Multi-omic profiling of tumor-derived extracellular vesicles enables early cancer detection

Margie Kinnersley, Tre Blohm, Tia Seibold, Rachel Short-Miller, Sean Lodmell, Amanda Mast, Riley Kemp, Kelley Van Vaerenberghe, Claire Seibold, and Katie Havranek

FYR Diagnostics, Missoula, MT USA

8049T

Introduction

- Lung Cancer is the leading cause of cancer-related death in the U.S., with fatalities exceeding those of breast, colorectal, and prostate cancers combined[1].
- Non-small cell lung cancer (NSCLC), the most prevalent type, has a 5-year survival rate of just 25%[2].
  - Lung adenocarcinoma (LUAD) is a form of NSCLC that originates in the periphery of the lung and accounts for 40% of all lung cancers[1,3].
  - Squamous cell carcinoma of the lung (LUSC) is a type of NSCLC that originates in the central part of the lung and accounts for 30% of all lung cancers[1,3].
- Early detection can boost NSCLC survival rates by 36%, yet the recommended screening method, low-dose CT, lacks specificity, provides no comprehensive prognostic information and is inaccessible to many patients[4,5].
- Liquid biopsies based on circulating DNA have emerged as convenient, cost-effective alternatives to traditional screening but often lack sensitivity and specificity for early-stage cancers[6,7].
- Blood plasma extracellular vesicles (EVs), which carry heterogeneous protein, nucleic acid, and metabolite cargos derived from various cell types including cancer cells, provide a richer diagnostic analyte base than cfDNA.
- Here we present a novel tumor-derived EV (TDEV) enrichment technique called SPARCs™, applied to plasma from early-stage NSCLC patients.

Objective

- The objective of this study was to identify RNA and protein biomarkers in enriched TDEVs from cancer vs. healthy donors to advance early screening capabilities for NSCLC.

Materials and Methods

- Plasma was processed from whole blood collected in Streck Cell-Free BCT preservative (Streck, La Vista, Nebraska) and stored at -80°C.
- For EV isolation, plasma was thawed, re-spun to clear debris and subjected to ion-exchange chromatography. Purified EVs were then characterized in concordance with the Minimal Information for Studies of Extracellular Vesicles (MISEV) 2023 guidelines with respect to particle concentration and size (Zetaview nanoparticle tracking system, ParticleMetrix, Ammersee, Germany), and presence/absence of category 1,2 and 3 protein markers (Jess automated western blot system, Biotechne, Minneapolis, MN).
- EVs were incubated with Tumor SPARCs™ to enrich tumor-derived EVs.
- Purified RNA was used to generate bulk RNAseq libraries and sequenced on an Element Biosciences AVITI system (San Diego, CA).
- SPARC-enriched EVs were subjected to digestion and subsequent LC-MS/MS on the Orbitrap Astral Instrument (Thermo Fisher Scientific, Waltham, MA) at Cedars Sinai Precision Biomarkers Laboratories (Beverly Hills, CA) using data independent acquisition.
- 10-fold cross-validation was used for Machine Learning model training, and the best model was selected based on AUC.
- Candidate biomarkers were selected based on leading model AUC.

EV Extraction and Characterization

- EV extractions yielded an average of 1.2x10[11] particles /mL plasma with a mean diameter of 155.3nm (Fig. 1).
- There was no significant difference in concentration or size between NSCLC and healthy donor particles.

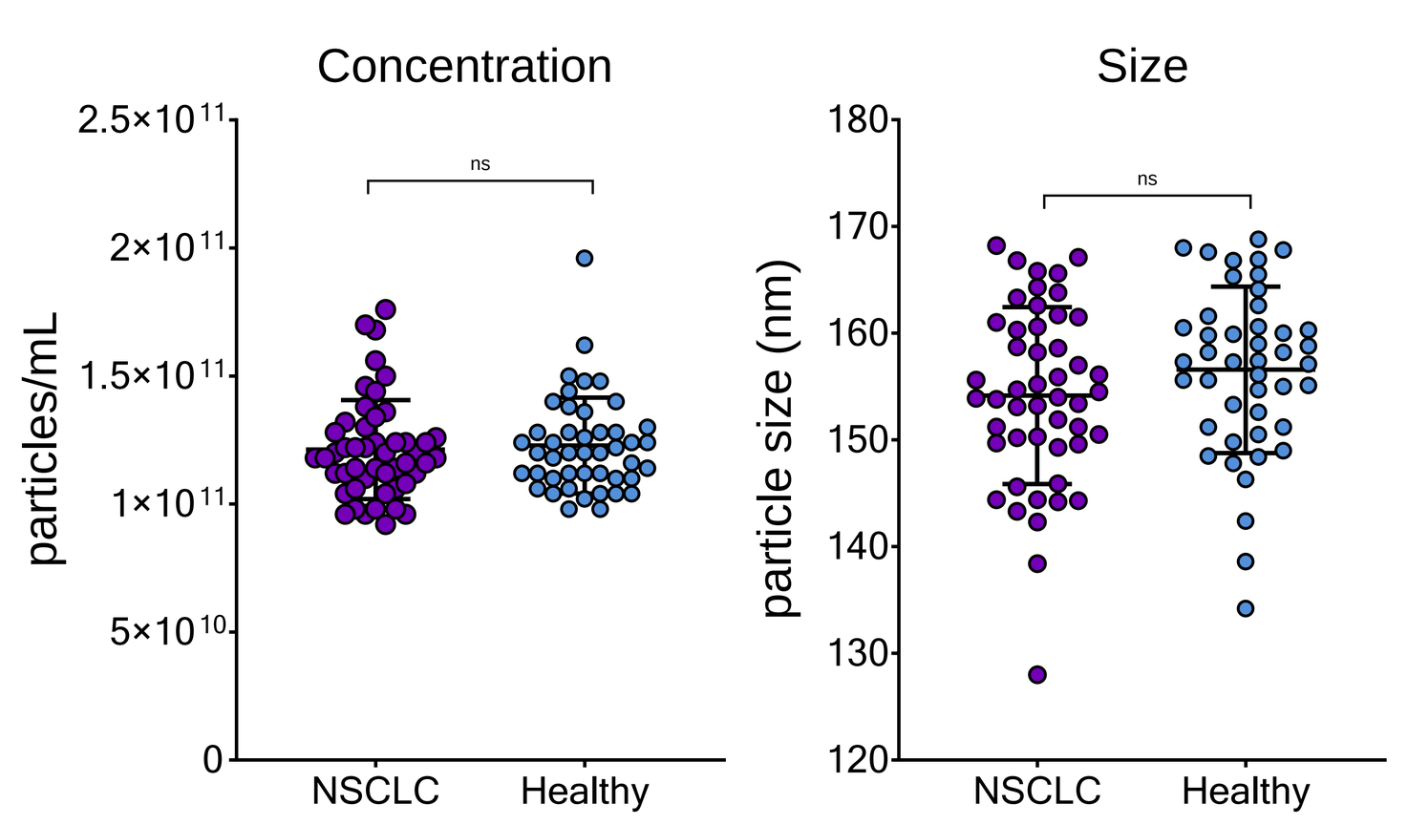

Fig. 1. Zetaview particle analysis, Kruskal-Wallis test p<0.05

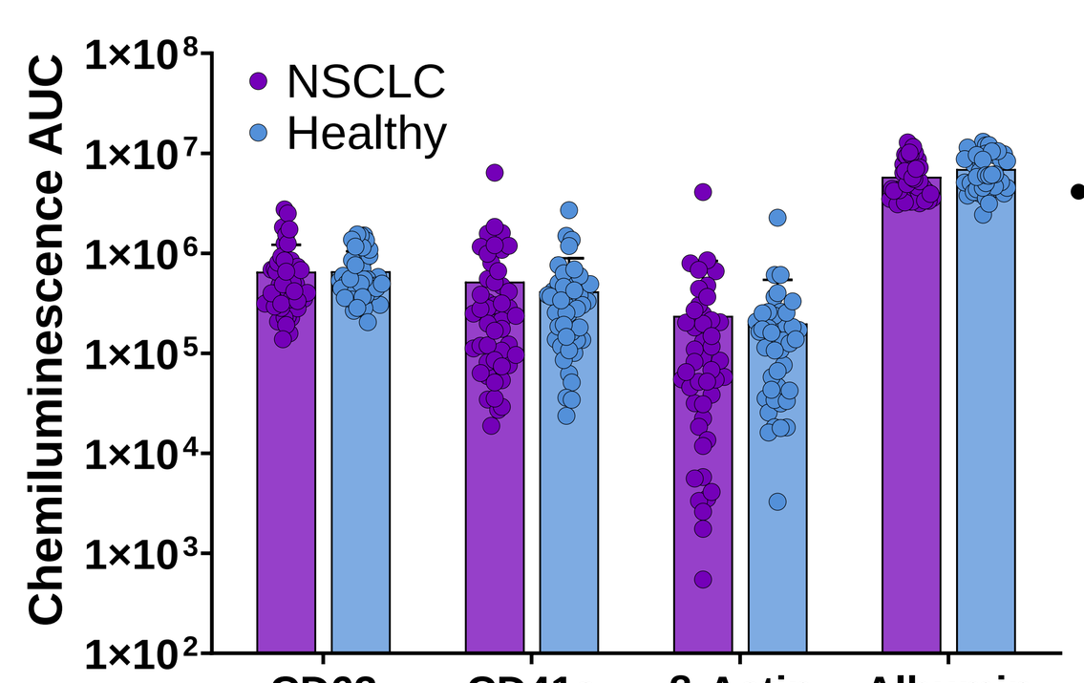

Fig. 2. Jess Automated Western Blot analysis

- Particle preparations contained MISEV marker proteins CD63 (category 1a), CD41a (category 1b), β-Actin (category 2b), and Albumin (category 3a), confirming the presence of extracellular vesicles (Fig. 2).

Study Design

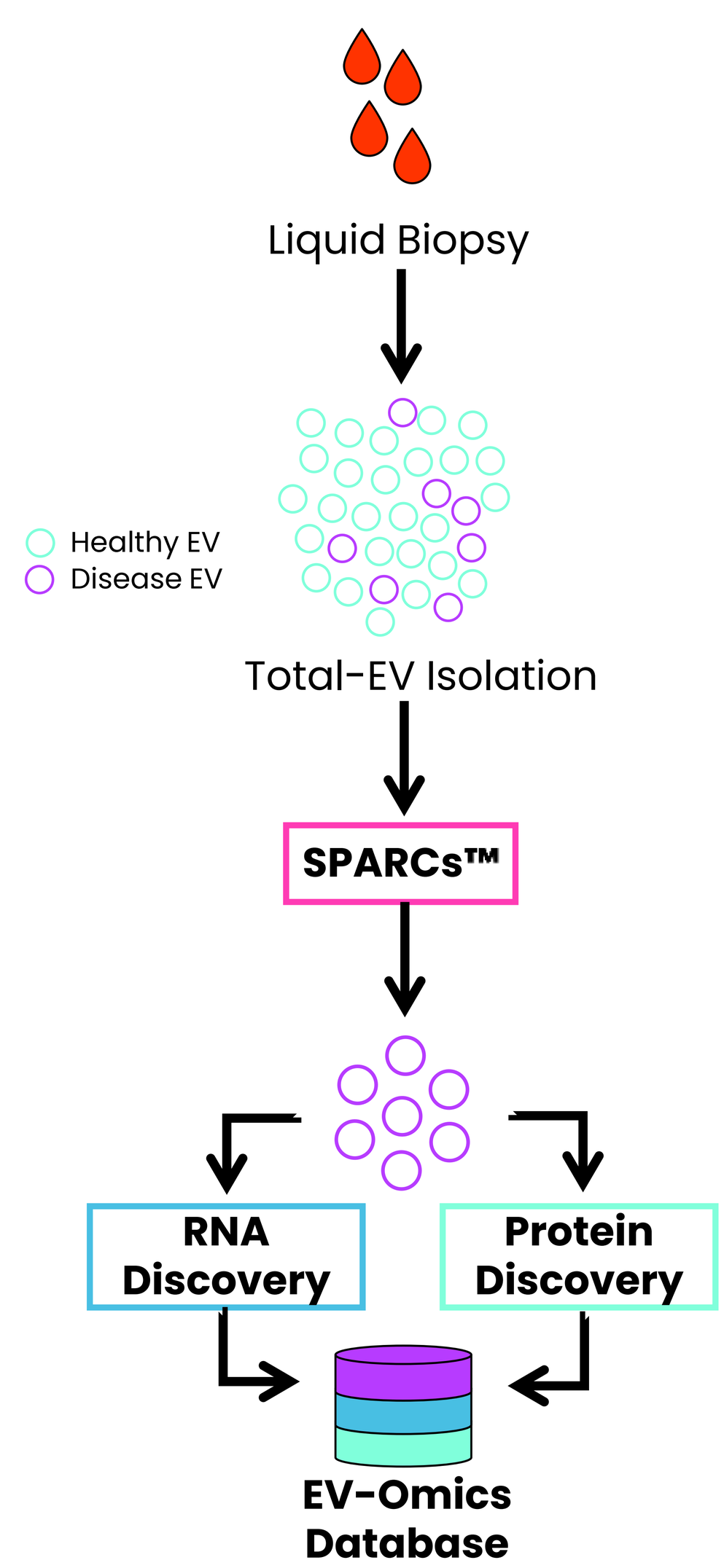

Table 1. Samples used in this study

	NSCLC	Healthy
N	49	45
Age		
<51	4	5
51-60	11	13
61-70	21	23
71-80	12	4
81-90	1	0
>90	0	0
Sex		
Male	28	26
Female	21	19
Race/Ethnicity		
Caucasian	49	37
Asian	0	2
Black	0	6
Staging		
Stage 1 (I, IA1, IA2, IA3, IB)	25	-
Stage 2 (IIA, IIB)	24	-
Subtype		
Adenocarcinoma (LUAD)	37	-
Squamous Cell Carcinoma (LUSC)	10	-
Adenosquamous	2	-
Specimen Providers	2	2
Smoking Status		
Unknown	33	8
Never used	4	23
Previous use	8	8
Current cigarette	4	4
Current tobacco	-	2

Proteomics

- Differential expression analysis identified 103 proteins significantly enriched and 49 proteins significantly depleted in NSCLC TDEVs relative to healthy controls at p_{adj}≤0.05 and log_2 fold-change of ±1 (Fig. 5).

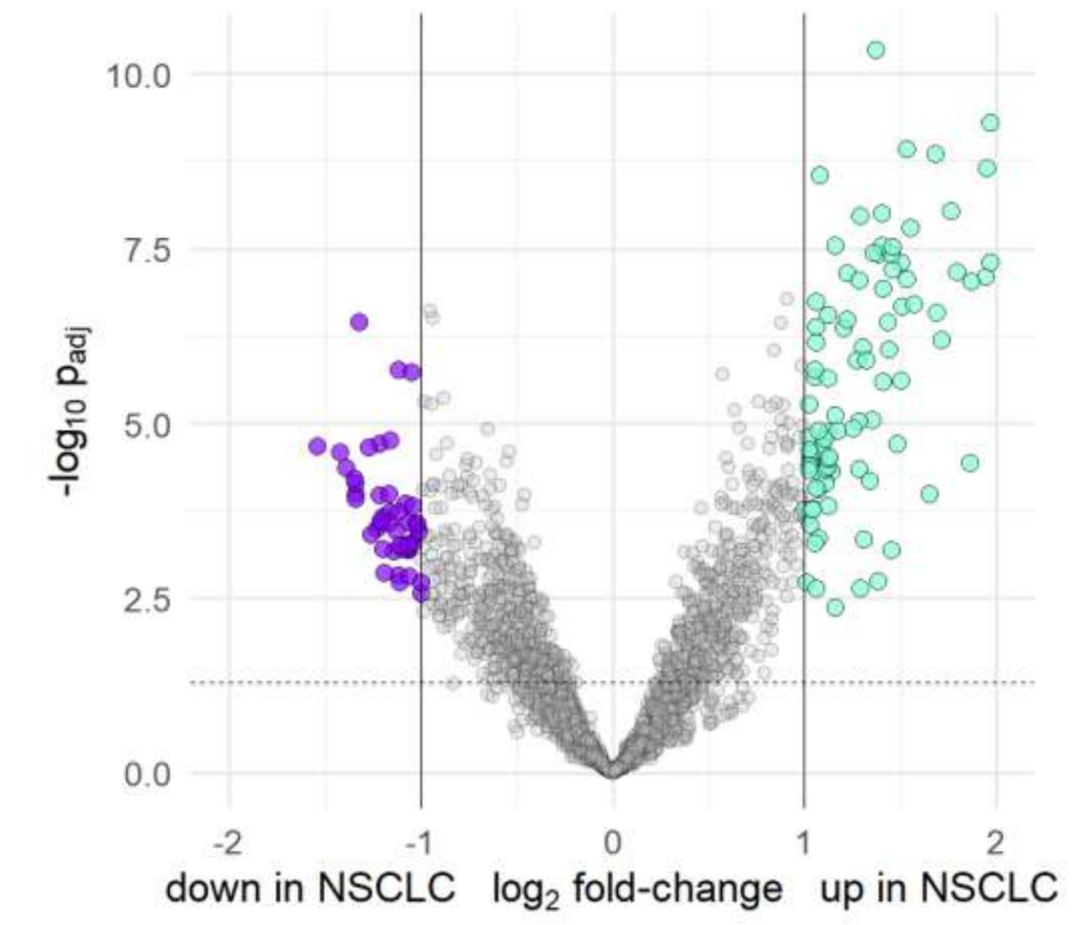

Fig. 5. Volcano plot of NSCLC vs. Healthy TDEV protein abundance.

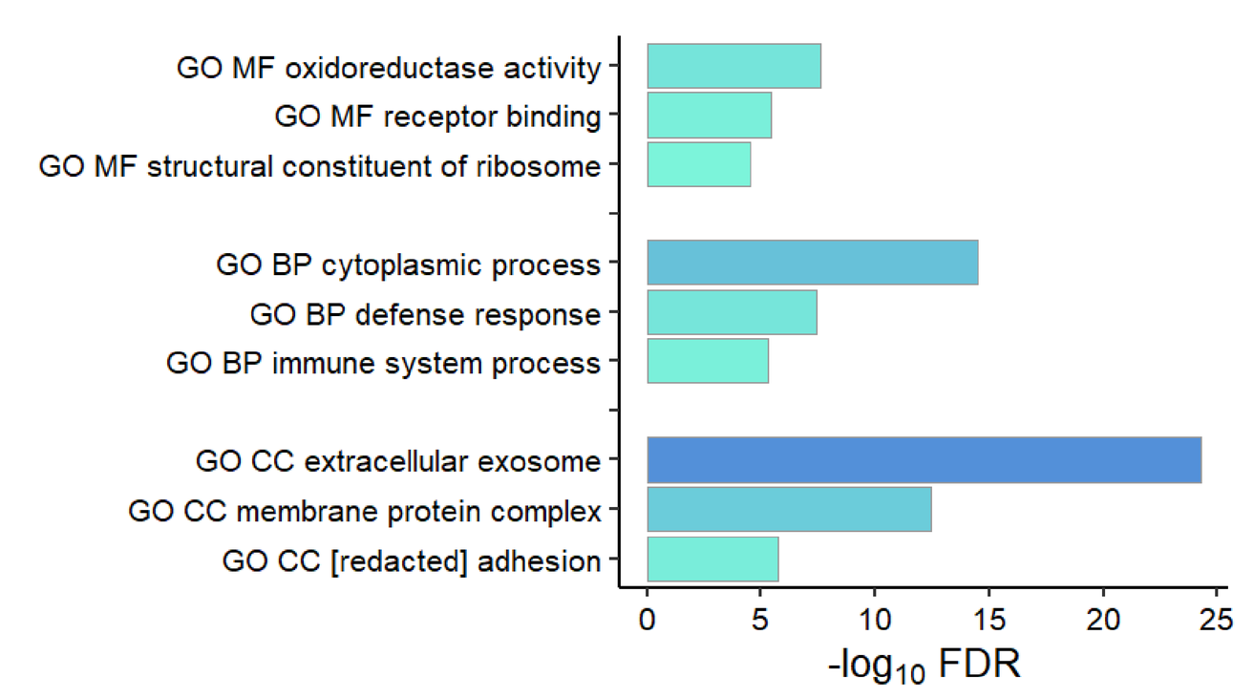

Fig. 6. NSCLC v healthy GO term enrichment

- Enriched or depleted proteins belonged to GO categories related to receptor binding, immune system processes and the exosomal compartment, consistent with tumor growth and spread (Fig. 6).

- 437 proteins distinguished LUAD and LUSC subtypes from one another and from healthy controls (Fig. 7, 8).
  - LUAD-specific proteins belonged to GO functional categories like cytoplasmic translation and were associated with extracellular exosomes.
  - LUSC-specific proteins belonged to GO biological processes like respiration and adhesion, and were associated with the cytoplasm.

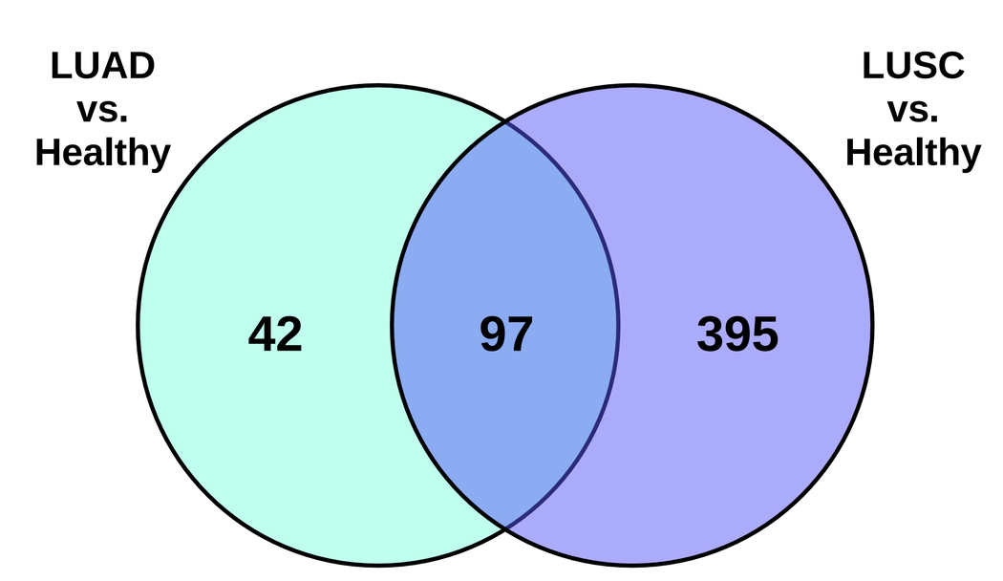

Fig. 7. Intersection of LUAD vs Healthy & LUSC vs. Healthy protein sets

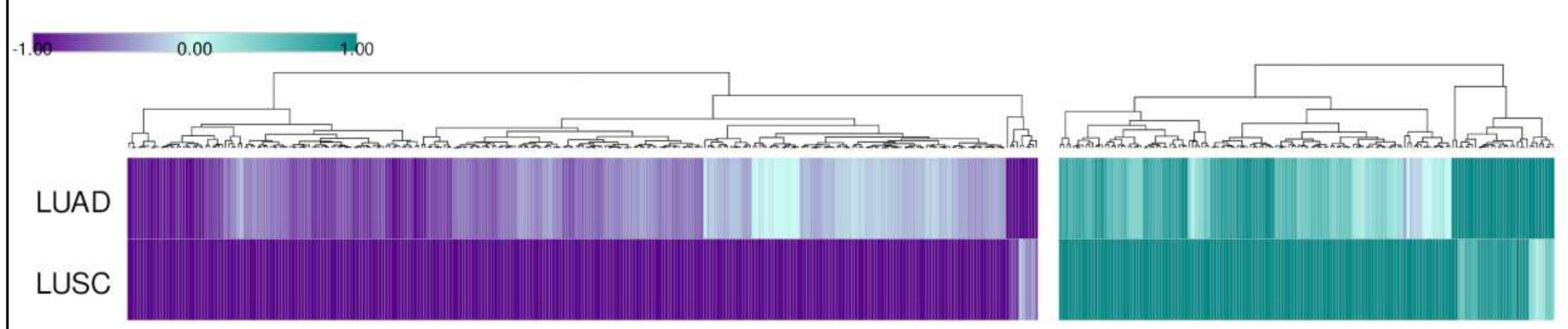

Fig. 8. Heatmap of log_2 relative abundance of proteins unique to LUAD and LUSC

RNA extraction, sequencing and LC-MS/MS

- Following EV isolation, SPARCs™ enrichment and RNA extraction, each mL of NSCLC input plasma yielded an average of 32 ng RNA, significantly more than healthy EVs and consistent with published observations that cancer cells over-secrete EVs (Fig. 3A)[8].

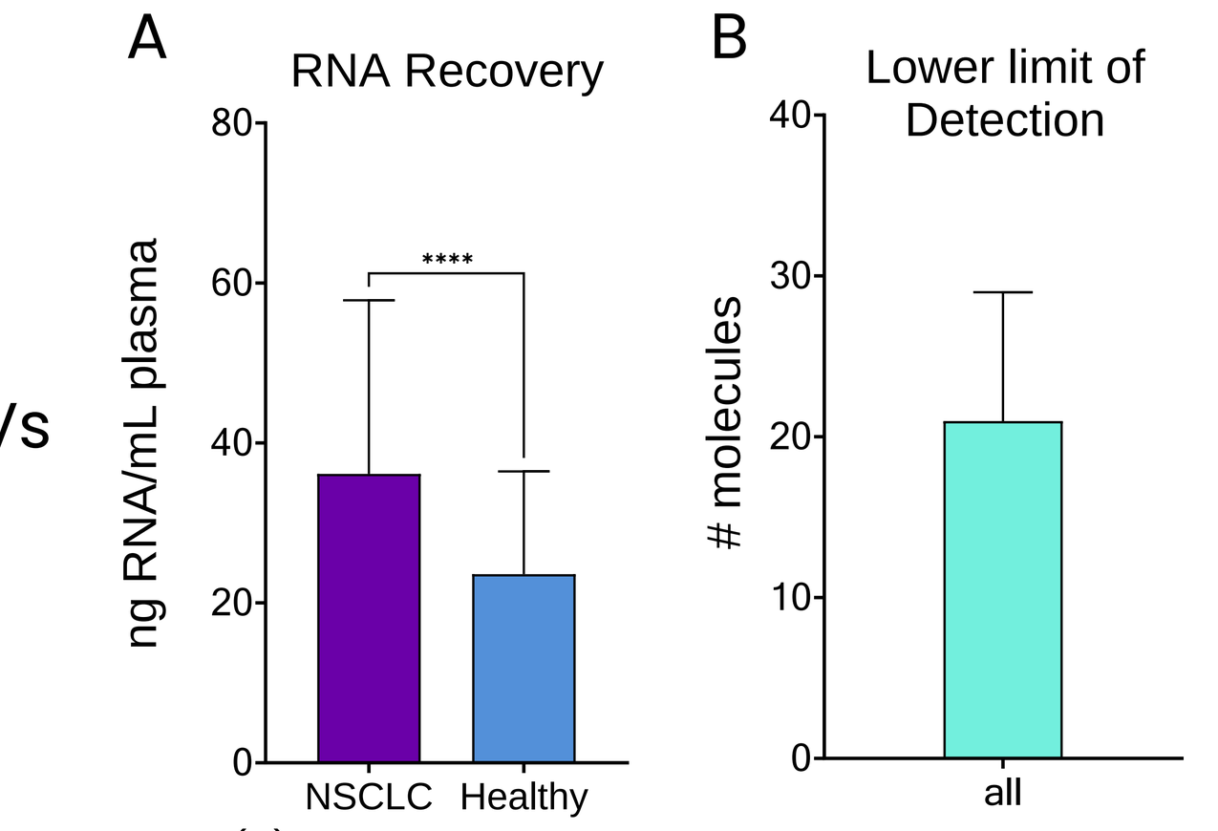

Fig. 3. (A) RNA recovery. Significance determined by Kruskal-Wallis test p≤0.05. (B) Sequencing lower limit of detection.

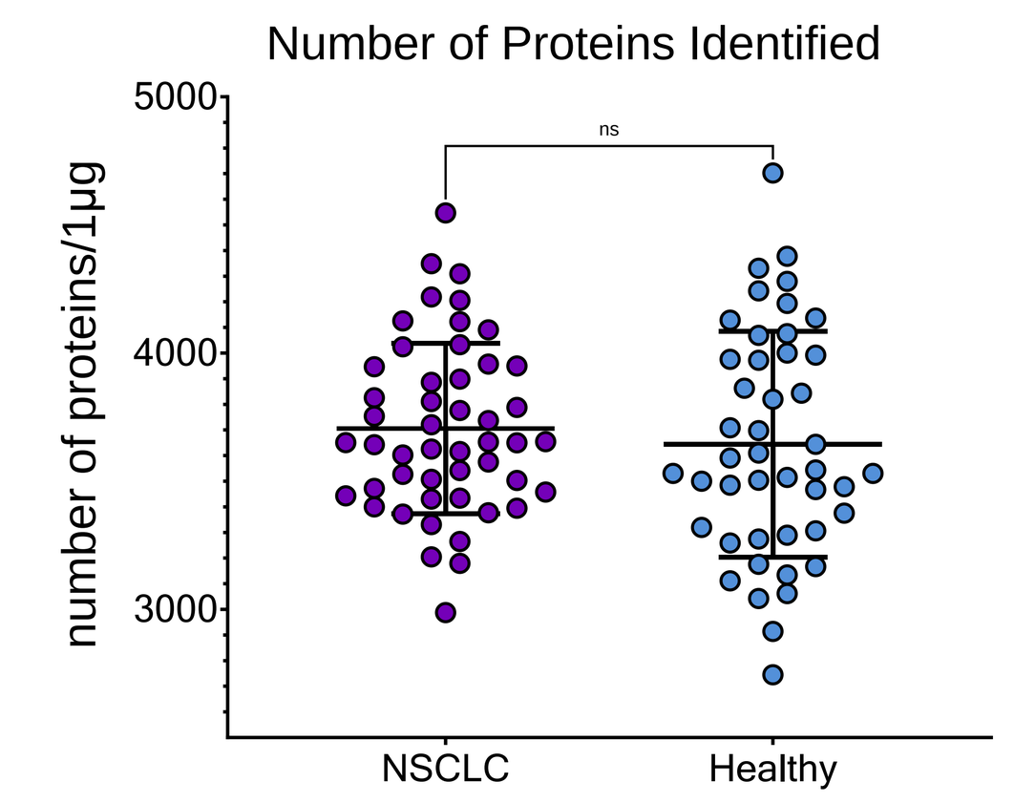

Fig. 4. LC-MS/MS protein identifications.

- RNAseq libraries were sequenced to a lower limit of detection of ≈ 21 molecules at RPKM =1 (Fig. 3B), with an average Q score of 42.
- LC-MS/MS identified an average of 3611 proteins per sample at 1 µg input (Fig. 4).

Transcriptomics

- Differential expression analysis identified 223 transcripts significantly enriched and 50 significantly depleted in NSCLC TDEVs relative to healthy controls at p_{adj}≤0.05 and log_2 fold-change of ±0.8 (Fig 9).

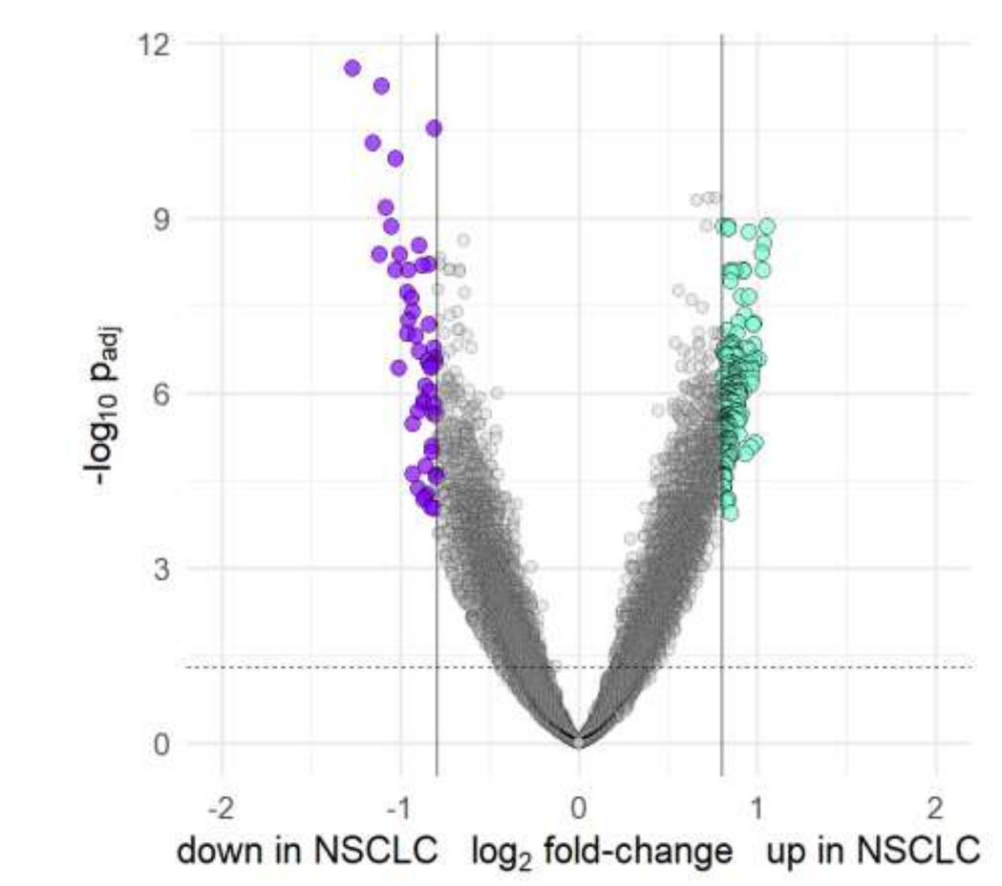

Fig. 9 Volcano plot of NSCLC vs. Healthy relative gene expression

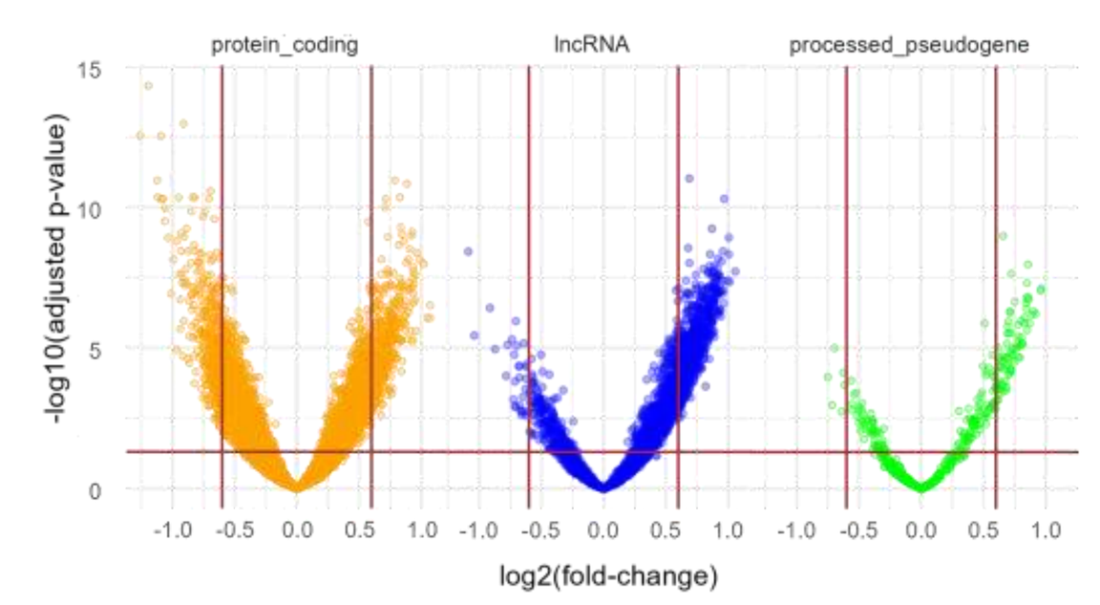

Fig. 10. Volcano plot of NSCLC vs. Healthy RNA biotypes

- Differentially packaged transcripts were from a variety of biotype classes including mRNAs, long non-coding RNAs and transposable element-associated pseudogenes (Fig. 10).

- Differentially packaged transcripts were from GO categories related to lung cancer progression like tumorigenesis-associated receptor activity and the inflammatory response (Fig. 11).

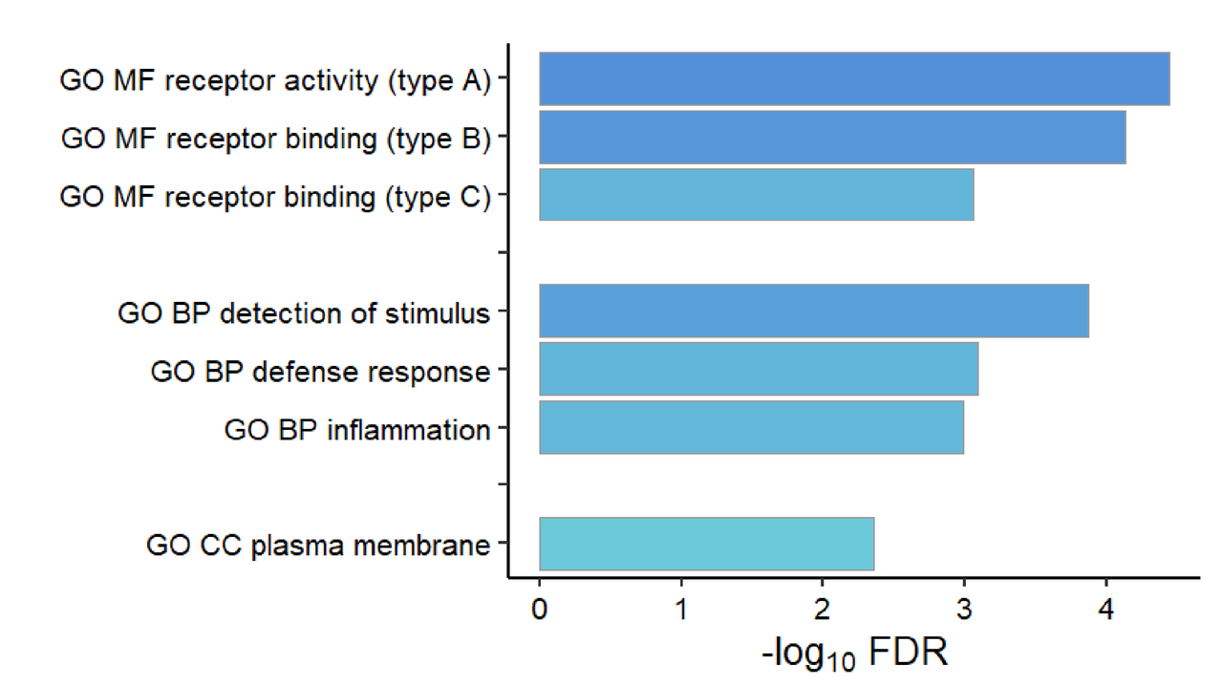

Fig. 11. NSCLC v healthy GO term enrichment

- 716 transcripts distinguished LUAD and LUSC from one another and from healthy controls (Fig. 12, 13).
  - LUAD-specific transcripts were from GO categories related to tumorigenesis-associated receptor activity and the extracellular space.
  - LUSC-specific transcripts belonged to functional categories like protein binding and the cytoplasmic compartment.

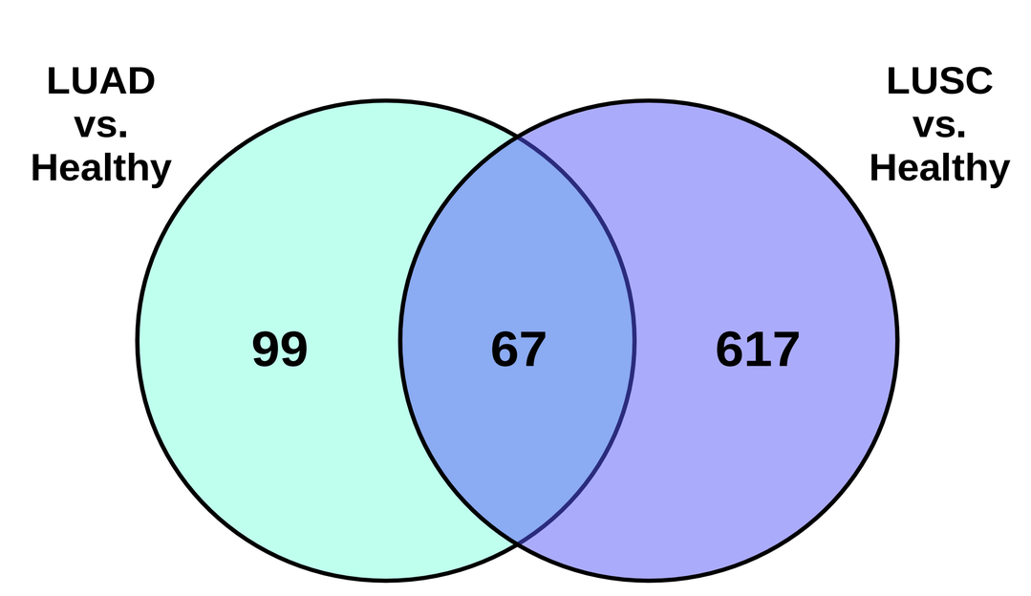

Fig. 12. Intersection of LUAD vs Healthy & LUSC vs. Healthy gene sets

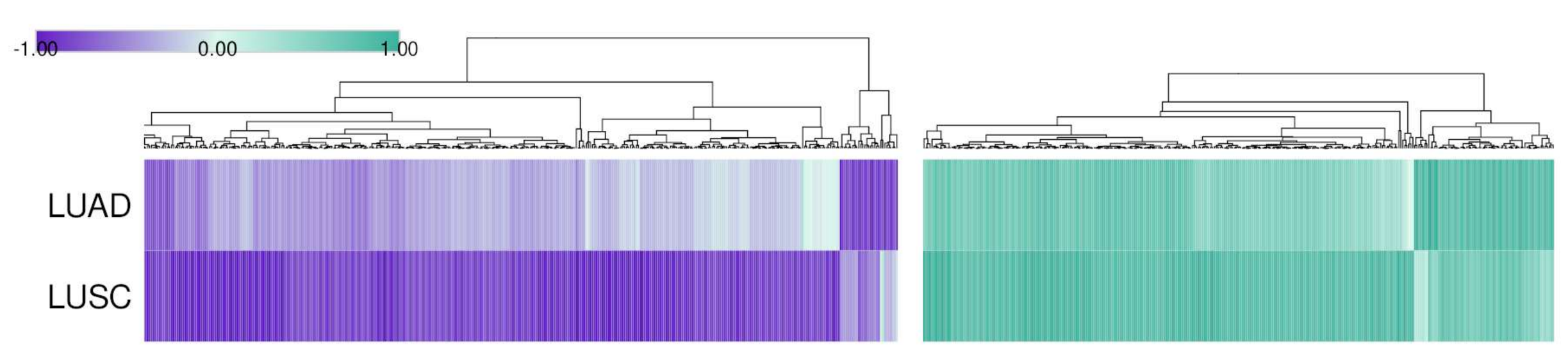

Fig. 13. Heatmap of log_2 relative expression of genes unique to LUAD and LUSC

Biomarker Discovery

- Supervised machine-learning algorithms were used to detect differences between NSCLC and Healthy samples.
- 10-fold cross-validation was used to calculate performance metrics (Table 2).
- The best model was selected based on AUC (Table 2, Fig. 14).

Table 2. Performance metrics for leading model: NSCLC v Healthy

	Protein	RNA
AUC	0.93	0.98
Combined Sensitivity	0.92	0.98
Stage I Sensitivity	0.92	0.96
Stage 2 Sensitivity	0.92	1
Specificity	0.87	0.98
No of Features	13	24

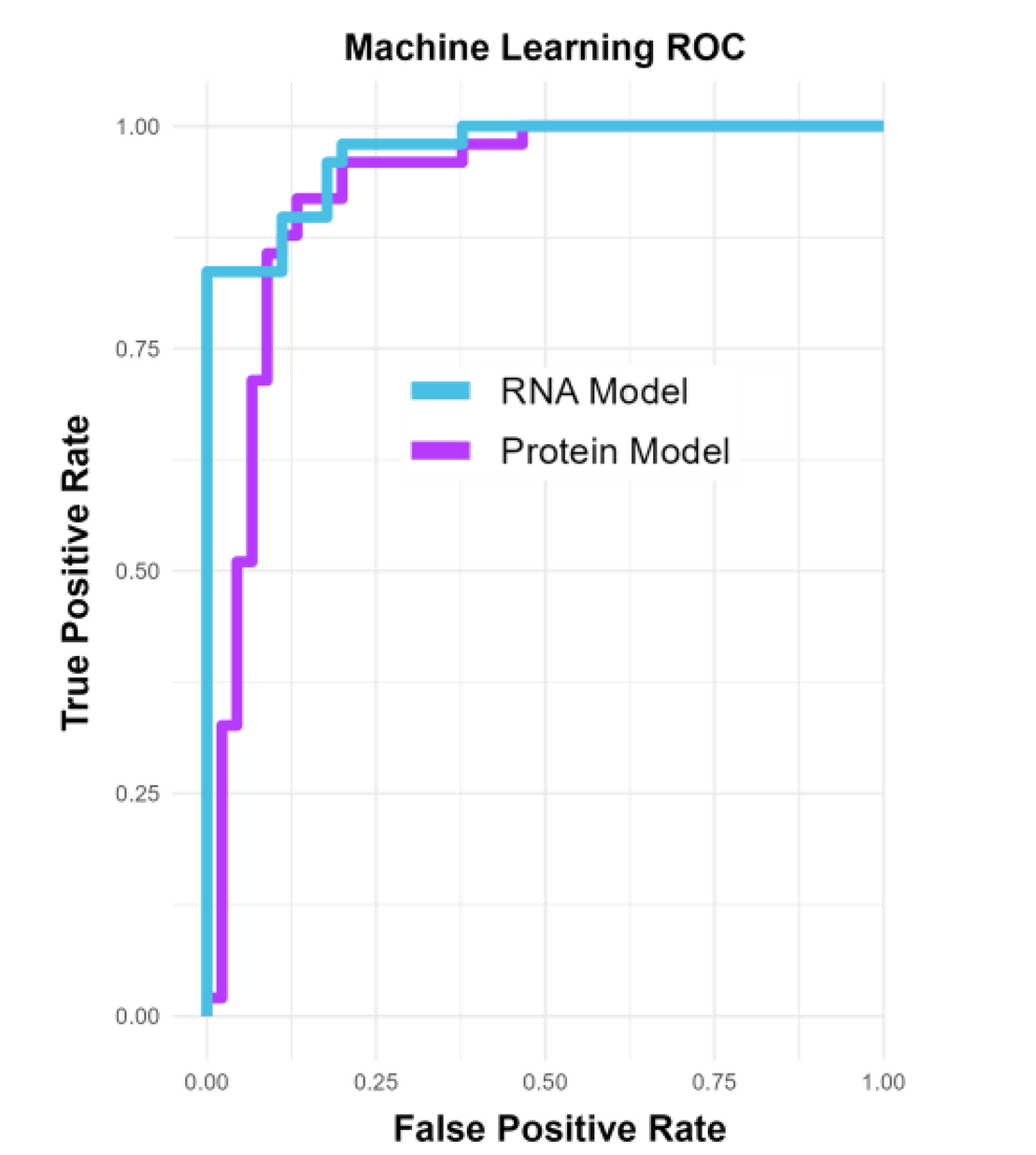

Figure 14. ROC curve for leading RNA and protein models

- Candidate biomarkers selected based on leading model AUC are cancer associated genes/proteins (Fig. 15).

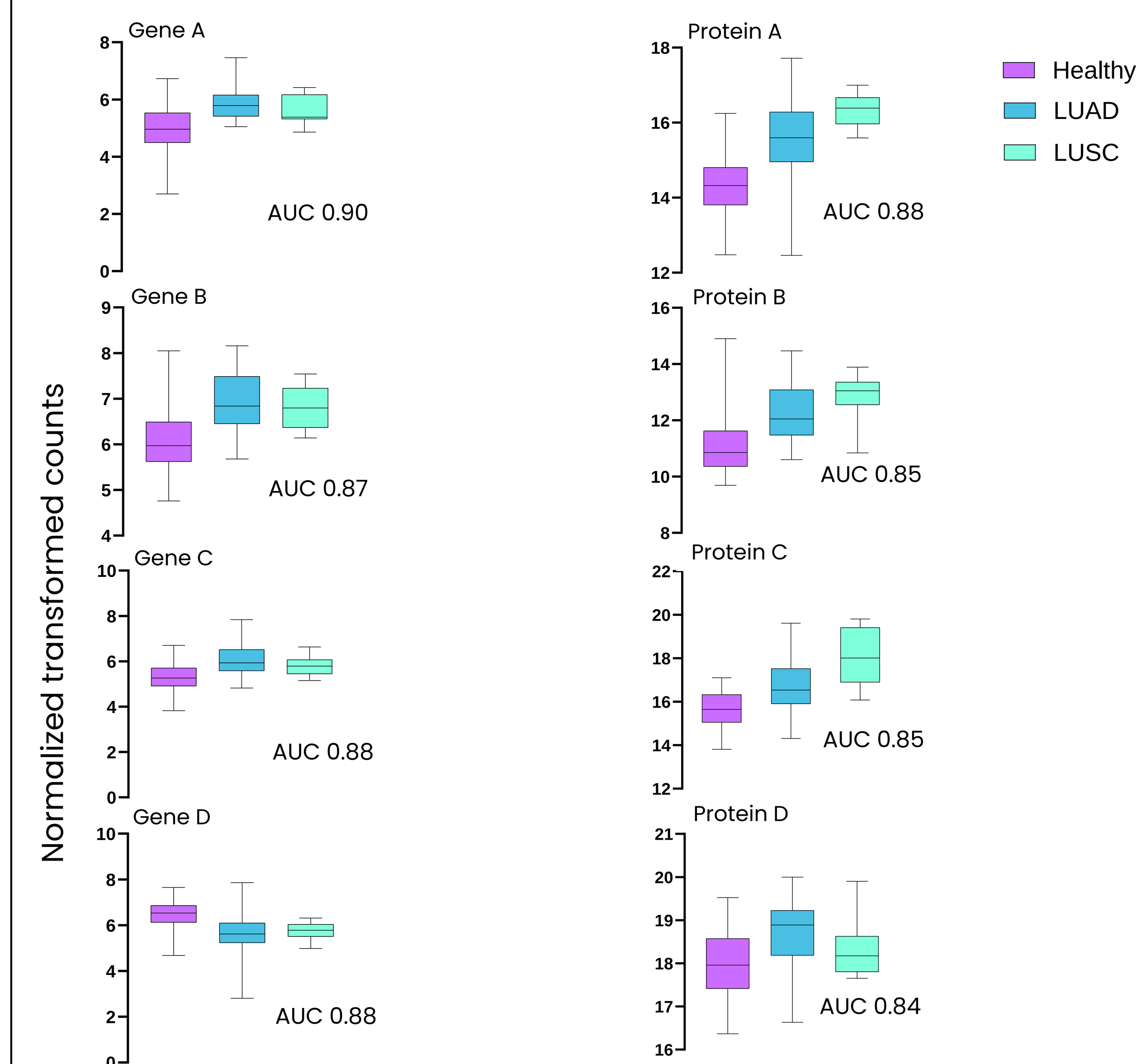

Figure 15. Normalized transformed counts for Healthy, LUAD and LUSC sample types for top candidate markers

Conclusions and Future Directions

- Blood plasma extracellular vesicles (EVs) are an attractive alternative to traditional diagnostics and provide a richer analyte base than ctDNA or cfDNA.
- SPARCs™-mediated enrichment and analysis identified a number of TDEV transcripts/proteins that can be used to distinguish NSCLC patients from healthy controls.
- FYR's leading machine-learning model shows high sensitivity and specificity in classifying patients and identifies candidate biomarkers that are cancer-associated genes/proteins.
- Future work will be focused on developing targeted assays that can be used to advance early screening capabilities for NSCLC.

References

1. *NCI Cancer Stat Facts: Common Cancer Sites*. Available from: https://seer.cancer.gov/statfacts/html/common.html.
2. Society, A.C., *Cancer Facts & Figures 2024.* Atlanta: American Cancer Society, 2024.
3. ALA. *Types of Lung Cancer*. https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/basics/lung-cancer.
4. Babar L, M.P., Anjum F., *Lung Cancer Screening*. StatPearls. 2024, Treasure Island, FL: StatPearls Publishing.
5. Poon, C., et al., *Why is the screening rate in lung cancer still low? A seven-country analysis of the factors affecting adoption.* Front Pub. Health, 2023. **11**: p. 1264342.
6. Connal, S., et al., *Liquid biopsies: the future of cancer early detection.* J Transl Med, 2023. **21**(1): p. 118.
7. Stejskal, P., et al., *Circulating tumor nucleic acids: biology, release mechanisms, and clinical relevance.* Mol Cancer, 2023. **22**(1): p. 15.
8. Bebelman, M.P., et al., *The forces driving cancer extracellular vesicle secretion.* Neoplasia, 2021. **23**(1): p. 149-157.