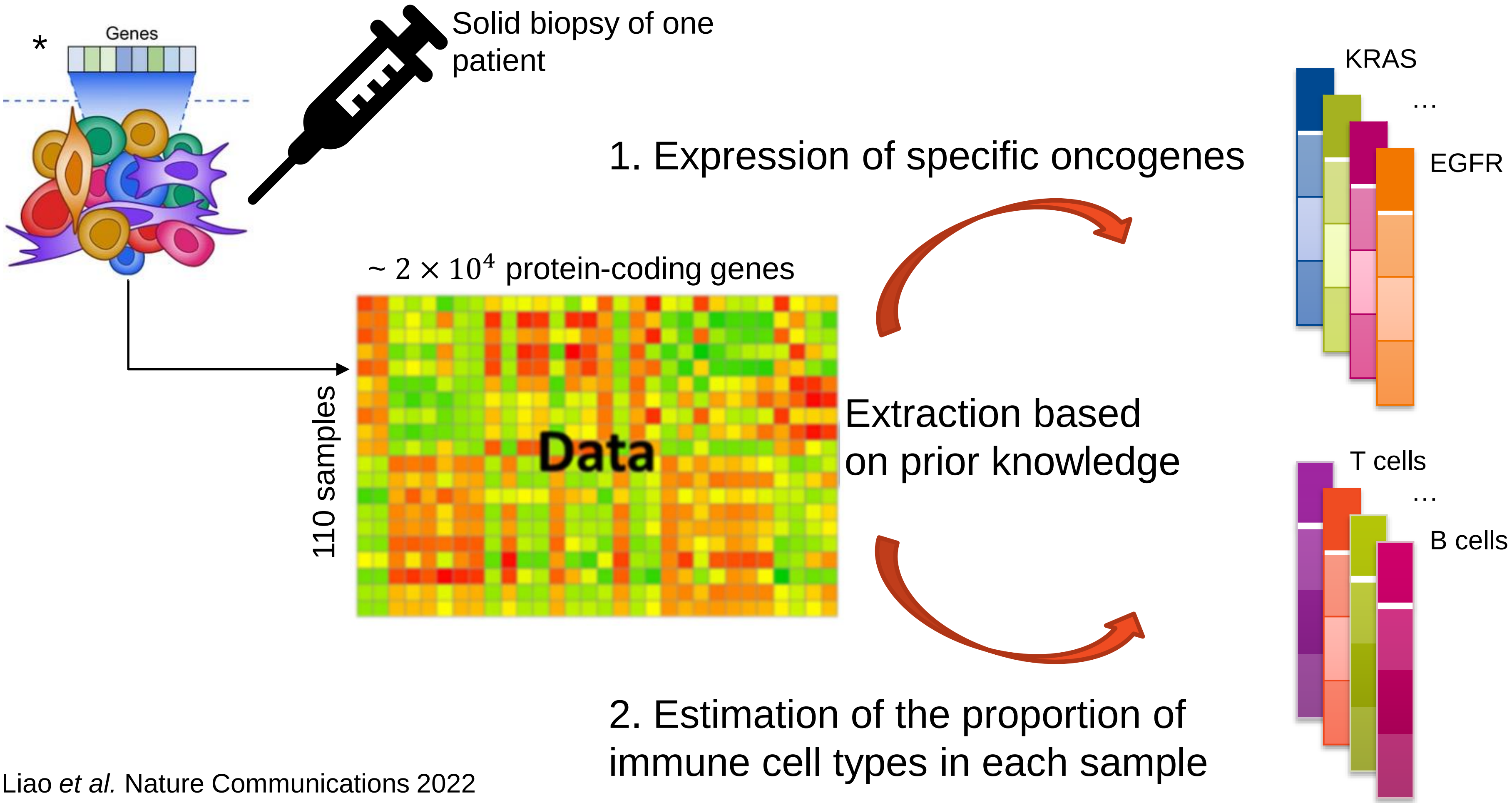


110 samples

Extraction based on prior knowledge

* Liao *et al.* Nature Communications 2022

Bulk RNAseq: handle highly dimensional data



Solid biopsy of one patient

~ 2×10^4 protein-coding genes

1. Expression of specific oncogenes

Extraction based on prior knowledge

2. Estimation of the proportion of immune cell types in each sample

* Liao *et al.* Nature Communications 2022

Outline

1. **Introduction:** predict the outcome of metastatic non-small cell lung cancer patients treated by immunotherapy
2. **Unimodal approaches:** extract relevant features from each modality
3. **Multimodal approach:** handle missing modalities with late fusion strategies

Late fusion strategy: aggregate unimodal predictions

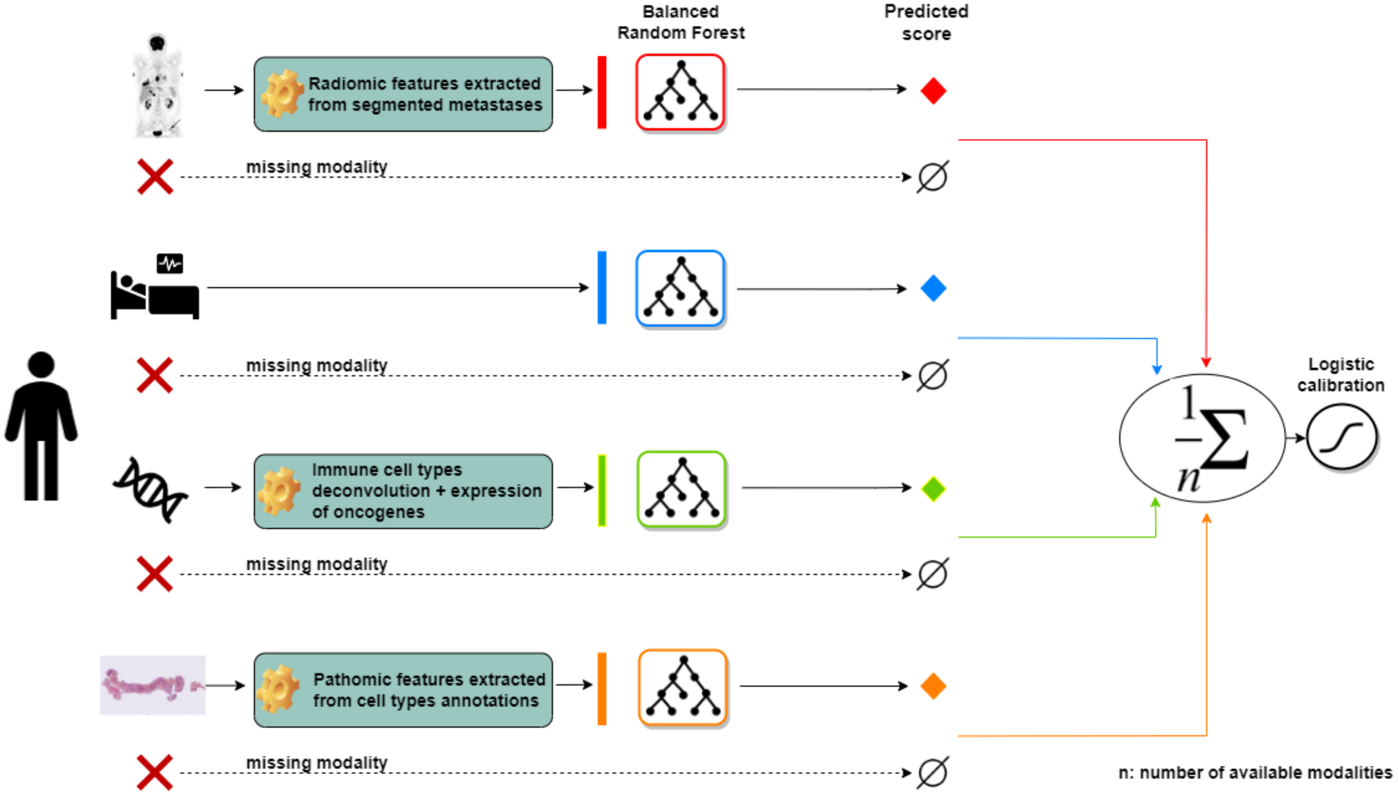
- Need multimodal models that could be trained without all the modalities for each patient

- If a new patient does not have all the modalities available, we still want to make a prediction

Late fusion strategy: aggregate unimodal predictions

- Need multimodal models that could be trained without all the modalities for each patient

- If a new patient does not have all the modalities available, we still want to make a prediction

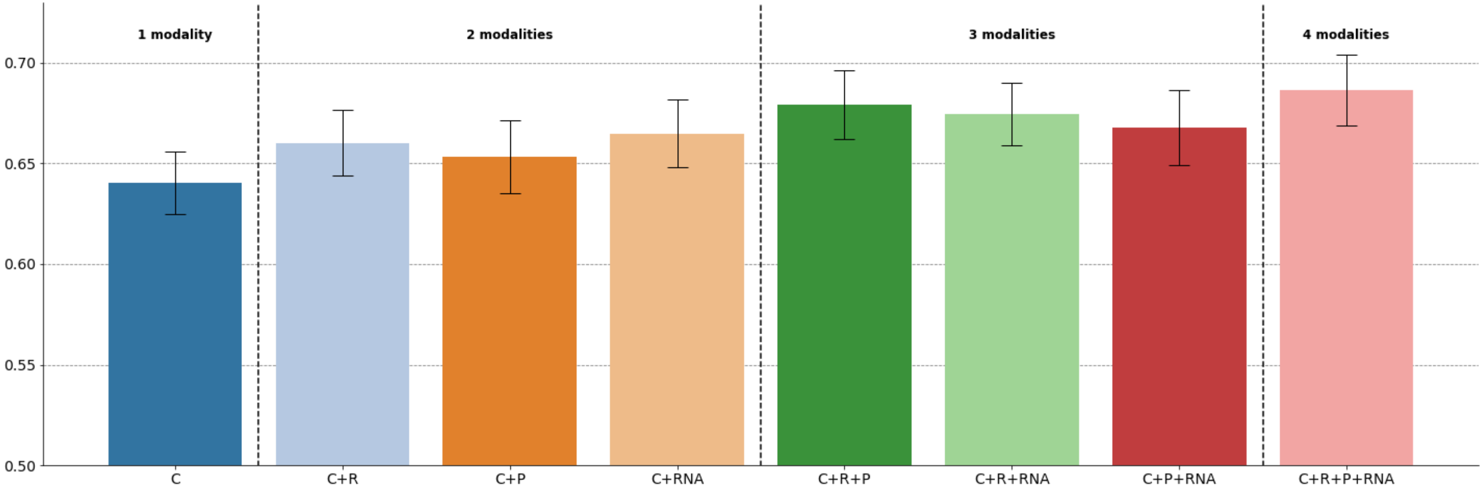


Prediction of the survival at 1 year (binary classification)

Does adding other modalities to the clinical data, whenever they are available, improve the prediction of the death before 1 year ?

Prediction of the survival at 1 year (binary classification)

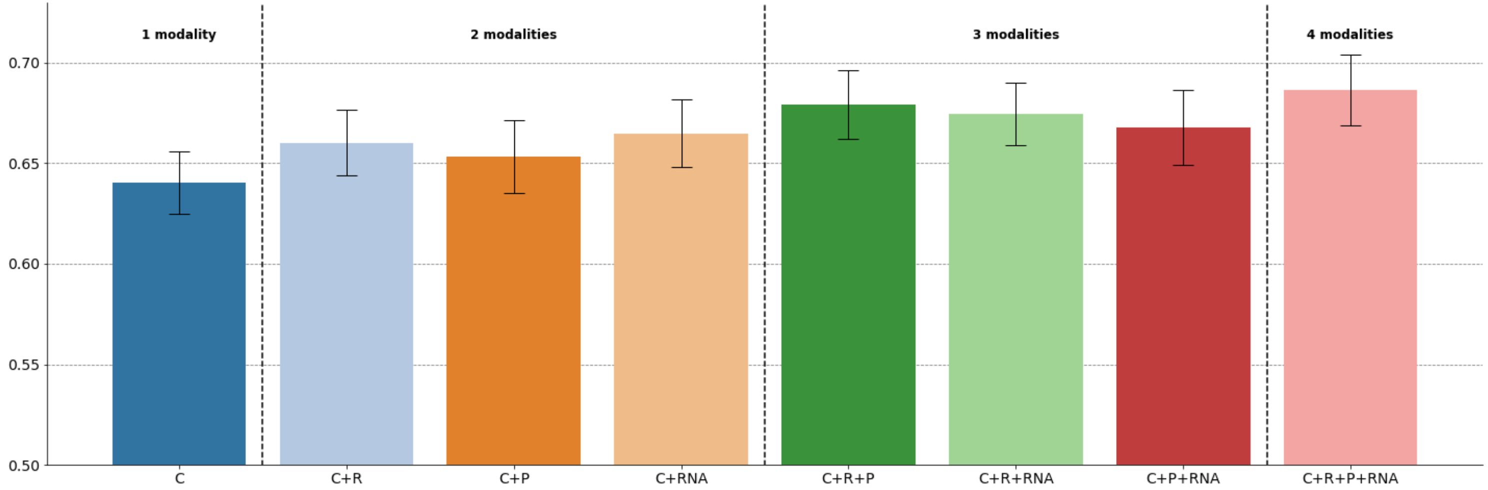
Does adding other modalities to the clinical data, whenever they are available, improve the prediction of the death before 1 year ?



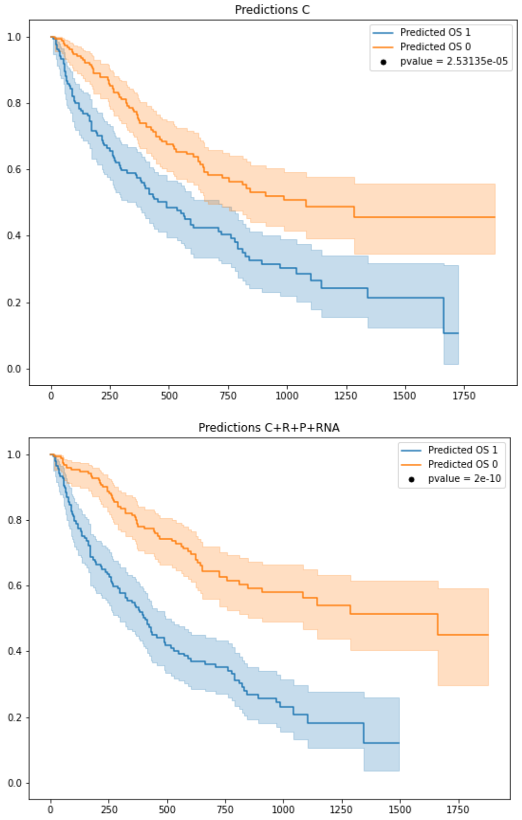
ROC AUC scores of different multimodal combinations averaged over 100 experiments

Prediction of the survival at 1 year (binary classification)

Does adding other modalities to the clinical data, whenever they are available, improve the prediction of the death before 1 year ?



ROC AUC scores of different multimodal combinations averaged over 100 experiments



Kaplan-Meier curves comparing the two predicted classes (clinical model vs multimodal model)

Conclusions & perspectives

- Adding available modalities to clinical data improve the prediction of the patient’s outcome (both for Overall Survival and Progression-Free Survival).

Conclusions & perspectives

- Adding available modalities to clinical data improve the prediction of the patient’s outcome (both for Overall Survival and Progression-Free Survival).

- It shows the potential of multimodal strategies to build powerful predictors for the prognosis of metastatic lung cancer patients that may also be predictive of the response to immunotherapy.

Conclusions & perspectives

- Adding available modalities to clinical data improve the prediction of the patient’s outcome (both for Overall Survival and Progression-Free Survival).

- It shows the potential of multimodal strategies to build powerful predictors for the prognosis of metastatic lung cancer patients that may also be predictive of the response to immunotherapy.

1. Collect more data to increase the number of patients with complete profiles (i.e 4 modalities collected).

Conclusions & perspectives

- Adding available modalities to clinical data improve the prediction of the patient’s outcome (both for Overall Survival and Progression-Free Survival).

- It shows the potential of multimodal strategies to build powerful predictors for the prognosis of metastatic lung cancer patients that may also be predictive of the response to immunotherapy.

1. Collect more data to increase the number of patients with complete profiles (i.e 4 modalities collected).

2. Collect new modalities (mutation data, multiplex imaging, Visium spatial transcriptomics) to gain more biological insights on the mechanisms that drive patient’s response.

Conclusions & perspectives

- Adding available modalities to clinical data improve the prediction of the patient’s outcome (both for Overall Survival and Progression-Free Survival).

- It shows the potential of multimodal strategies to build powerful predictors for the prognosis of metastatic lung cancer patients that may also be predictive of the response to immunotherapy.

1. Collect more data to increase the number of patients with complete profiles (i.e 4 modalities collected).

2. Collect new modalities (mutation data, multiplex imaging, Visium spatial transcriptomics) to gain more biological insights on the mechanisms that drive patient’s response.

3. Explore and design new features for the different modalities to use their full potential.

Acknowledgements

- **Curie-Montsouris Chest center:** Nicolas Girard, Sarah Lagha, Anne-Sophie Tedesco
- **Imaging department Curie:** Hervé Brisse, Marie Luporsi, Toulsie Ramtohul
- **Pathology department Curie:** Clément Beaulaton
- **Data department Curie:** Paulette Salamoun Feghali
- **LITO Curie:** Irène Buvat, Fanny Orlhac, Narinée Hovhannisyan, Nicolas Captier
- **SysBIO U900 Curie:** Emmanuel Barillot, Andrei Zinovyev, Christine Lonjou, Nicolas Captier
- **CBIO U900 Curie-Mines:** Thomas Walter, Marvin Lerousseau