



# miRNA Biomarker for Lung Cancer Diagnostics

Selecting a test panel for  
patient classification

Dr. Patrick Neubert  
13.05.2020

**01**

## **miRNA**

Introduction on miRNA

**02**

## **LUNG CANCER**

Lung Cancer - Key Numbers

Lung Cancer - Current Diagnostics

**03**

## **PROJECT GOALS & OBJECTIVES**

Project Goals & Objectives

**04**

## **DATA EXPLORATION & WORK FLOW**

Data Exploration on Annotation & miRNA Data  
Work-Flow for Predictive Modeling

**05**

## **RESULTS**

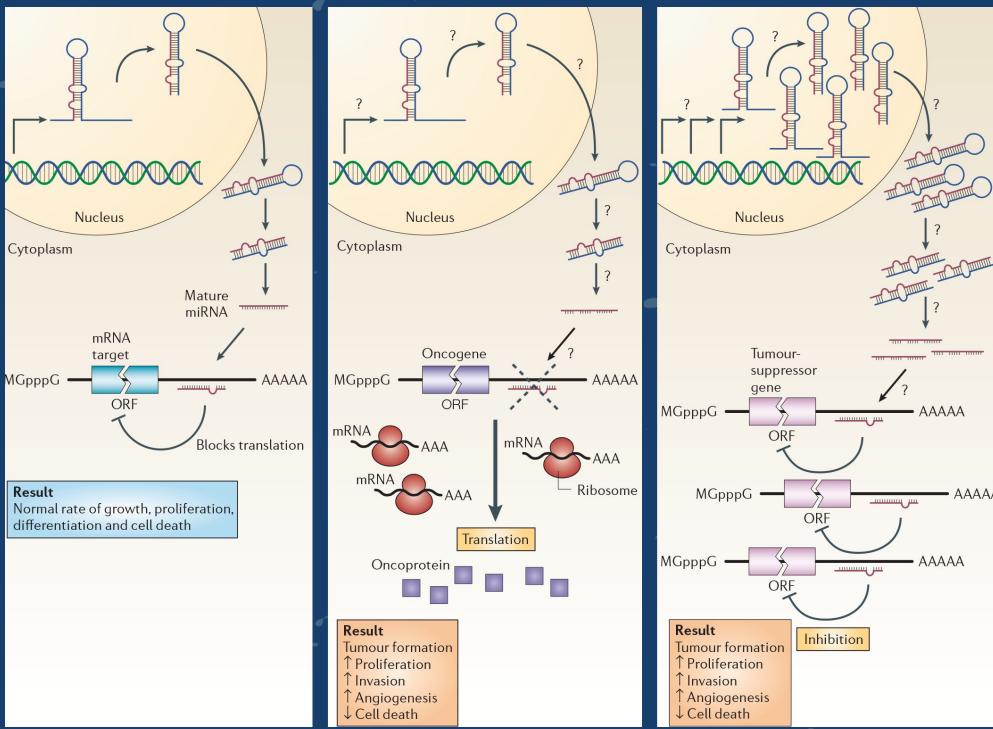
Feature Selection, Model Selection, Model  
Optimization, Feature Reduction, Model Evaluation

**06**

## **CONCLUSION & FUTURE WORK**

Conclusion & Future Work

# miRNA



Esquela-Kerscher & Slack, 2006

- Small non-protein coding RNA (~22 nucleotides)
- Post-transcriptional negative gene regulators
- Multi target genes
- Oncomirs (tumor-suppressive or oncogenic)

# LUNG CANCER - KEY NUMBERS

**2,090,000**

Worldwide new cases of lung cancer  
(Ferlay et al., 2018)

**1,760,000**

Worldwide deaths from lung cancer  
(Ferlay et al., 2018)

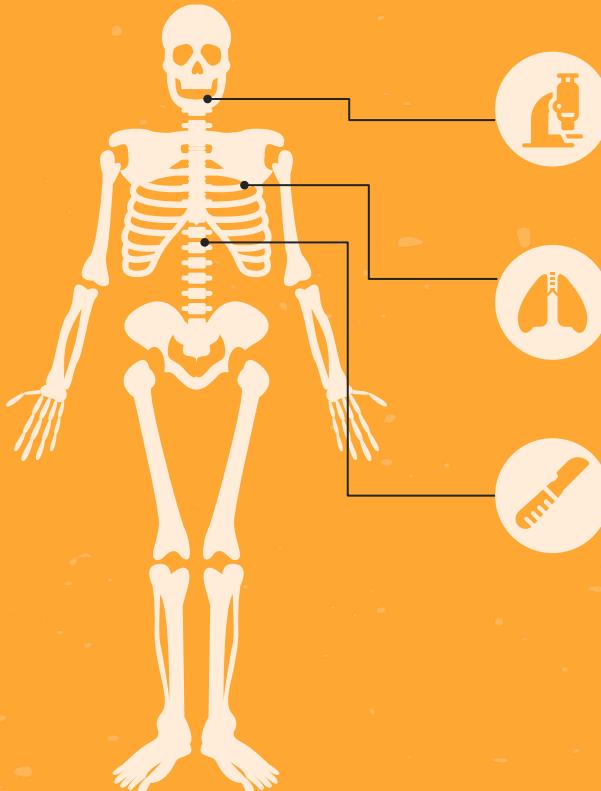
**70%**

5-year overall survival if detected at  
earlier stages (Robinson et al., 2011)

**5%**

5-year overall survival if detected at  
advanced stage (Siegel et al., 2012)

# LUNG CANCER - CURRENT DIAGNOSTICS



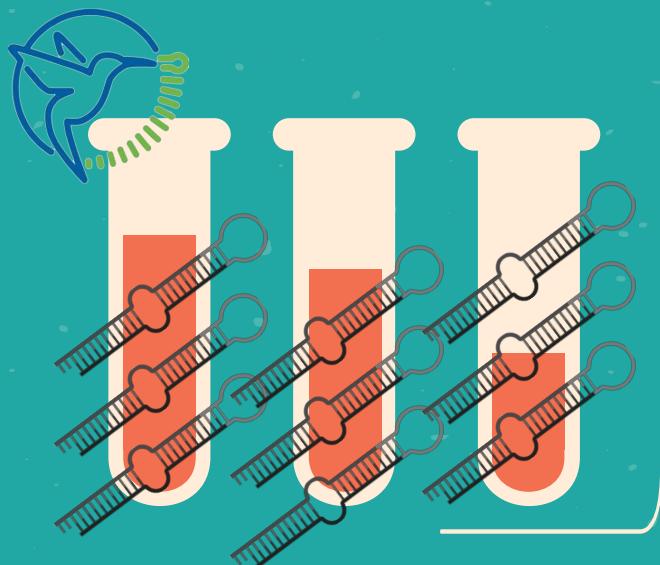
Sputum cytology

X-ray imaging

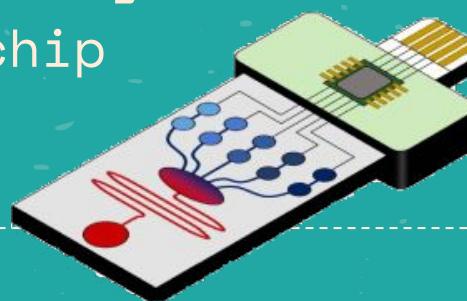
Tissue sample  
(Biopsy)

- No early detection screening methods
- High false positive rates & radiation exposure

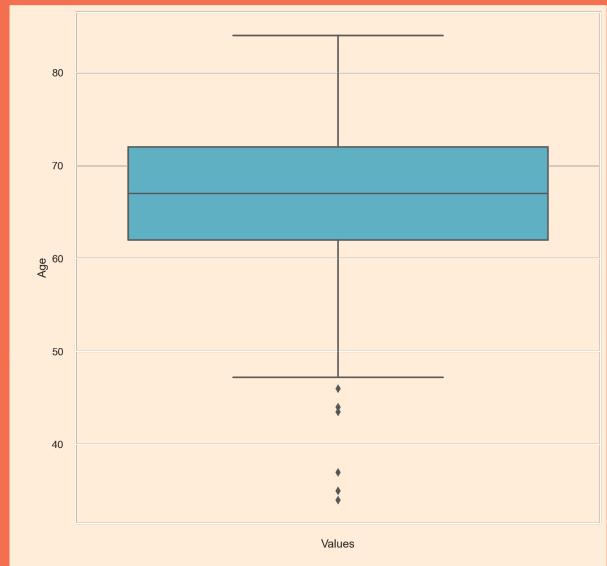
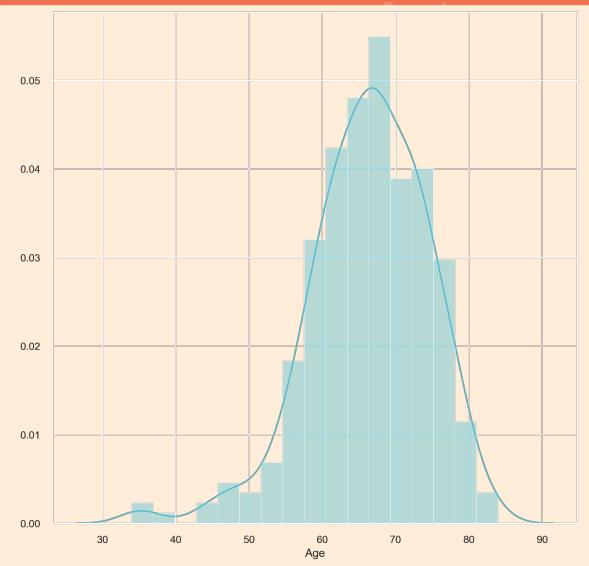
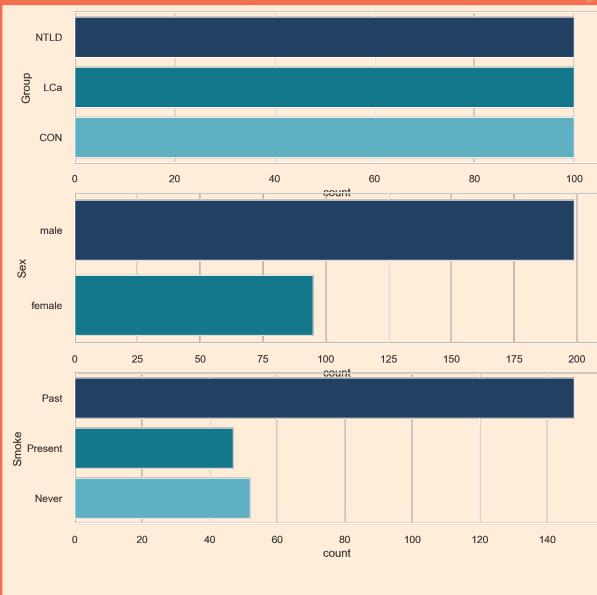
# PROJECT GOALS AND OBJECTIVES



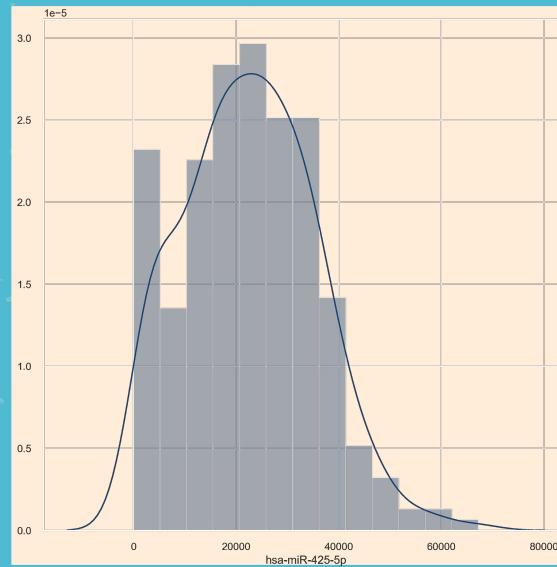
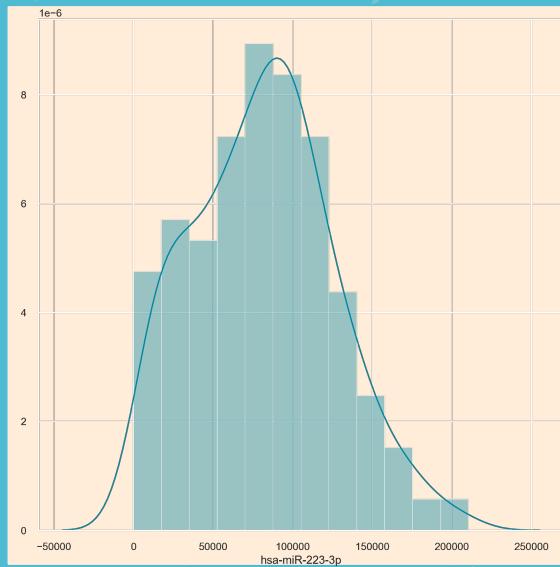
- miRNA Expression Levels of 300 blood samples
- Annotation Data with 3 classes (LCa, NTLD, CON)
- Identify 10-20 miRNA for biochip



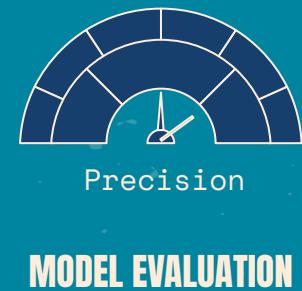
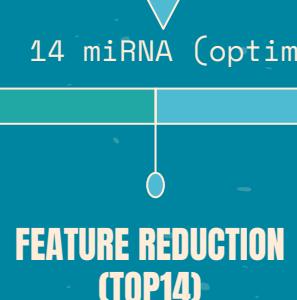
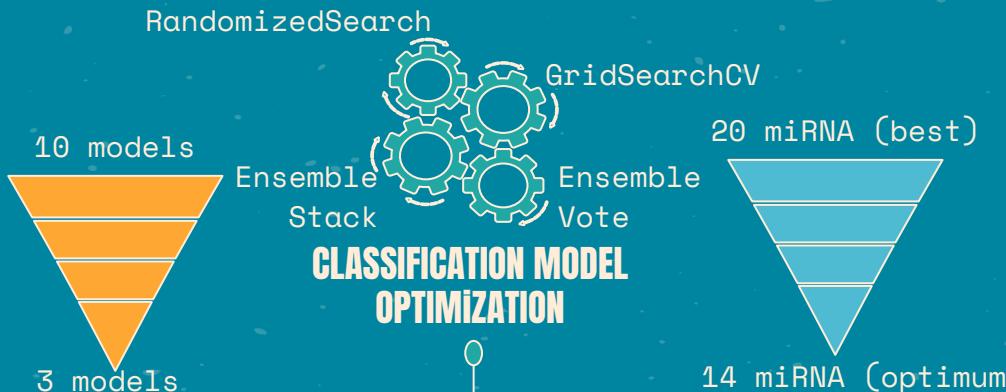
# DATA EXPLORATION - ANNOTATION



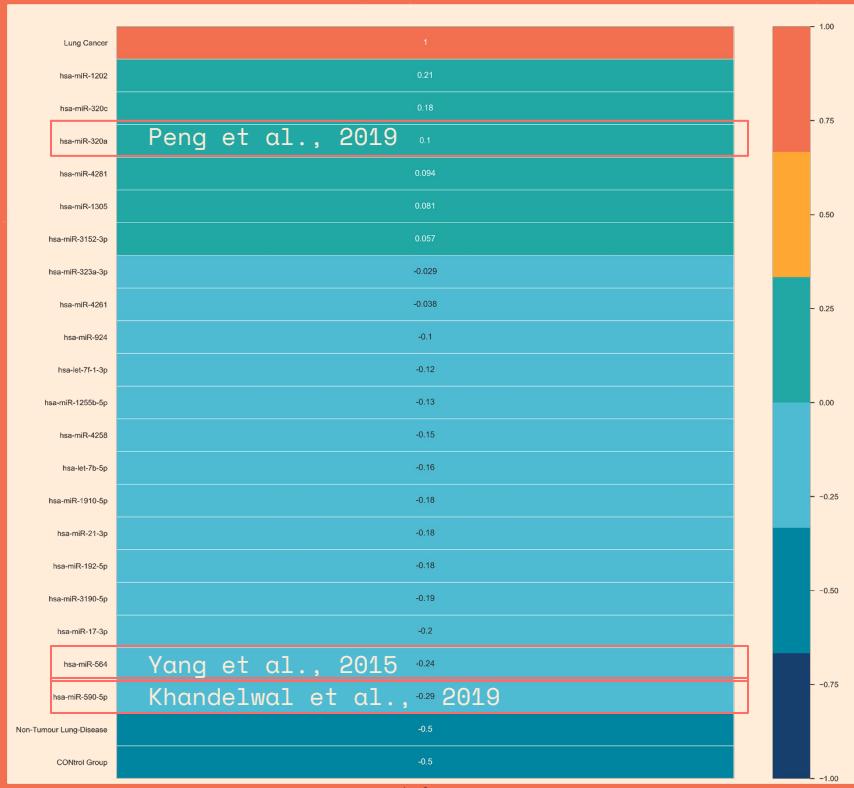
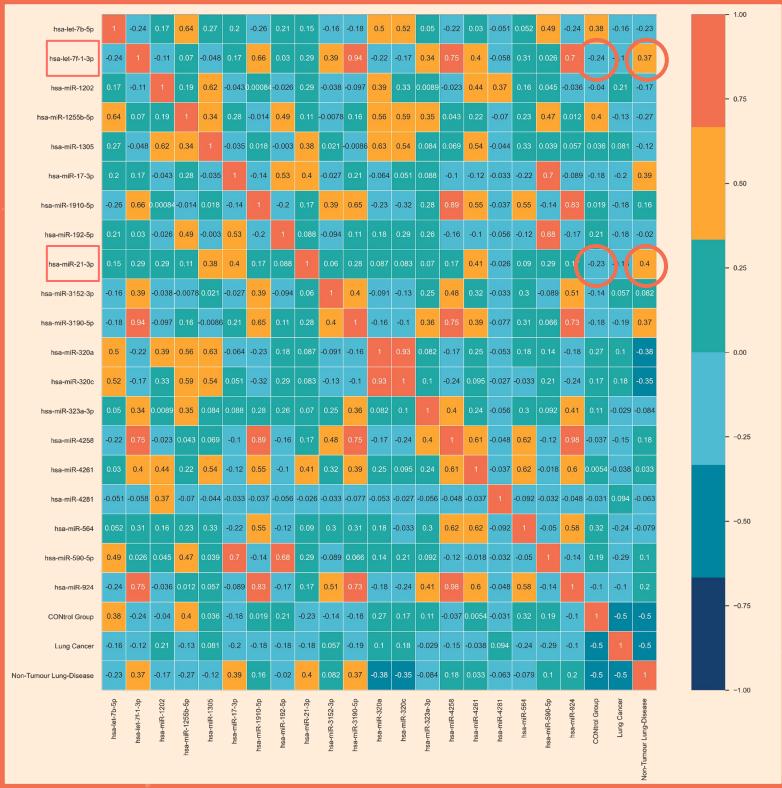
# DATA EXPLORATION - miRNA



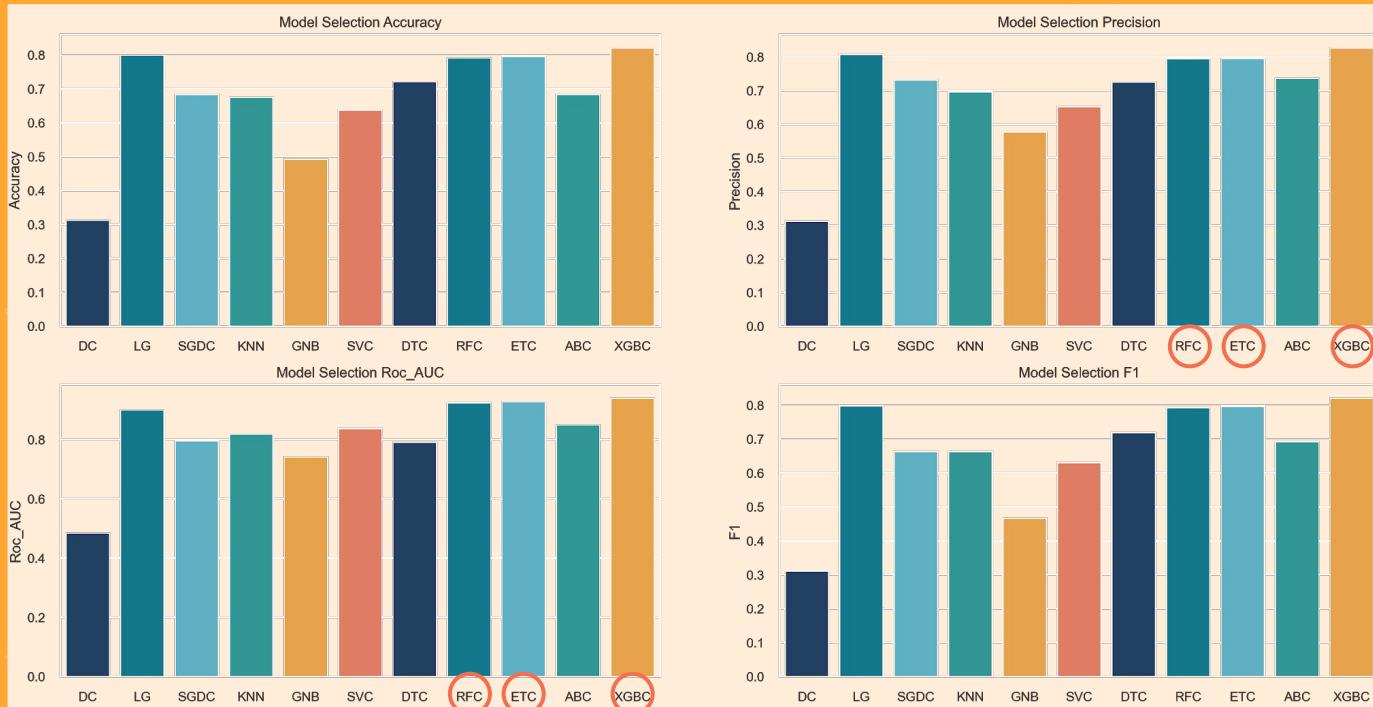
# WORK-FLOW



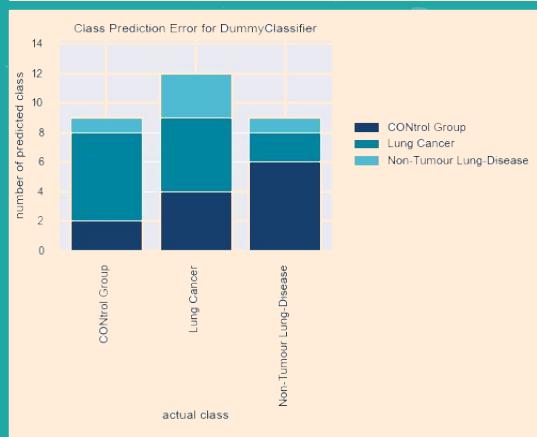
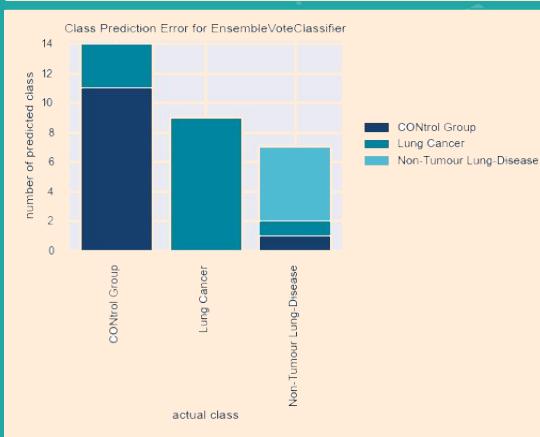
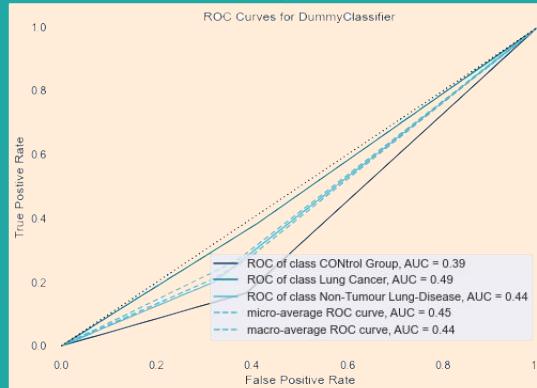
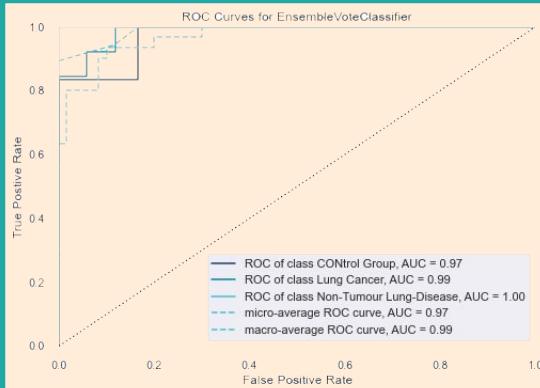
# RESULTS - FEATURE SELECTION



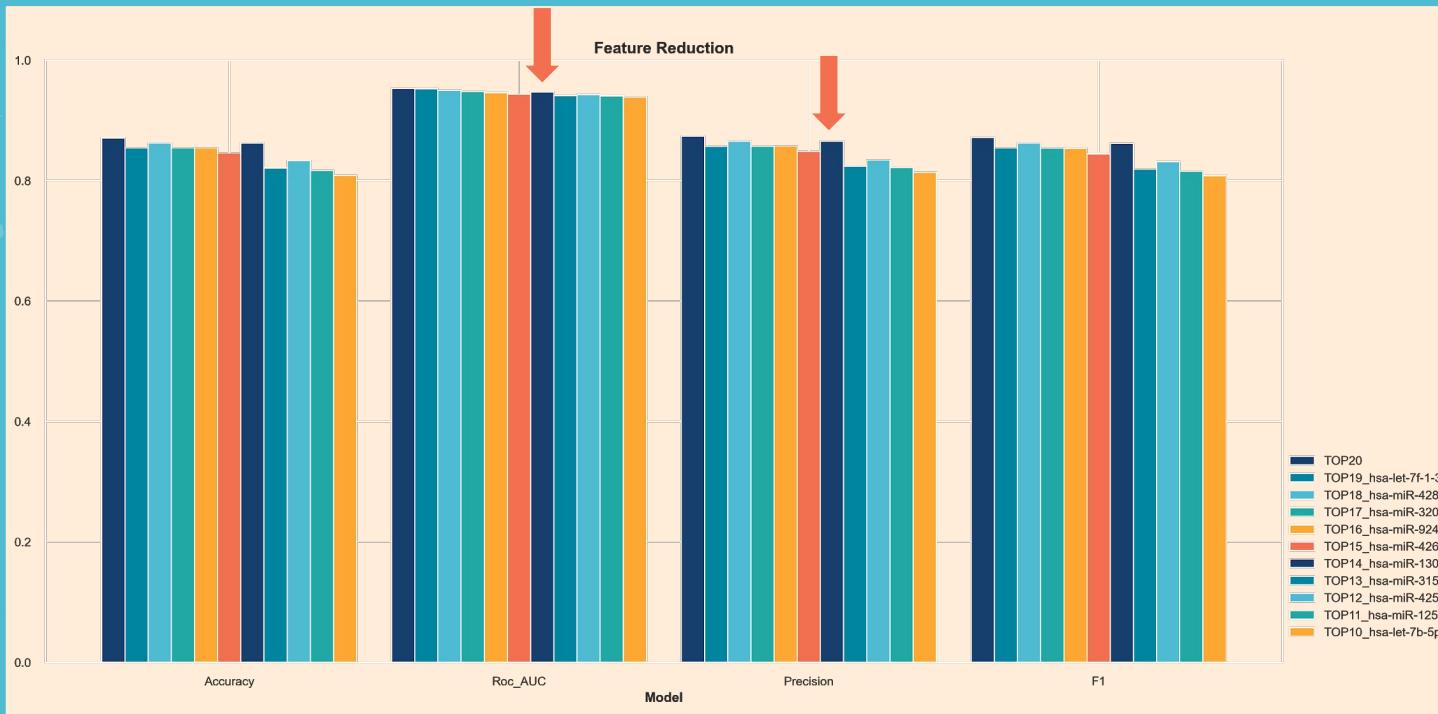
# RESULTS - MODEL SELECTION



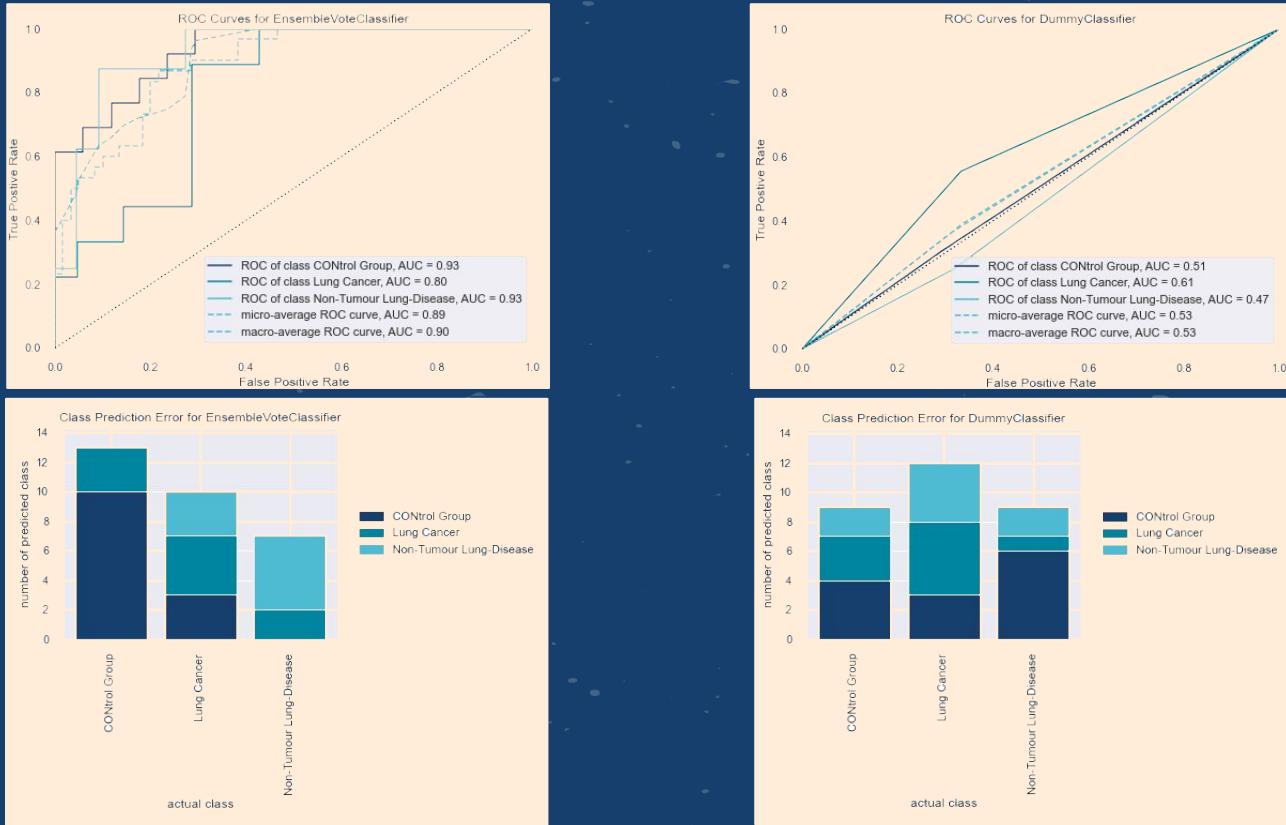
# RESULTS - MODEL OPTIMIZATION



# RESULTS - FEATURE REDUCTION



# RESULTS - MODEL EVALUATION



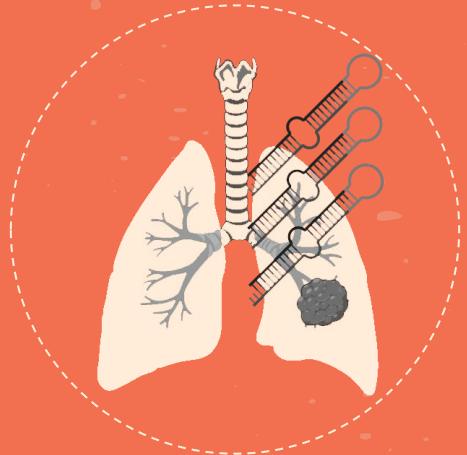
# CONCLUSION & FUTURE WORK



- Defined a TOP20 (best) and TOP14 (optimum) panel of miRNA applicable in a biochip test system for lung cancer
- 3 of high-correlating-LCa miRNAs were already described in literature for their implication in Lung Tumors
- Built a classification model with medium performance to predict LCa, NTLD and CON classes



- Reduce classification to binary problem
- Use different models for prediction of each class
- Implement more regularization penalizing overfitting and other preprocessing methods
- Collect more samples to improve modeling (e.g. deep learning)



# THANKS!

Do you have any questions?



[patrick.neubert@t-online.de](mailto:patrick.neubert@t-online.de)

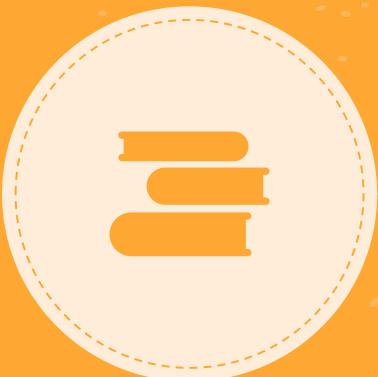


[https://github.com/Patrick-Neubert/miRNA\\_LCa\\_Diagnostics](https://github.com/Patrick-Neubert/miRNA_LCa_Diagnostics)

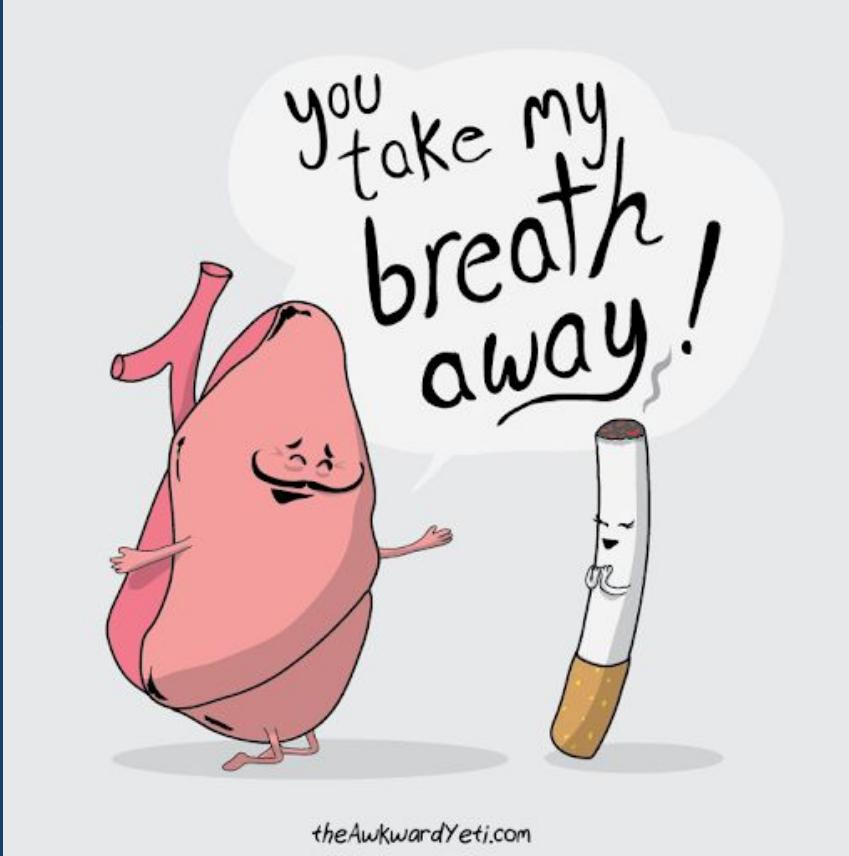


<https://www.linkedin.com/in/dr-patrick-neubert-91910b111/>

# REFERENCES



- Esquela-Kerscher & Slack (2006). Oncomirs – microRNAs with a role in cancer
- Ferley et al. (2018). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods
- Robinson et al. (2011). Node-Negative Non-small Cell Lung Cancer
- Siegel et al. (2012). Cancer Statistics
- Khandelwal et al. (2019). Circulating microRNA-590-5p functions as a liquid biopsymarker in non-small cell lung cancer
- Yang et al. (2015). MiR-564 functions as a tumor suppressor in human lung cancer by targeting ZIC3
- Peng et al. (2019). Correlation of plasma exosomal microRNAs with the efficacy of immunotherapy in EGFR/ALK wildtype advanced non-small cell lung cancer



theAwkwardYeti.com