LUXEMBOURG INSTITUTE OF HEALTH

Noncoding RNAs and machine learning for COVID-19 patients

Yvan Devaux, PhD, FESC
Luxembourg Institute of Health

12th National Biobank Symposium - NAPKON

23/09/2024

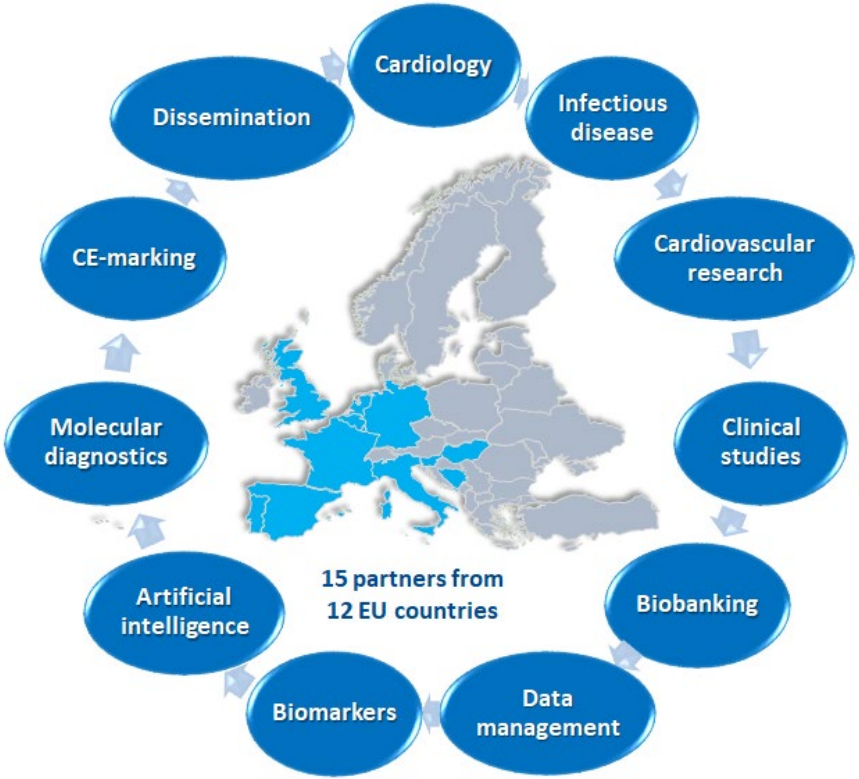
A diagnostic test to improve surveillance and care of COVID-19 patients

(H2020 SC1-PHE-CORONAVIRUS-2B projects)

European Commission

Horizon 2020
European Union funding
for Research & Innovation

COVIRNA multidisciplinary team



COVIRNA concept

- **Biomedical issue**

COVID-19 pandemic

Predicting clinical outcome of COVID-19 patients is a challenge

- **Unmet clinical need**

Early predictors of outcome in COVID-19 patients

- **Solution/goal of the COVIRNA project**

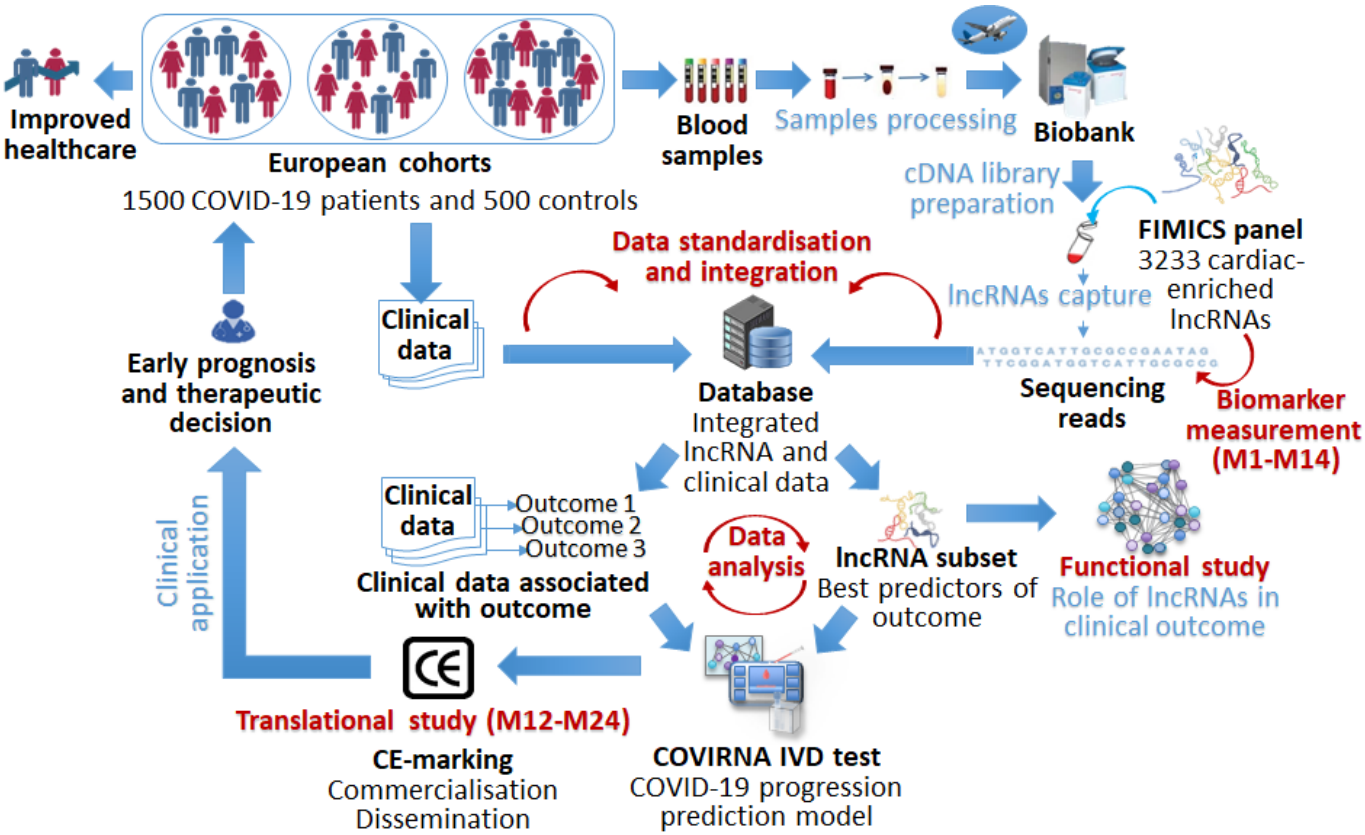
A molecular diagnostic kit to predict outcome of COVID-19 patients

→ Simple, minimally invasive and rapid blood test

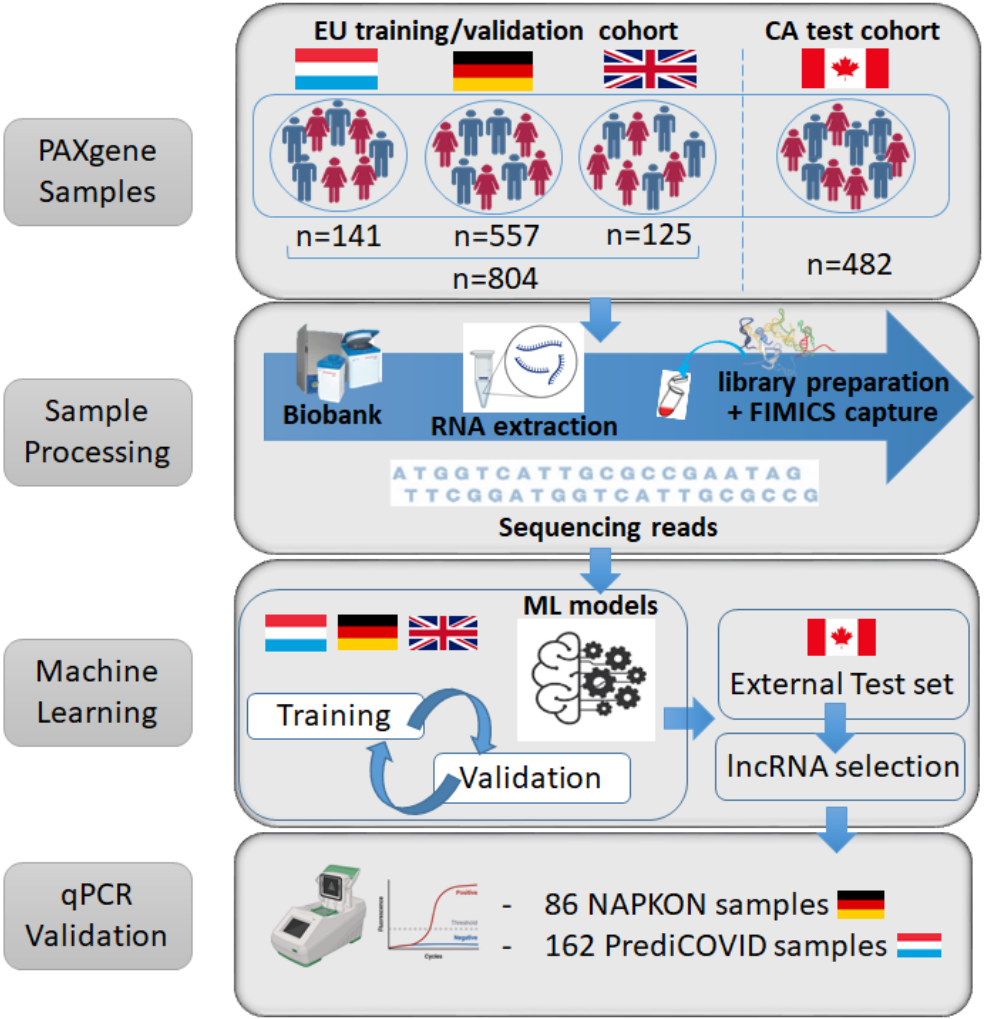
→ Based on the FIMICS panel of cardiac-enriched long noncoding RNAs (lncRNAs) for which the consortium has the intellectual property and exclusive licenses

→ Suitable for PCR diagnostic platforms and cost-effective.

Approach


Improved healthcare
European cohorts
1500 COVID-19 patients and 500 controls
Blood samples
Samples processing
Biobank
cDNA library preparation
FIMICS panel
3233 cardiac-enriched lncRNAs
lncRNAs capture
Sequencing reads
Biomarker measurement (M1-M14)
Data standardisation and integration
Clinical data
Early prognosis and therapeutic decision
Database
Integrated lncRNA and clinical data
Clinical application
Clinical data
Outcome 1
Outcome 2
Outcome 3
Clinical data associated with outcome
Data analysis
lncRNA subset
Best predictors of outcome
Functional study
Role of lncRNAs in clinical outcome
CE
Translational study (M12-M24)
CE-marking
Commercialisation
Dissemination
COVIRNA IVD test
COVID-19 progression prediction model

COVIRNA study workflow



PAXgene samples handling

The German National Pandemic Cohort Network (NAPKON):

- 583 COVID 19 positive PAXgene samples
- MTA signed June 2021
- Samples were sent to Firalis SA (France) for extraction and sequencing

PAXgene samples extraction and sequencing

- Batches of 64 samples were randomized for RNA extraction at Firalis
- RNA extraction using the KingFisher Apex instrument (fully automatized)
- RNA quantification with a Qubit Fluorometer
- Library preparation and capture using the FIMICS panel containing more than 55000 probes of 120 nucleotides each covering specific regions of 3233 lncRNAs, cardiac-enriched or associated with heart failure (Heliyon 2023; Noncoding RNA Res 2023)

Datasets available for analysis

- A technical issue in one run during targeted sequencing led to a loss of 26 samples, resulting in 557 RNAseq datasets available for analysis

- RNAseq datasets were combined with clinical data and data analysis was performed

Study on in-hospital mortality

nature communications

Article

https://doi.org/10.1038/s41467-024-47557-1

Development of a long noncoding RNA-based machine learning model to predict COVID-19 in-hospital mortality

Received: 11 December 2023

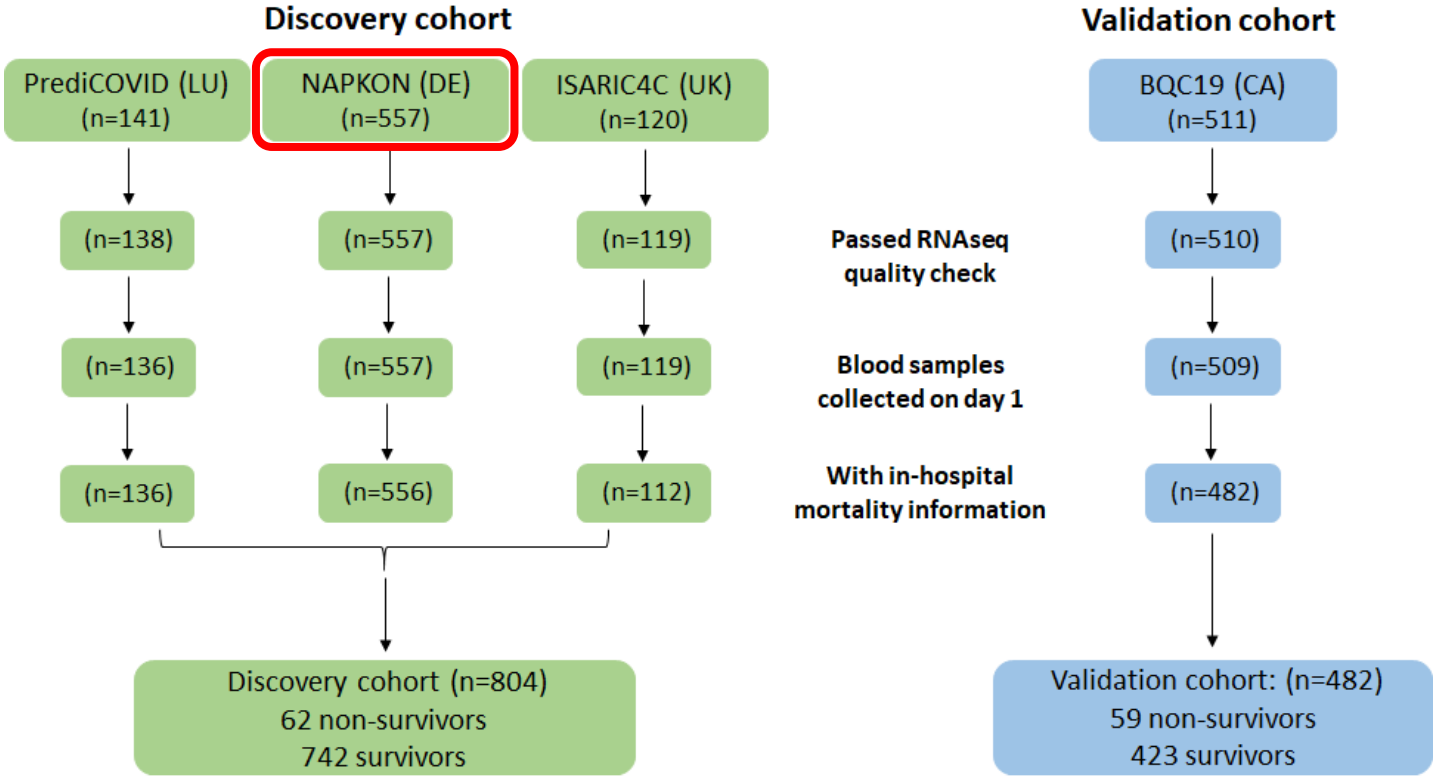
Accepted: 3 April 2024

Published online: 20 May 2024

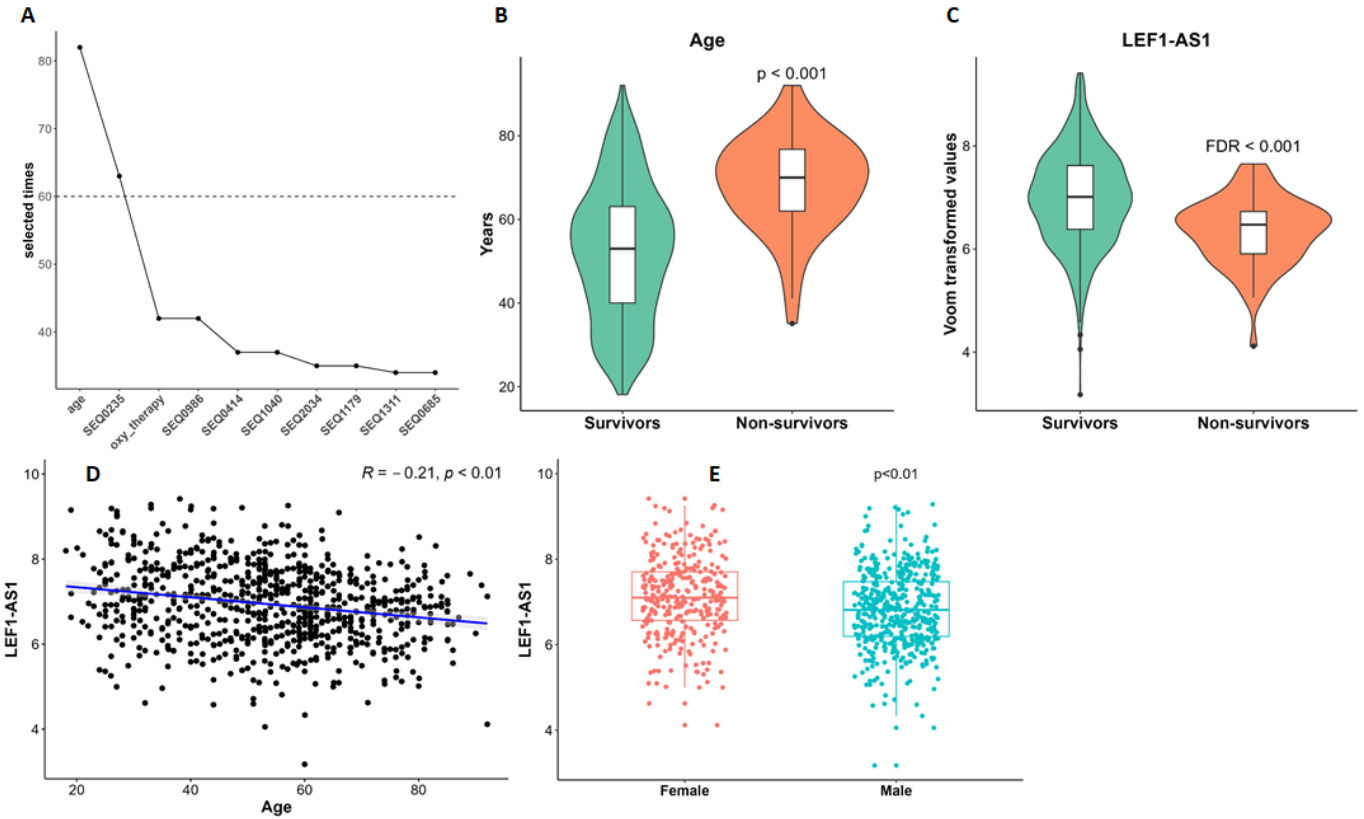
Check for updates

Yvan Devaux[1], Lu Zhang[2], Andrew I. Lumley[1], Kanita Karaduzovic-Hadziabdic[3], Vincent Mooser[4], Simon Rousseau[5], Muhammad Shoaib[6], Venkata Satagopam[6], Muhamed Adilovic[3], Prashant Kumar Srivastava[7], Costanza Emanueli[7], Fabio Martelli[8], Simona Greco[8], Lina Badimon[9], Teresa Padro[9], Mitja Lustrek[10], Markus Scholz[11], Maciej Rosolowski[11], Marko Jordan[10], Timo Brandenburger[12], Bettina Benczik[13], Bence Agg[13], Peter Ferdinandy[13], Jörg Janne Vehreschild[14,15,16,17], Bettina Lorenz-Depiereux[18], Marcus Dörr[19], Oliver Witzke[20], Gabriel Sanchez[21], Seval Kul[21], Andy H. Baker[22,23], Guy Fagherazzi[24], Markus Ollert[25,26], Ryan Wereski[27], Nicholas L. Mills[27,28] & Hüseyin Firat[21]

Study design



Feature selection (discovery cohort)

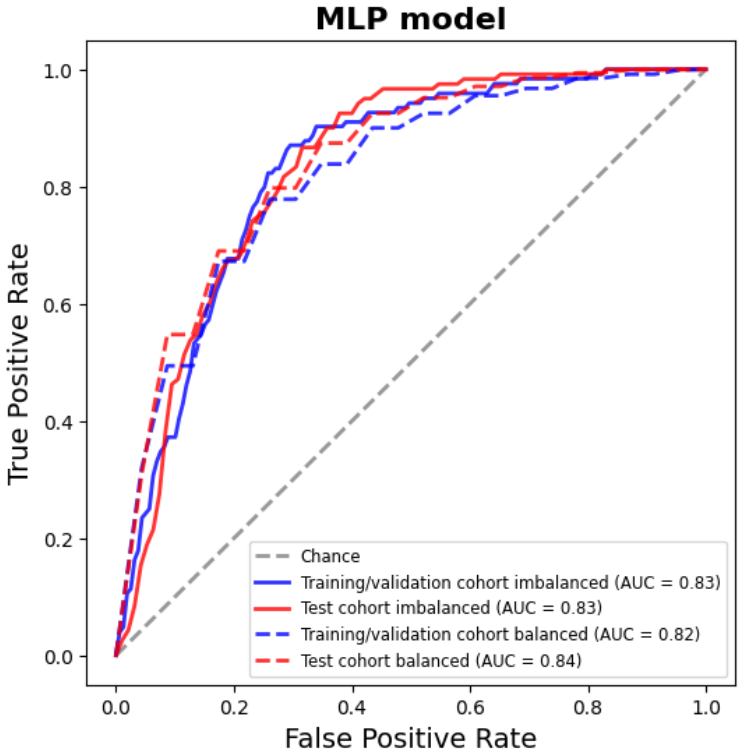


Machine learning models for in-hospital mortality

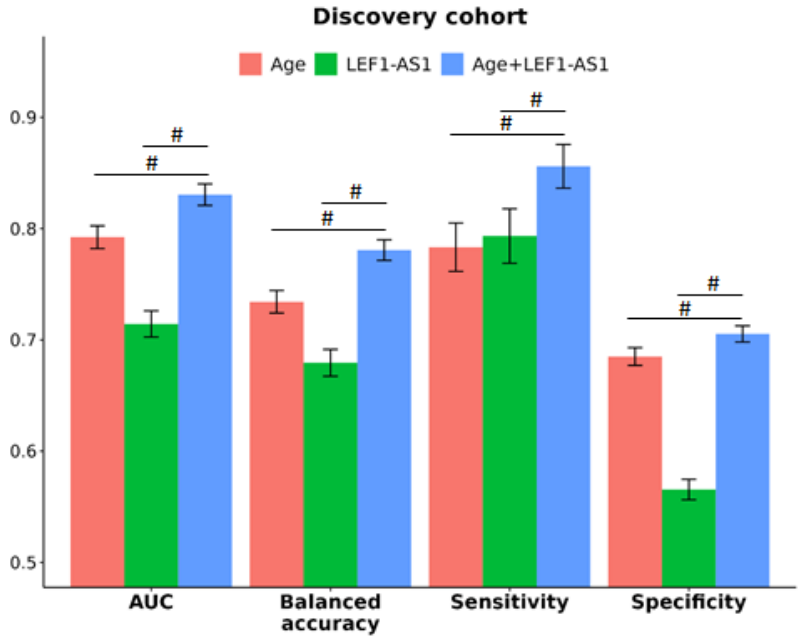
Classifier	AUC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Brier score (95% CI)
RF	0.78 (0.76–0.80)	0.73 (0.71–0.75)	0.76 (0.73–0.78)	0.71 (0.68–0.74)	0.19 (0.18–0.20)
kNN	0.81 (0.79–0.83)	0.75 (0.73–0.77)	0.85 (0.83–0.87)	0.65 (0.62–0.68)	0.18 (0.17–0.19)
Logit	0.81 (0.79–0.83)	0.76 (0.74–0.77)	0.81 (0.78–0.84)	0.70 (0.67–0.73)	0.18 (0.17–0.19)
MLP	0.82 (0.80–0.84)	0.77 (0.75–0.79)	0.82 (0.80–0.84)	0.72 (0.69–0.75)	0.18 (0.17–0.18)
SVM	0.67 (0.62–0.72)	0.74 (0.72–0.76)	0.82 (0.80–0.84)	0.67 (0.63–0.70)	0.21 (0.20–0.22)
XGB	0.74 (0.72–0.76)	0.68 (0.66–0.70)	0.69 (0.66–0.71)	0.67 (0.64–0.70)	0.25 (0.24–0.26)

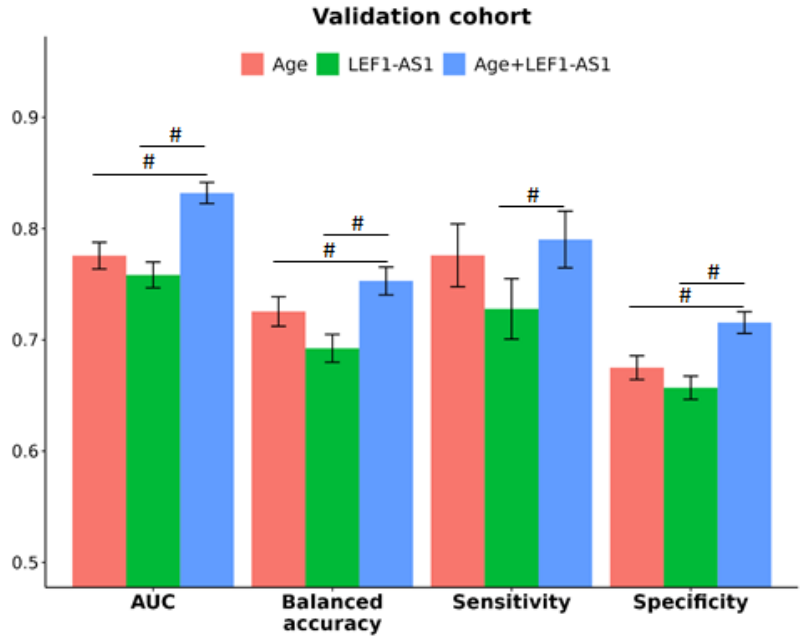
Performance of different classifiers to predict in-hospital mortality in the discovery cohort using two selected features (age and LEF1-AS1)

MLP model performance

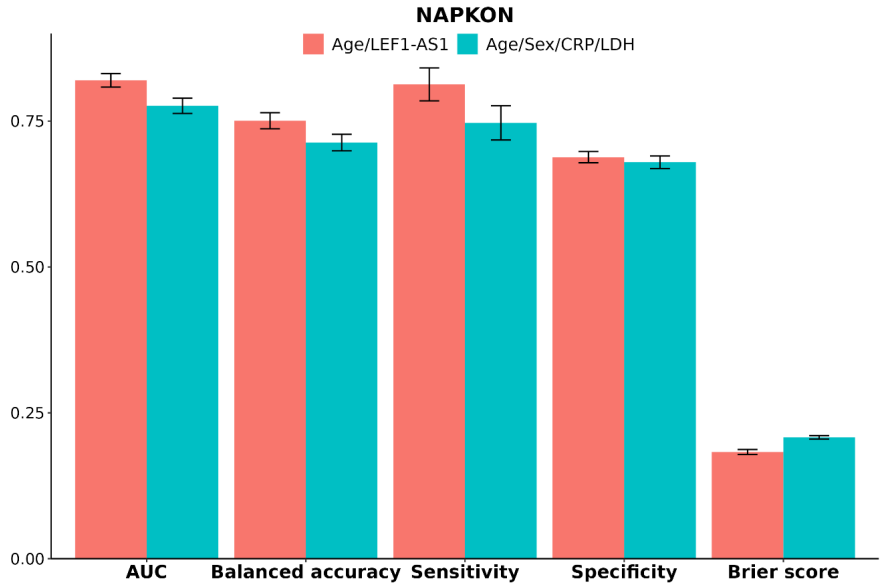


MLP model performance



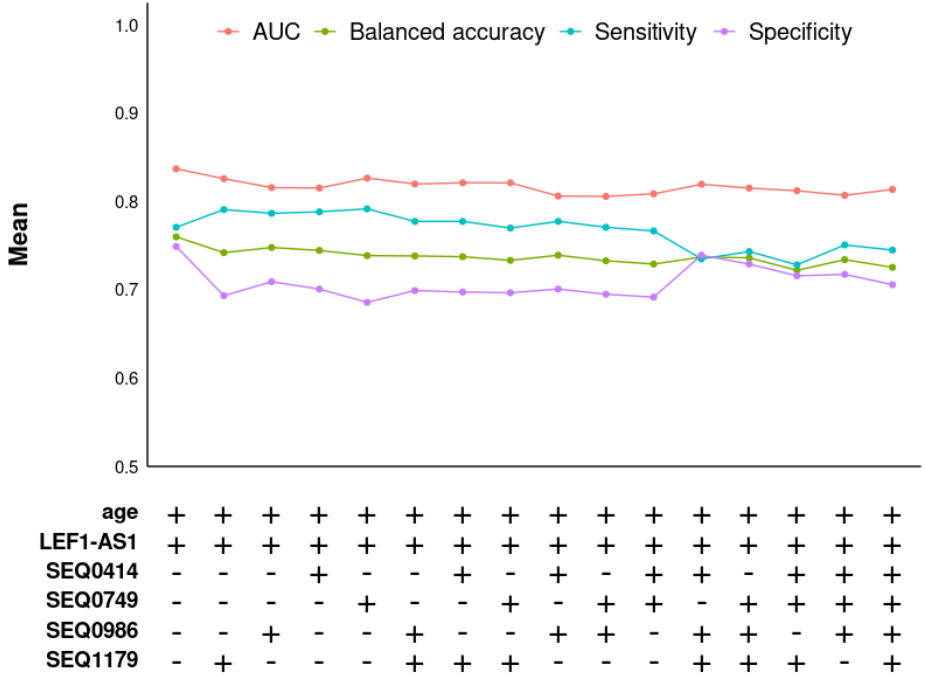


MLP model performance



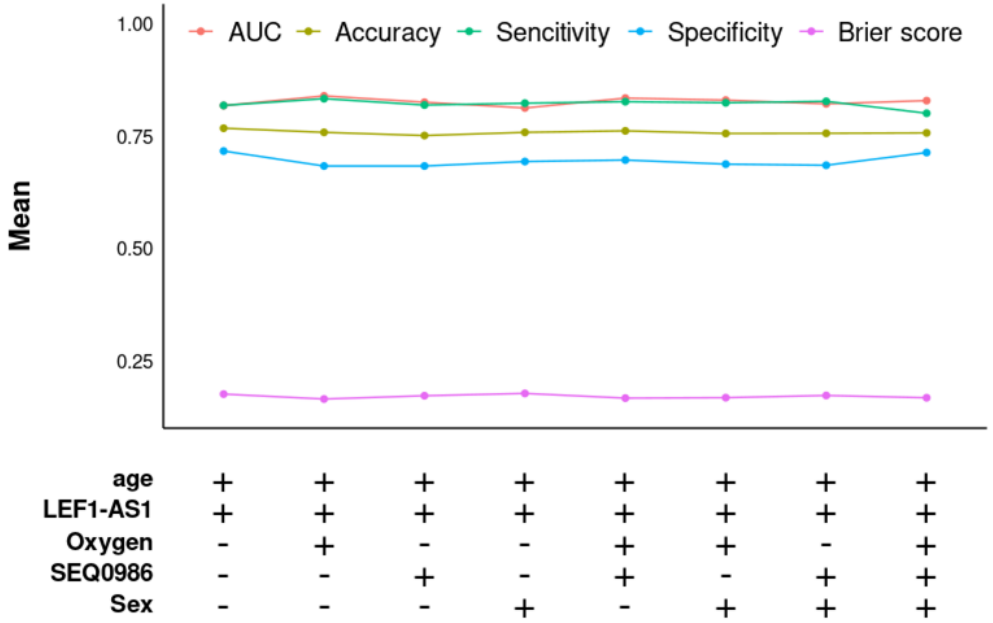
Comparison of our model with age and LEF1-AS1 with a previously published model with age, sex, CRP and LDH

MLP model performance



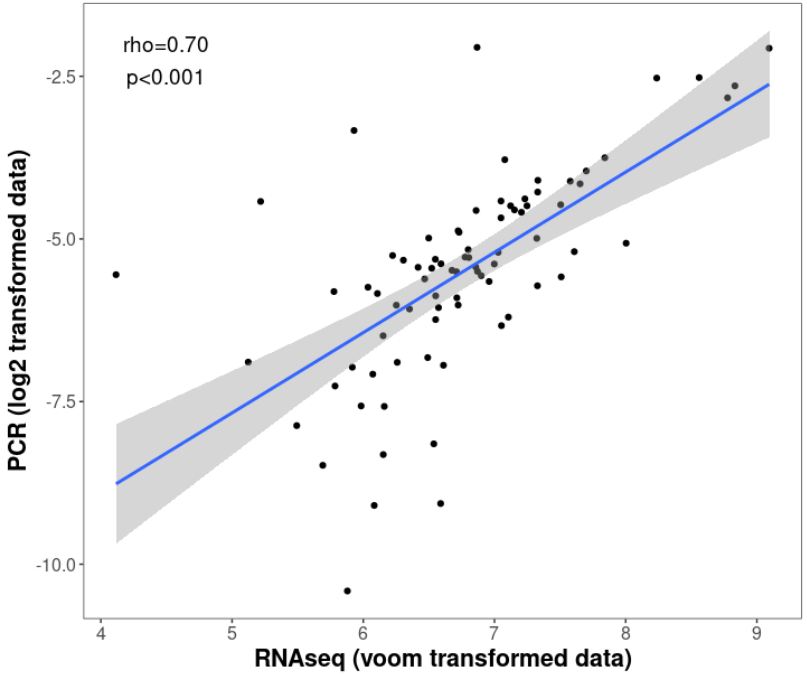
Lack of added value of other lncRNAs to the model with age and LEF1-AS1

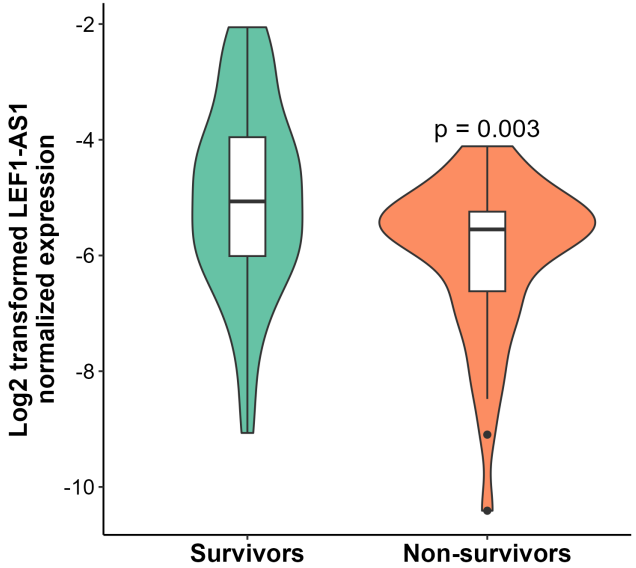
MLP model performance



Lack of added value of other data to the model with age and LEF1-AS1

PCR validation

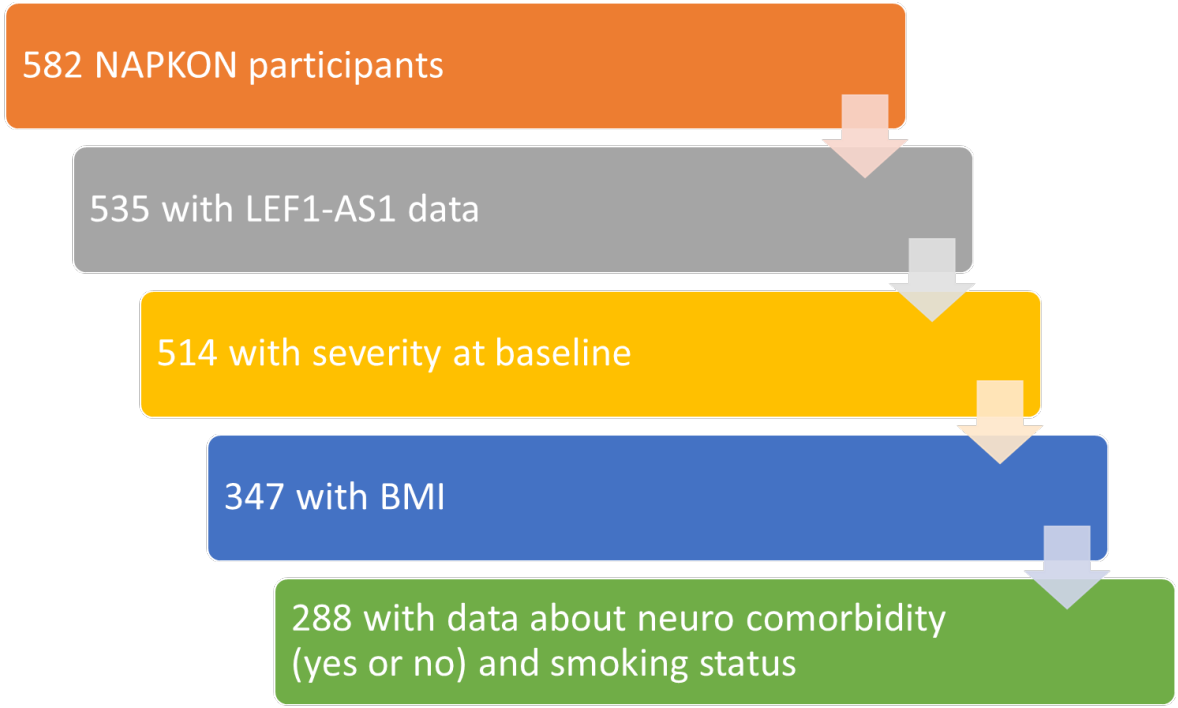




LEF1-AS1 expression in 84 PAXgene samples from the NAPKON cohort. LEF1-AS1 expression assessed by PCR was correlated with RNAseq data and was higher in survivors (n= 41) vs. non-survivors (n=43).

Study on long COVID

No imputation



Males and females

number participants	288			347			514			535		
male	178 (61%)			209 (60%)			308 (60%)			320 (60%)		
female	110			138			206			215		
CVD complication	142 (49%)			166 (48%)			252 (49%)			264 (49%)		
Univariate analysis for CVD	AUC	OR	p-value	AUC	OR	p-value	AUC	OR	p-value	AUC	OR	p-value
Gender	0.526	0.801	0.361	0.523	0.825	0.382	0.516	0.878	0.471	0.515	0.886	0.491
Age	0.583	1.331	0.019	0.58	1.309	0.015	0.569	1.247	0.014	0.574	1.265	0.008
BMI	0.556	1.189	0.146	0.539	1.126	0.274						
Comorb_neuro	0.526	1.636	0.18									
Current_smoker	0.501	1.031	0.941									
Former_smoker	0.501	1.011	0.965									
severity_no_compl	0.523	0.825	0.429	0.519	0.849	0.46	0.53	0.766	0.152			
severity_compl	0.502	0.978	0.93	0.503	0.97	0.894	0.515	1.161	0.442			
severity_severe_compl	0.525	2.981	0.067	0.523	2.478	0.073	0.515	1.823	0.142			
LEF1_AS1	0.567	0.763	0.027	0.545	0.826	0.08	0.557	0.786	0.008	0.556	0.785	0.006

Age and LEF1-AS are significantly associated with development of long CVD symptoms (12 months), but AUC is <0.60

Males only

number MALE participants	178			209			308			320		
CVD complication	84 (47%)			96 (46%)			147 (48%)			154 (48%)		
Univariate analysis for CVD	AUC	OR	p-value	AUC	OR	p-value	AUC	OR	p-value	AUC	OR	p-value
Age	0.555	1.177	0.286	0.573	1.241	0.130	0.561	1.202	0.113	0.563	1.211	0.093
BMI	0.559	1.191	0.251	0.531	1.096	0.511						
Comorb_neuro	0.521	1.874	0.288									
Current_smoker	0.500	0.994	0.991									
Former_smoker	0.514	1.129	0.701									
severity_no_compl	0.510	1.089	0.779	0.505	1.041	0.887	0.527	0.795	0.329			
severity_compl	0.542	0.691	0.242	0.539	0.705	0.235	0.505	1.046	0.855			
severity_severe_compl	0.532	3.193	0.095	0.534	3.169	0.058	0.522	2.296	0.106			
LEF1_AS1	0.552	0.869	0.356	0.512	0.977	0.870	0.520	0.921	0.475	0.520	0.916	0.437

No significant predictor for long CVD symptoms in males (neither age nor LEF1-AS1)

Females only

number FEMALE participants	110			138			206			215		
CVD complication	58 (53%)			70 (51%)			105 (51%)			110 (51%)		
Univariate analysis for CVD	AUC	OR	p-value	AUC	OR	p-value	AUC	OR	p-value	AUC	OR	p-value
Age	0.626	1.615	0.021	0.586	1.404	0.057	0.580	1.311	0.059	0.588	1.344	0.036
BMI	0.515	1.111	0.586	0.522	1.107	0.553						
Comorb_neuro	0.526	1.380	0.505									
Current_smoker	0.505	1.132	0.859									
Former_smoker	0.518	0.850	0.690									
severity_no_compl	0.583	0.457	0.065	0.561	0.573	0.131	0.538	0.693	0.234			
severity_compl	0.567	1.962	0.124	0.554	1.696	0.166	0.534	1.436	0.263			
severity_severe_compl	0.516	2.782	0.382	0.507	1.478	0.674	0.504	1.212	0.779			
LEF1_AS1	0.594	0.618	0.023	0.598	0.638	0.016	0.613	0.615	0.002	0.612	0.618	0.001

Age and LEF1-AS1 are predictors of CVD complications in females

Multivariable analysis

n=288 – Females only (n=110)

variables	AIC	wald_p	AUC
LEF1_AS1 + Age	146.836	0.010	0.659
LEF1_AS1 + Age + severity_severe_compl	148.054	0.020	0.671
LEF1_AS1 + Age + severity_compl	148.149	0.022	0.663
Age	150.370	0.023	0.626
LEF1_AS1 + Age + BMI	148.493	0.023	0.664
LEF1_AS1	150.497	0.025	0.594
LEF1_AS1 + Age + Comorb_neuro	148.681	0.025	0.660
LEF1_AS1 + Age + Former_smoker	148.713	0.026	0.664
LEF1_AS1 + Age + Current_smoker	148.814	0.026	0.660
LEF1_AS1 + Age + severity_compl + severity_severe_compl	149.134	0.034	0.675
LEF1_AS1 + Age + BMI + severity_severe_compl	149.587	0.036	0.679
LEF1_AS1 + Age + BMI + severity_compl	149.893	0.041	0.671
LEF1_AS1 + Age + Current_smoker + severity_severe_compl	149.990	0.042	0.669
LEF1_AS1 + Age + Comorb_neuro + severity_severe_compl	149.992	0.042	0.669
LEF1_AS1 + Age + Former_smoker + severity_severe_compl	149.936	0.043	0.674
LEF1_AS1 + Age + Comorb_neuro + severity_compl	149.900	0.043	0.666
LEF1_AS1 + severity_compl	150.898	0.043	0.627
Age + severity_compl	151.137	0.044	0.635
LEF1_AS1 + Age + Current_smoker + severity_compl	150.060	0.045	0.666
LEF1_AS1 + Age + Former_smoker + severity_compl	150.049	0.045	0.661
LEF1_AS1 + Age + BMI + Former_smoker	150.304	0.047	0.677
LEF1_AS1 + Age + BMI + Comorb_neuro	150.378	0.047	0.664
LEF1_AS1 + Age + BMI + Current_smoker	150.449	0.048	0.664
LEF1_AS1 + severity_severe_compl	151.167	0.048	0.609

Age + LEF1-AS1 is the best model to predict CVD in females, but AUC is low (0.66)

Conclusions

- NAPKON study helped to identify LEF1-AS1 as a predictor of in-hospital mortality
- In NAPKON, the capacity of baseline levels of LEF1-AS1 to predict the development of cardiovascular symptoms at 12 months is low (AUC < 0.60)

Acknowledgements

- NAPKON study investigators
- The clinical-scientific infrastructure of NAPKON (Nationales Pandemie Kohorten Netz, German National Pandemic Cohort Network) and NUKLEUS (NUM Klinische Epidemiologie- und Studienplattform, NUM Clinical Epidemiology and Study Platform) of the Network University Medicine (NUM)
- All study participants and NAPKON sites who contributed clinical data and/or biosamples for this analysis
- The NAPKON Steering Committee: University Hospital Giessen and Marburg, Giessen (Herold S), University of Wuerzburg, Wuerzburg (Heuschmann P), Charité - Universitaetsmedizin Berlin, Berlin (Heyder R), University Medicine Greifswald, Greifswald (Hoffmann W), Hannover Unified Biobank, Hannover Medical School, Hannover (Illig T), University Hospital Schleswig-Holstein, Kiel (Schreiber S), University Hospital Cologne and University Hospital Frankfurt, Cologne and Frankfurt (Vehreschild JJ), Charité - Universitaetsmedizin Berlin, Berlin (Witzenrath M).

Acknowledgements

- Founders: ACBB, the Augsburg Central BioBank (www.biobank-augsburg.de), the CCS Biobank at the University Heart and Vascular Center Hamburg (https://www.uke.de/kliniken-institute/kliniken/kardiologie/forschung/), the Central Biobank Erlangen as a core unit of the Friedrich-Alexander-University Erlangen-Nürnberg in cooperation with the Department of Medicine 1, University Hospital Erlangen, the Central Biobank UMG as a core facility of the University Medical Center Goettingen (Germany), the HOM.BMB (Biobank Internal Medicine V, Saarland University, Homburg), the Institute of Clinical Chemistry and Laboratory Medicine - Integrated Research Biobank, University Medicine Greifswald, the RWTH centralized Biomaterial Bank (RWTH cBMB) of the Medical Faculty of RWTH Aachen University (https://www.cbmb.ukaachen.de/), the West German Biobank Essen (https://www.uni-due.de/med/biobank/), the ZBR, the Central Biobank Regensburg, the iBioTUM, the Central Interdisciplinary Biomaterial Bank as a Core Unit of the TUM School of Medicine and the University Hospital of the Technical University of Munich, the interdisciplinary Biobank and Database Frankfurt (https://ibdf-frankfurt.de/).

European Commission

Horizon 2020
European Union funding
for Research & Innovation

Thank you