GHENT UNIVERSITY

CRIG

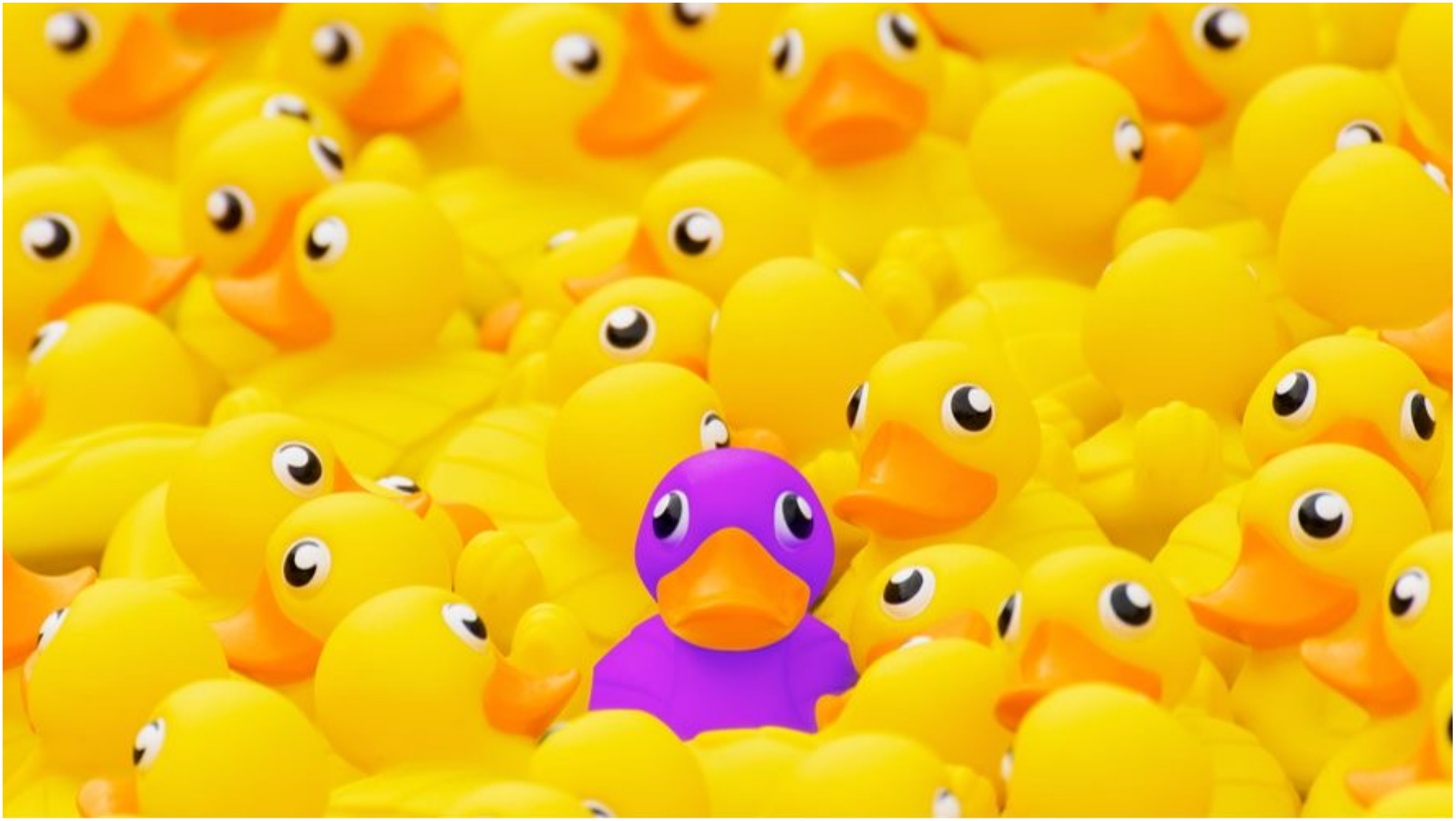
OncoRNALab

Extracellular RNA in biofluids: a tale of cancer biomarker tail genes

Jo Vandesompele

December 4, 2025

CNAPS, Hong Kong



Disclosures

- co-founder and CSO at pxlence
  - universal Rainbow detection probes for digital PCR
  - patent application EP4613876A1
- professor at Ghent University
  - co-inventor tail gene concept
  - patent application WO 2024/200811
- research funding from Qiagen, Illumina, Bio-Rad

OncoRNALab

Why study extracellular RNA?

- dynamic analyte
  - linked to health and disease states
  - pharmacodynamic biomarker
- versatile & complex
  - small/long – (un)spliced – linear/circular – (non)polyA – (non-)coding
  - informative at abundance & structural level
    - DGE/GSEA/deconvolution – splice isoforms/RNA-editing/somatic mutations
- human biofluids are full of exRNA – some more abundant than cfDNA
- outside oncology “cfRNA is all we have!” [in oncology: stroma/host]

- various (historic) reasons for lack of enthusiasm for mRNA
  - degraded / lack of methods / large dynamic range

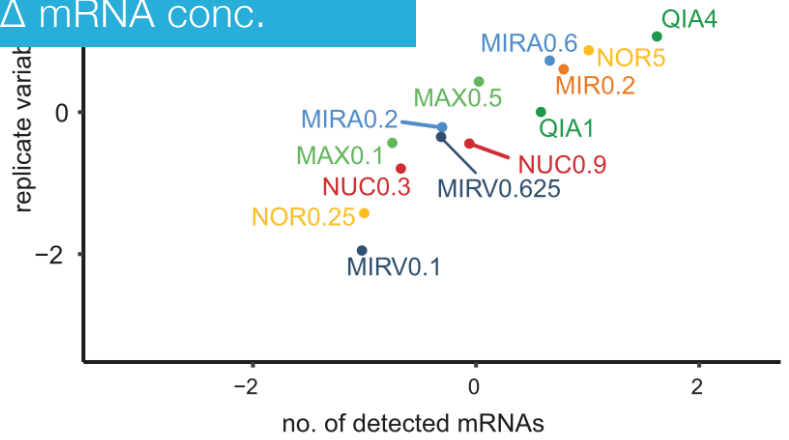
OncoRNALab

Problem 1: impact of pre-analytical variables

- large scale benchmarking
  - 456 exRNA transcriptomes
  - 10 blood collection tubes x 3 time intervals
  - 8 RNA purification kits
  - 11 performance metrics

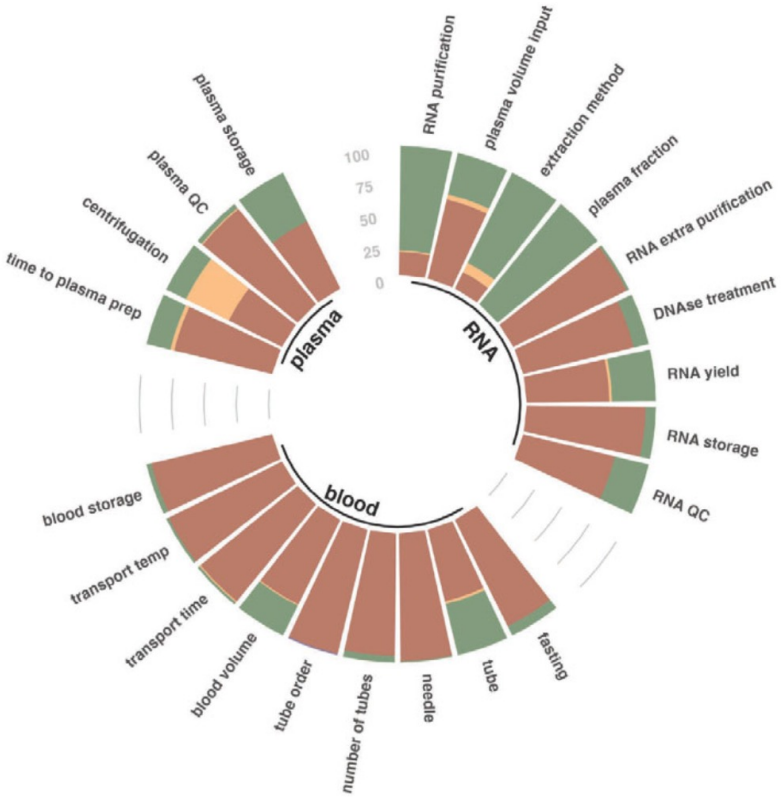
- recommendations for users and manufacturers
  - standardize workflow
  - maximize input in extraction
  - avoid ‘preservation’ tubes
    - EDTA/citrate processed withing 4 h

- 11x Δ detected genes
- 30x Δ mRNA yield
- 76x Δ mRNA conc.



exRNAQC consortium, Nature Communications, 2025

Problem 2: lack of reporting



Reported
Reported in insufficient detail
Not applicable
Unknown

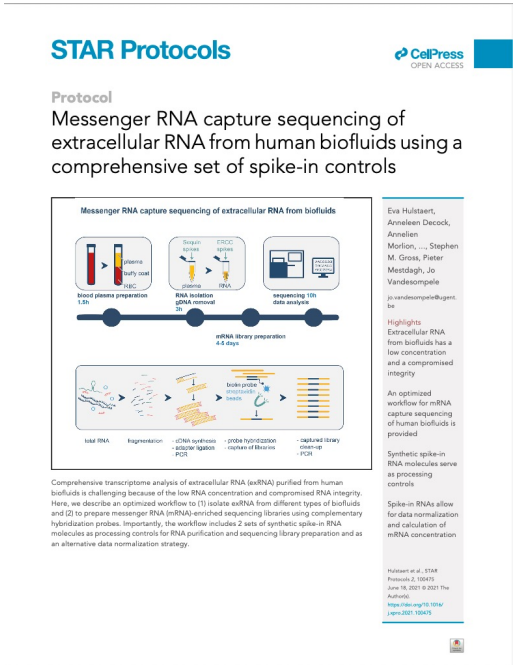
- 200 peer-reviewed articles in 2018/2023 on ‘blood plasma + RNA’
- evaluated 22 pre-analytical parameters
  - median level of reporting is 7%
  - only 6/22 variables reported in >50% articles
- reporting checklist for exRNA-based studies > reliable interpretation & replication
- CEN/TS 17742:2022

Van Der Schueren, Molecular Oncology, 2024

OncoRNALab

Problem 3: lack of methods

Hulstaert et al.
STAR Protocols, 2021
mRNA capture seq



STAR Protocols

CellPress
OPEN ACCESS

Protocol

Messenger RNA capture sequencing of extracellular RNA from human biofluids using a comprehensive set of spike-in controls

Messenger RNA capture sequencing of extracellular RNA from biofluids

blood plasma preparation 1.5h
RNA isolation gDNA removal 2h
sequencing 10h data analysis
mRNA library preparation 4-5 days

Comprehensive transcriptome analysis of extracellular RNA (exRNA) purified from human biofluids is challenging because of the low RNA concentration and compromised RNA integrity. Here, we describe an optimized workflow to (1) isolate exRNA from different types of biofluids and (2) to prepare messenger RNA (mRNA)-enriched sequencing libraries using complementary hybridization probes. Importantly, the workflow includes 2 sets of synthetic spike-in RNA molecules as processing controls for RNA purification and sequencing library preparation and as an alternative data normalization strategy.

Eva Hulstaert, Anneleen Decock, Annelien Morlion, ..., Stephen M. Gross, Pieter Mestdagh, Jo Vandesompele

jo.vandesompele@ugent.be

Highlights

Extracellular RNA from biofluids has a low concentration and a compromised integrity

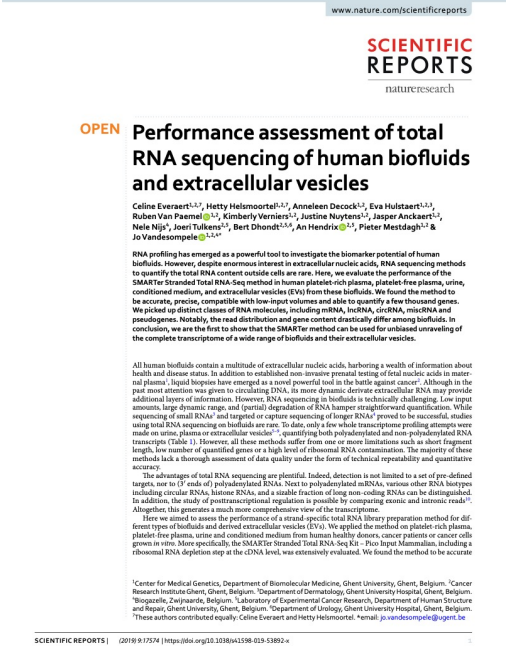
An optimized workflow for mRNA capture sequencing of human biofluids is provided

Synthetic spike-in RNA molecules serve as processing controls

Spike-in RNAs allow for data normalization and calculation of mRNA concentration

Hulstaert et al., STAR Protocols 2, 100475 June 18, 2021 © 2021 The Author(s). https://doi.org/10.1016/j.xpro.2021.100475

Everaert et al.
Scientific Reports, 2019
total RNA seq



www.nature.com/scientificreports

SCIENTIFIC REPORTS
nature research

OPEN

Performance assessment of total RNA sequencing of human biofluids and extracellular vesicles

Celine Everaert[1,2,7], Hetty Helsmoortel[1,2,7], Anneleen Decock[1,2], Eva Hulstaert[1,2,3], Ruben Van Paemel[1,2], Kimberly Verniers[1,2], Justine Nuytens[1,2], Jasper Anckaert[1,2], Nele Nijs[4], Joeri Tulkens[1,2], Bert Dhondt[1,5,6], An Hendrix[1,2], Pieter Mestdagh[1,2] & Jo Vandesompele[1,2,*]

RNA profiling has emerged as a powerful tool to investigate the biomarker potential of human biofluids. However, despite enormous interest in extracellular nucleic acids, RNA sequencing methods to quantify the total RNA content outside cells are rare. Here, we evaluate the performance of the SMARTer Stranded Total RNA-Seq method in human platelet-rich plasma, platelet-free plasma, urine, conditioned medium, and extracellular vesicles (EVs) from these biofluids. We found the method to be accurate, precise, compatible with low-input volumes and able to quantify a few thousand genes. We picked up distinct classes of RNA molecules, including mRNA, lncRNA, circRNA, miscRNA and pseudogenes. Notably, the read distribution and gene content drastically differ among biofluids. In conclusion, we are the first to show that the SMARTer method can be used for unbiased unraveling of the complete transcriptome of a wide range of biofluids and their extracellular vesicles.

All human biofluids contain a multitude of extracellular nucleic acids, harboring a wealth of information about health and disease status. In addition to established non-invasive prenatal testing of fetal nucleic acids in maternal plasma[1], liquid biopsies have emerged as a novel powerful tool in the battle against cancer[2]. Although in the past most attention was given to circulating DNA, its more dynamic derivate extracellular RNA may provide additional layers of information. However, RNA sequencing in biofluids is technically challenging. Low input amounts, large dynamic range, and (partial) degradation of RNA hamper straightforward quantification. While sequencing of small RNAs[3] and targeted or capture sequencing of longer RNAs[4] proved to be successful, studies using total RNA sequencing on biofluids are rare. To date, only a few whole transcriptome profiling attempts were made on urine, plasma or extracellular vesicles[5–7], quantifying both polyadenylated and non-polyadenylated RNA transcripts (Table 1). However, all these methods suffer from one or more limitations such as short fragment length, low number of quantified genes or a high level of ribosomal RNA contamination. The majority of these methods lack a thorough assessment of data quality under the form of technical repeatability and quantitative accuracy.

The advantages of total RNA sequencing are plentiful. Indeed, detection is not limited to a set of pre-defined targets, nor to (3′ ends of) polyadenylated RNAs. Next to polyadenylated mRNAs, various other RNA biotypes including circular RNAs, histone RNAs, and a sizable fraction of long non-coding RNAs can be distinguished. In addition, the study of posttranscriptional regulation is possible by comparing exonic and intronic reads[8,9]. Altogether, this generates a much more comprehensive view of the transcriptome.

Here we aimed to assess the performance of a strand-specific total RNA library preparation method for different types of biofluids and derived extracellular vesicles (EVs). We applied the method on platelet-rich plasma, platelet-free plasma, urine and conditioned medium from human healthy donors, cancer patients or cancer cells grown *in vitro*. More specifically, the SMARTer Stranded Total RNA-Seq Kit – Pico Input Mammalian, including a ribosomal RNA depletion step at the cDNA level, was extensively evaluated. We found the method to be accurate

[1]Center for Medical Genetics, Department of Biomolecular Medicine, Ghent University, Ghent, Belgium. [2]Cancer Research Institute Ghent, Ghent, Belgium. [3]Department of Dermatology, Ghent University Hospital, Ghent, Belgium. [4]Biogazelle, Zwijnaarde, Belgium. [5]Laboratory of Experimental Cancer Research, Department of Human Structure and Repair, Ghent University, Ghent, Belgium. [6]Department of Urology, Ghent University Hospital, Ghent, Belgium. [7]These authors contributed equally: Celine Everaert and Hetty Helsmoortel. *email: jo.vandesompele@ugent.be

SCIENTIFIC REPORTS | (2019) 9:17574 | https://doi.org/10.1038/s41598-019-53892-x

Morlion et al.
RNA Biology, 2021
lncRNA capture seq

RNA BIOLOGY
2021, VOL. 18, NO. S1, 215–222
https://doi.org/10.1080/15476286.2021.1971438

Taylor & Francis
Taylor & Francis Group

TECHNICAL PAPER

OPEN ACCESS

Custom long non-coding RNA capture enhances detection sensitivity in different human sample types

Annelien Morlion[a,b,*], Celine Everaert[a,b,*], Justine Nuytens[a,b], Eva Hulstaert[a,b,c], Jo Vandesompele[a,b], and Pieter Mestdagh[a,b]

[a]OncoRNALab, Center for Medical Genetics, Department of Biomolecular Medicine, Ghent University, B-9000 Ghent, Belgium; [b]Cancer Research Institute Ghent (CRIG), Ghent, Belgium; [c]Department of Dermatology, Ghent University Hospital, Ghent, Belgium

ABSTRACT

Long non-coding RNAs (lncRNAs) are a heterogeneous group of transcripts that lack protein coding potential and display regulatory functions in various cellular processes. As a result of their cell- and cancer-specific expression patterns, lncRNAs have emerged as potential diagnostic and therapeutic targets. The accurate characterization of lncRNAs in bulk transcriptome data remains challenging due to their low abundance compared to protein coding genes. To tackle this issue, we describe a unique short-read custom lncRNA capture sequencing approach that relies on a comprehensive set of 565,878 capture probes for 49,372 human lncRNA genes. This custom lncRNA capture approach was evaluated on various sample types ranging from artificial high-quality RNA mixtures to more challenging formalin-fixed paraffin-embedded tissue and biofluid material. The custom enrichment approach allows the detection of a more diverse repertoire of lncRNAs, with better reproducibility and higher coverage compared to classic total RNA-sequencing.

ARTICLE HISTORY
Received 12 March 2021
Revised 4 August 2021
Accepted 18 August 2021

KEYWORDS
lncRNA; RNA sequencing; probes; lncRNome; RNA abundance; RNA expression; FFPE; biofluid

Introduction

While the majority of the human genome is actively transcribed into RNA transcripts, most of these transcripts do not code for proteins [1]. The non-coding RNA transcripts longer than 200 nucleotides belong to the heterogeneous group of long non-coding RNAs (lncRNAs), half of which are not polyadenylated [2]. These lncRNAs are known to influence gene expression at both the transcriptional and post-transcriptional level through a variety of mechanisms [3,4]. Moreover, lncRNAs often show a particular cell- or cancer-type specific expression pattern [5], which adds to their biomarker potential.

In the past, several high-throughput methods have been developed to profile the long non-coding RNA transcriptome, study their structure or define their function [6,7]. Because of their generally low abundance compared to protein coding genes, quantification of lncRNAs in bulk transcriptome data remains challenging. Enrichment strategies favouring lncRNAs over the more abundant mRNAs could therefore result in more lncRNAs being detected with a better transcript coverage, improving downstream analysis. A promising method is RNA capture sequencing, a short-read sequencing method that can enrich RNA targets of interest using oligonucleotide probes that are specifically designed to tile the target sequences. These RNA capture sequencing technologies have mainly been applied for deep sequencing of a selection of lncRNAs [8,9]. Recently, the GENCODE consortium extended this method by applying long-read sequencing after capturing about 14,470 lncRNAs genes to improve their structural annotation (RNA Capture Long Seq, RNA CLS) [10].

In this study, we describe a custom lncRNA capture sequencing approach that targets a very comprehensive human lncRNome. This custom capture approach was evaluated on various sample types ranging from high-quality RNA mixtures to more challenging formalin-fixed paraffin-embedded (FFPE) tissue and biofluid material.

Material and methods

Probe design

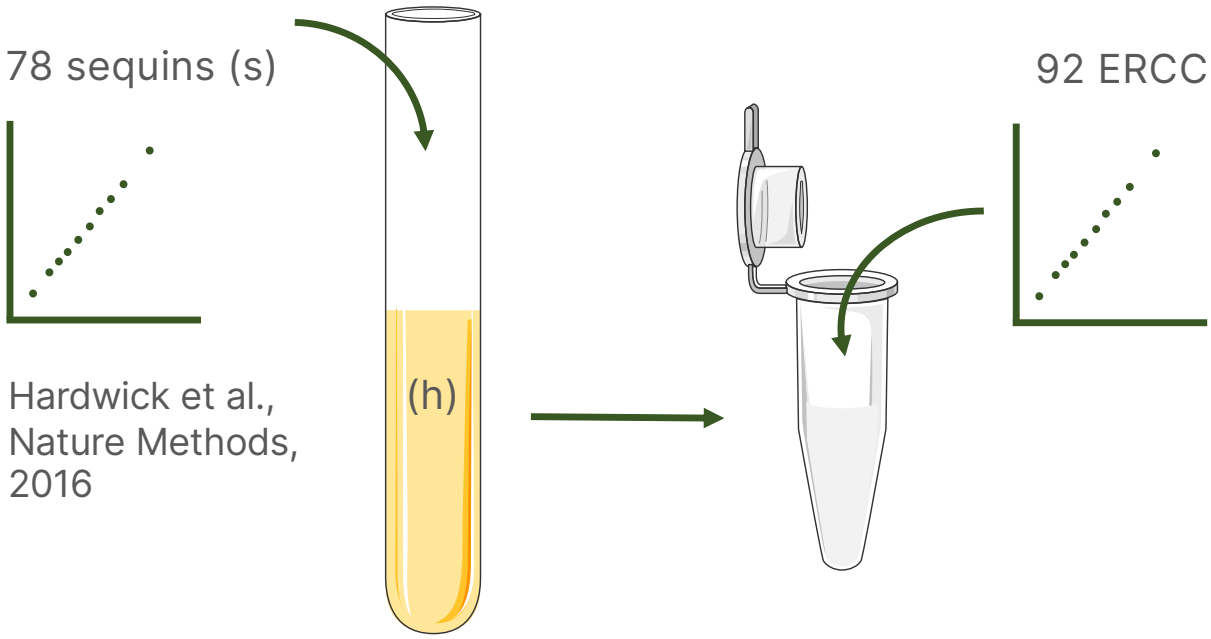
Probes were designed against the highly confident set of LNCipedia 5.2 (hg19 genome build). First, extended exons were created by concatenating each set of overlapping exons. For each of these extended exons, probes of 120 nucleotides were tiled, resulting in (number of nucleotides)-119 probes per concatenated exon. These exon tiling probes were mapped against repeat regions and protein coding genes to filter out these that would capture off-target fragments.

The resulting probe pool was extended with probes designed to capture both the Sequin and ERCC spikes. These probes are 120-mers designed by tiling the spike-in

CONTACT Pieter Mestdagh, pieter.mestdagh@ugent.be OncoRNALab, Center for Medical Genetics, Department of Biomolecular Medicine, Ghent University, Medical Research Building 1, Corneel Heymanslaan 10, B-9000 Ghent, Belgium

*Joint first authors.
Supplemental data for this article can be accessed here

Spike-in RNA as processing controls and normalization tool



Hardwick et al., Nature Methods, 2016

fluid input volume (f) for RNA purification

RNA eluate volume (r) in library prep

$$\frac{\text{human}}{\text{sequin}} = \text{absolute RNA conc.}$$

$$\frac{sequin}{ERCC} = purification\ eff.$$

19 small RNA spike-in controls (RC, LP)

OncoRNALab

Problem 4: Extract RNA or DNA from precious sample?

- Why not both?
- several commercial methods that co-purify cfDNA and exRNA
- evidence for increased mutation detection sensitivity (refs. 8-11)

RESEARCH Open Access

Digital PCR-based evaluation of nucleic acid extraction kit performance for the co-purification of cell-free DNA and RNA

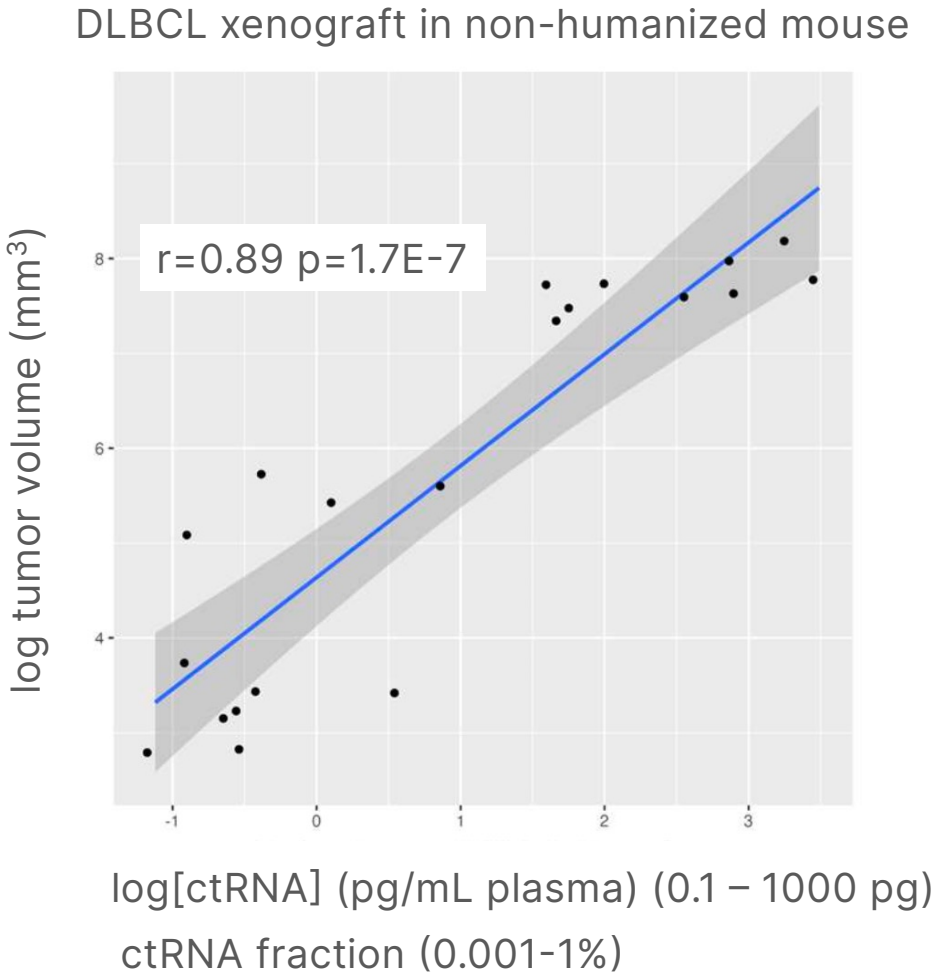
Jill Deleu[1,2†], Kathleen Schoofs[1,2,3,4†], Anneleen Decock[1,2], Kimberly Verniers[1,2], Sofie Roelandt[2,3,4], Angie Denolf[2,3], Joke Verreth[1,2], Bram De Wilde[1,2,5], Tom Van Maerken[1,2,6], Katleen De Preter[2,3,4] and Jo Vandesompele[1,2*]

Deleu et al., Human Genomics, 2022

OncoRNALab

Tale 1 – xenograft insights into ctRNA (1)

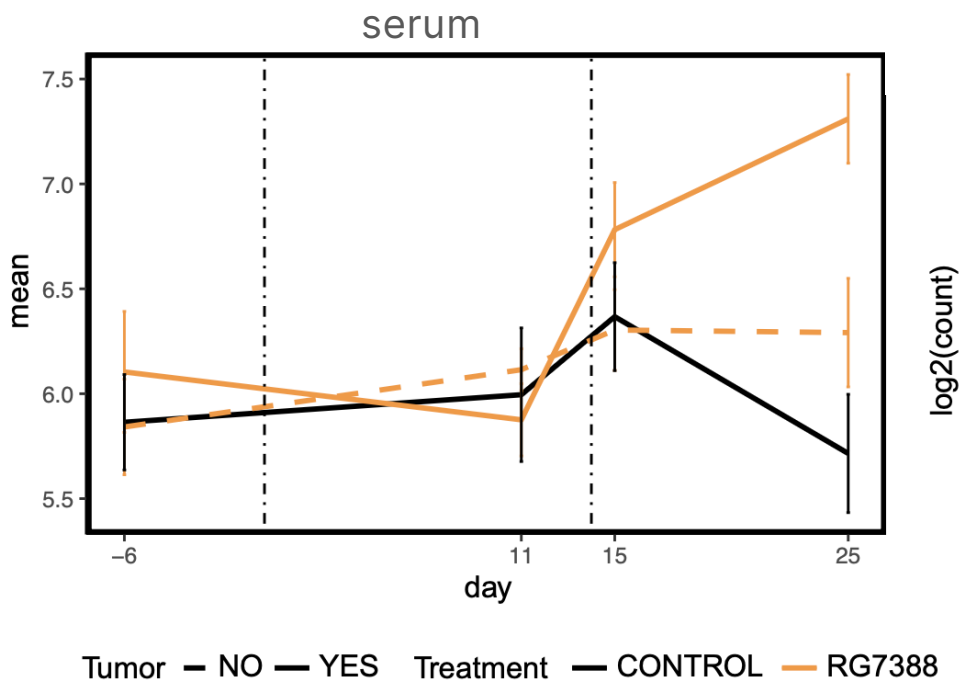
- human RNA in blood plasma = ctRNA
- highest conc. of ctRNA is in platelet-devoid plasma – TEP debunked
- [ctRNA] depends on
  - cancer type, model system, volume, vascularity, stage, treatment
  - similar ranges as ctDNA in humans
- individual ctRNA gene levels ~ tumor expression levels
  - no evidence for specific sorting

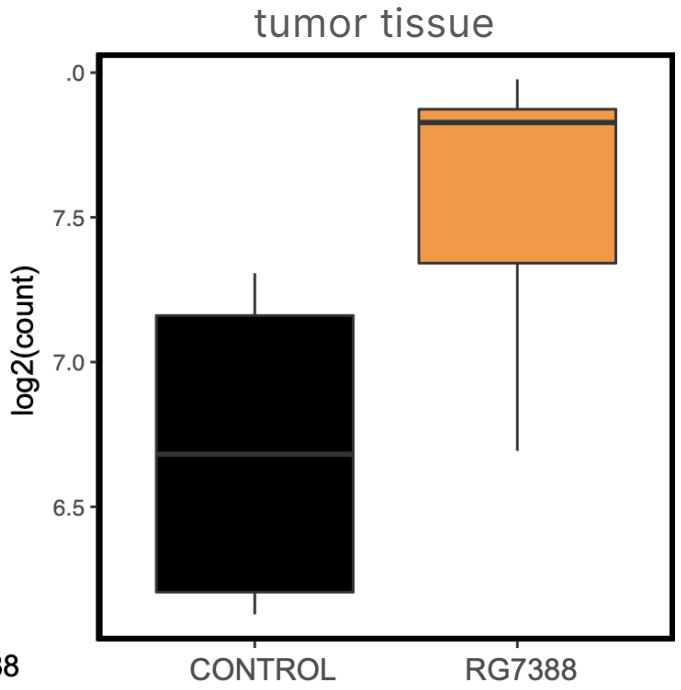


- Vermeirssen et al., NAR Cancer, 2022
- Decruyenaere et al., Frontiers in Oncology, 2023
- Deleu et al., exRNA, 2024

Tale 1 – xenograft insights into ctRNA (2)

- miR34a is a pharmacodynamic biomarker of on-target drug activity
  - idasanutlin (RG7388) treatment of neuroblastoma

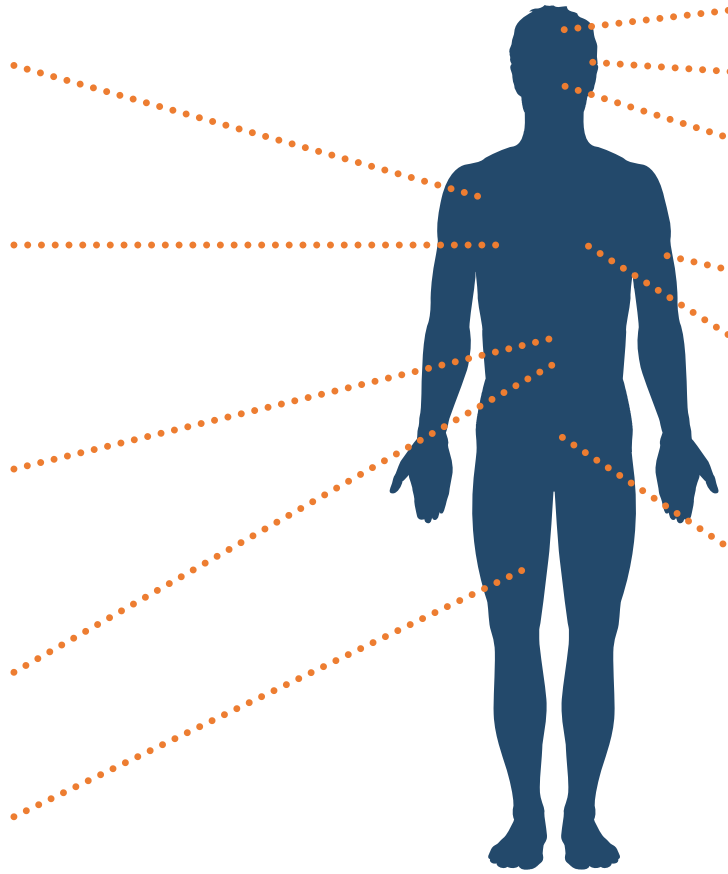




- Van Goethem et al., NAR Cancer, 2023

OncoRNALab

Tale 2: human biofluid RNA Atlas (1)



sweat

colostrum
mature breast milk
amniotic fluid

gastric fluid
bile
peritoneal fluid

pancreatic cyst fluid

synovial fluid

cerebrospinal fluid

tears
aqueous humor

saliva

blood plasma (PRP, PFP, PPP)
serum

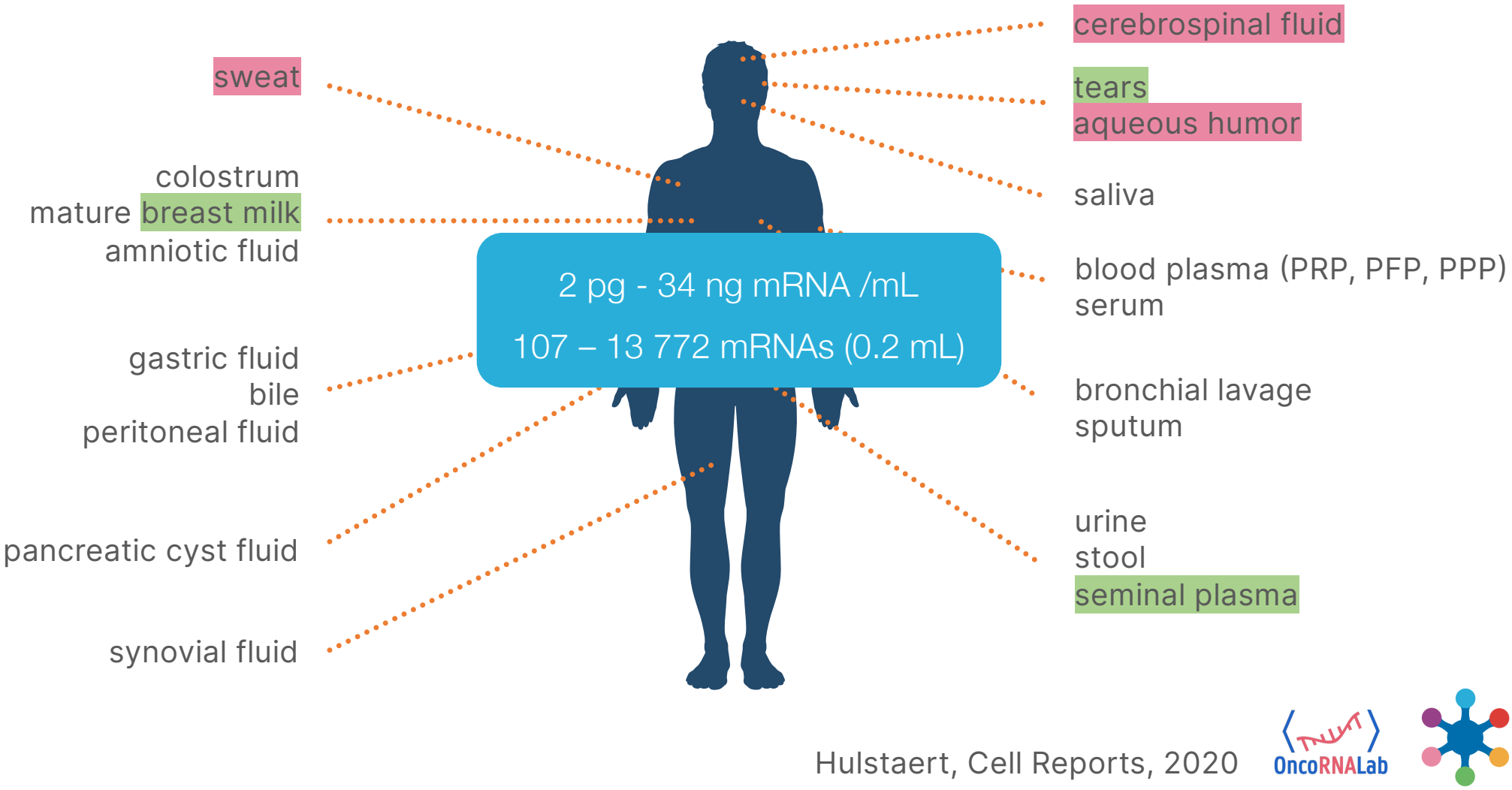
bronchial lavage
sputum

urine
stool
seminal plasma

Hulstaert, Cell Reports, 2020

OncoRNALab

Tale 2: human biofluid RNA Atlas (1)



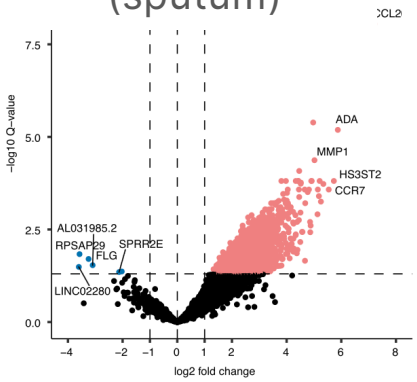
Hulstaert, Cell Reports, 2020

OncoRNALab

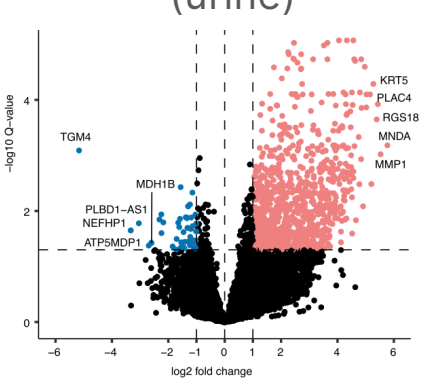
Tale 2: human biofluid RNA Atlas (2)

- enrichment of circular RNA (CiLiQuant tool)
- presence of bacterial RNA [0-30%]
  - stool, saliva, sputum, sweat
  - e.g. Campylobacter concisus in saliva
- biomarker potential

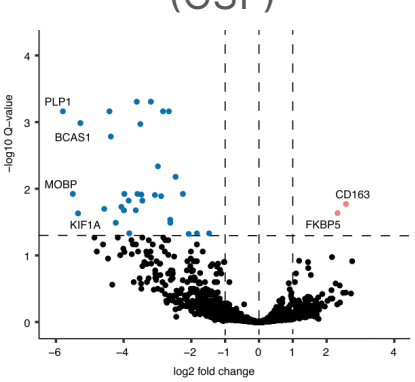
COPD (sputum)



bladder cancer (urine)



glioblastoma (CSF)

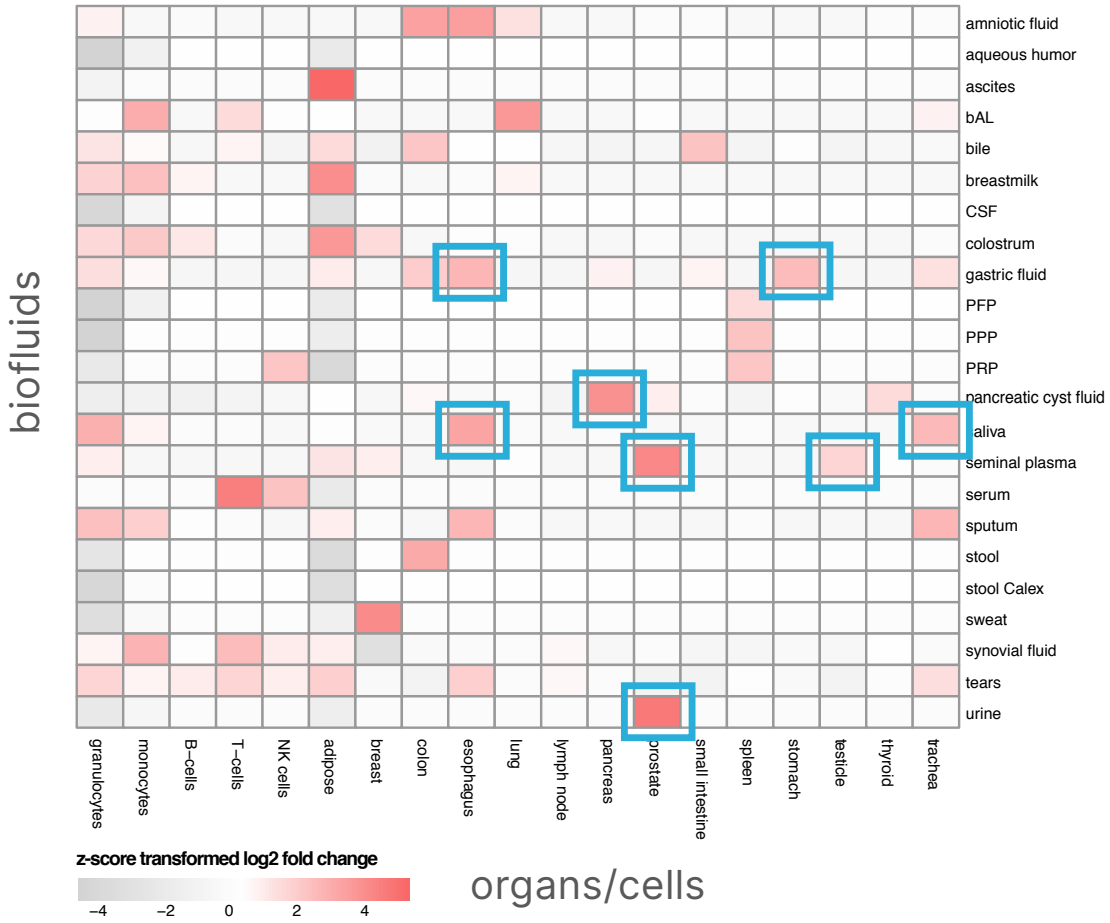


Morlion et al., Frontiers in Bioinformatics, 2022
Hulstaert, Cell Reports, 2020

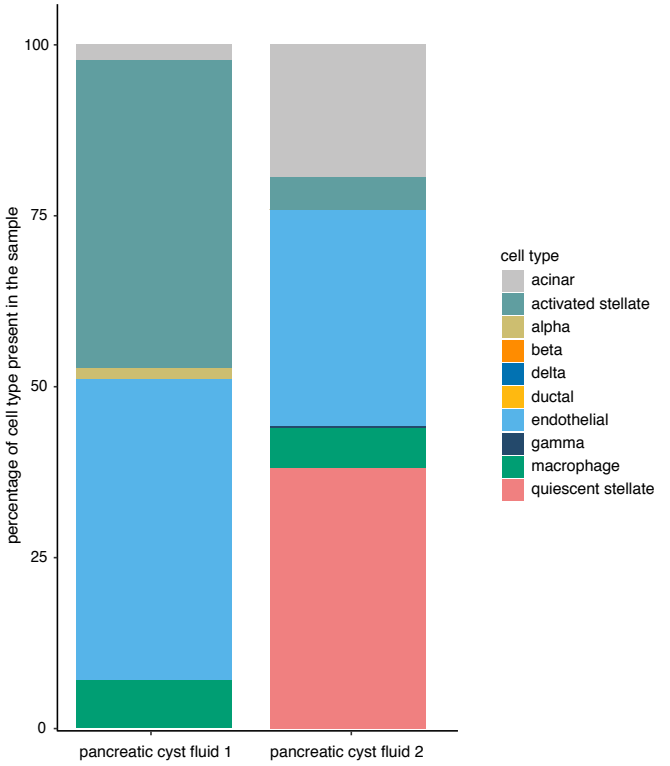
OncoRNALab

Tale 2: human biofluid RNA Atlas (3)

organ-of-origin / transport



pancreatic cyst cell type contribution

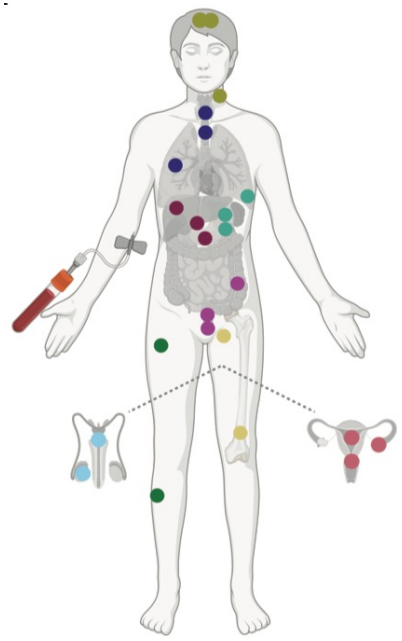


Hulstaert, Cell Reports, 2020

OncoRNALab

Tale 3: a pan-cancer cohort study (1)

- blood plasma from 25 cancer types (locally advanced or metastatic) vs. age-matched healthy controls (each n=8)



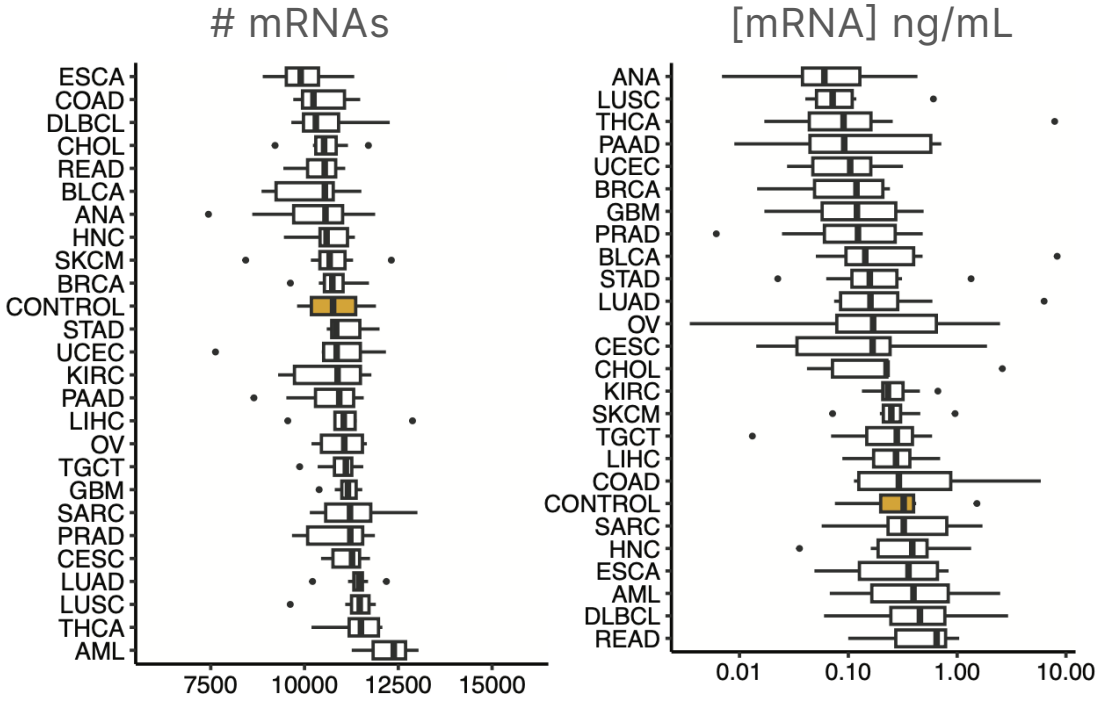
- glioblastoma multiforme (GBM)
- anaplastic astrocytoma (ANA)
- head and neck cancer (HNC)
- thyroid carcinoma (THCA)
- esophageal carcinoma (ESCA)
- lung cancer: squamous cell (LUSC) & adenocarcinoma (LUAD)
- breast carcinoma (BRCA)
- stomach adenocarcinoma (STAD)
- kidney renal clear cell carcinoma (KIRC)
- liver hepatocellular carcinoma (LIHC)
- cholangiocarcinoma (CHOL)
- pancreatic adenocarcinoma (PAAD)
- colon adenocarcinoma (COAD)
- bladder carcinoma (BLCA)
- rectal adenocarcinoma (READ)
- diffuse large B-cell lymphoma (DLBCL)
- acute myeloid leukemia (AML)
- sarcoma (SARC)
- skin cutaneous melanoma (SKCM)
- prostate adenocarcinoma (PRAD)
- testicular germ cell tumor (TGCT)
- ovarian carcinoma (OV)
- cervical squamous cell carcinoma (CESC)
- uterine corpus endometrial carcinoma (UCEC)

Morlion et al., medRxiv, 2024

OncoRNALab

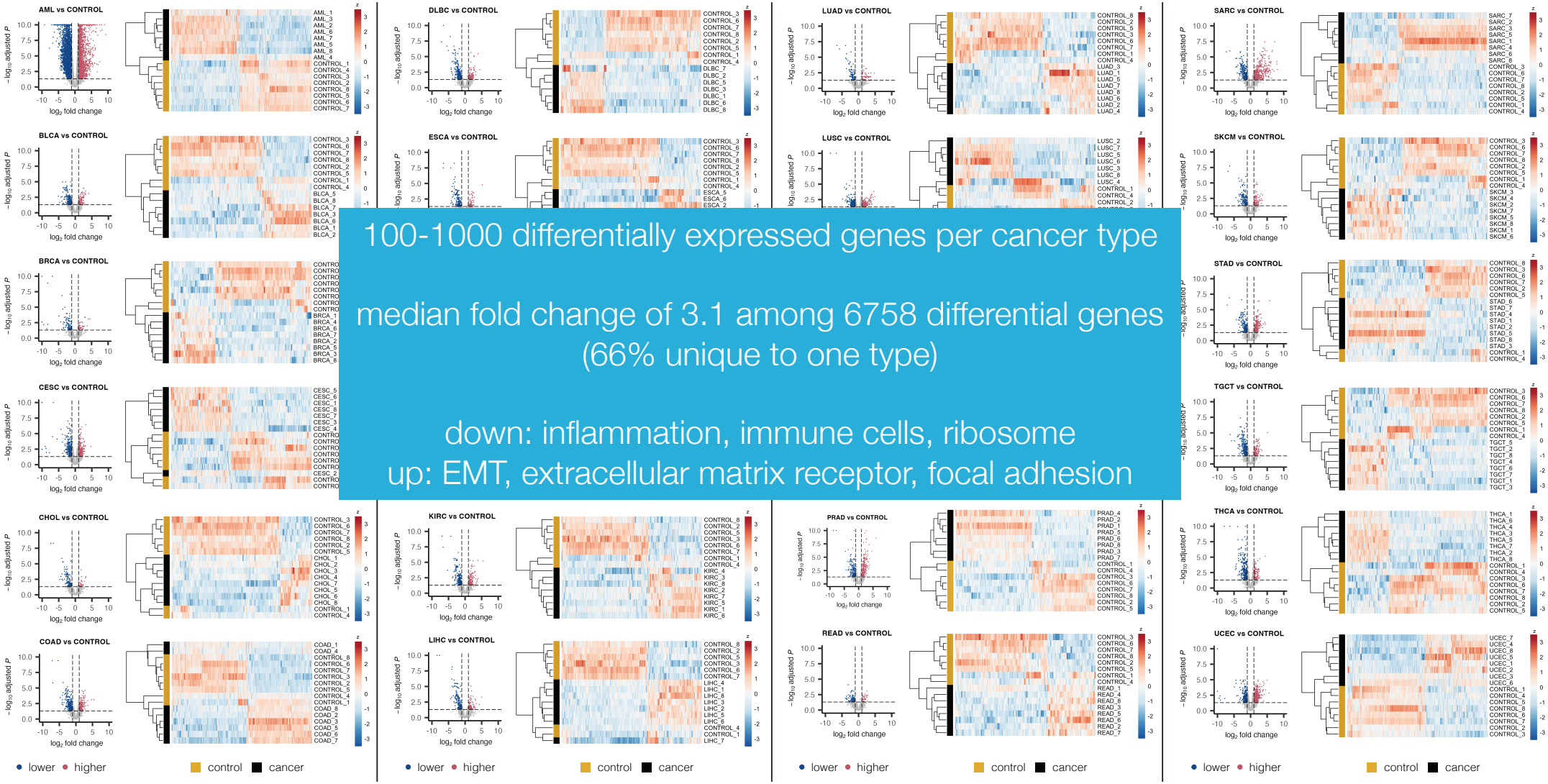
Tale 3: a pan-cancer cohort study (2)

- 10,000 – 12,500 mRNAs detected (median/type)
- 0.1 – 1 ng mRNA/mL (note: mRNA is only 1% of the genome)



Morlion et al., medRxiv, 2024

OncoRNALab



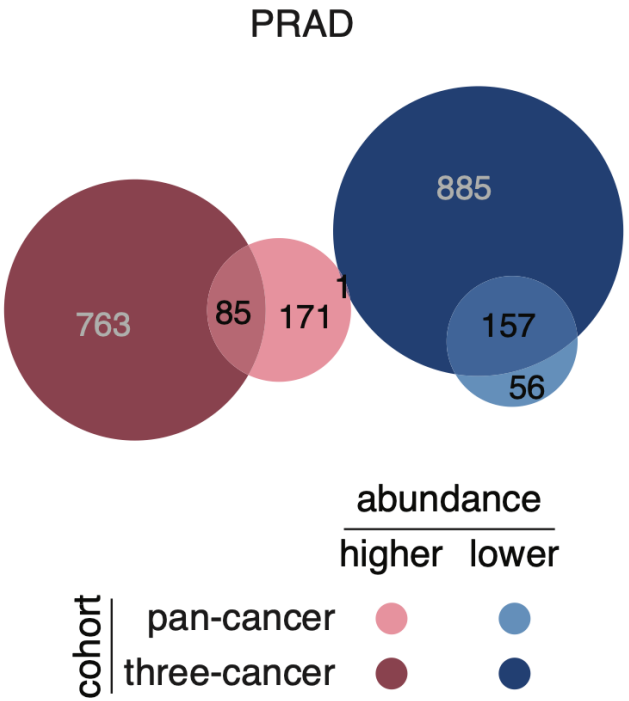
100-1000 differentially expressed genes per cancer type

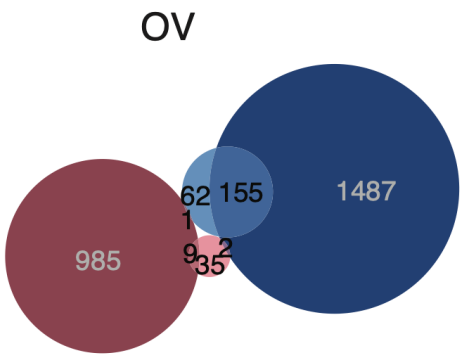
median fold change of 3.1 among 6758 differential genes (66% unique to one type)

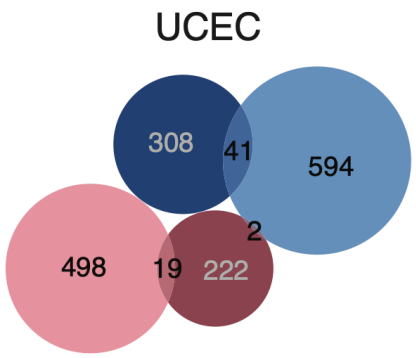
down: inflammation, immune cells, ribosome
up: EMT, extracellular matrix receptor, focal adhesion

So far so good, but ....

- cell-of-origin: only liver cancer
- evidence for ctRNA: only AML (PML:RARA fusion)*
- little overlap in DGE in independent cohorts for prostate, ovarian, and uterine cancer





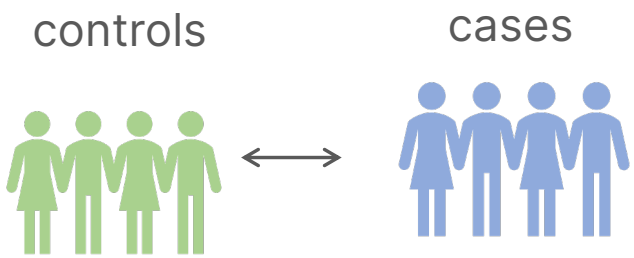


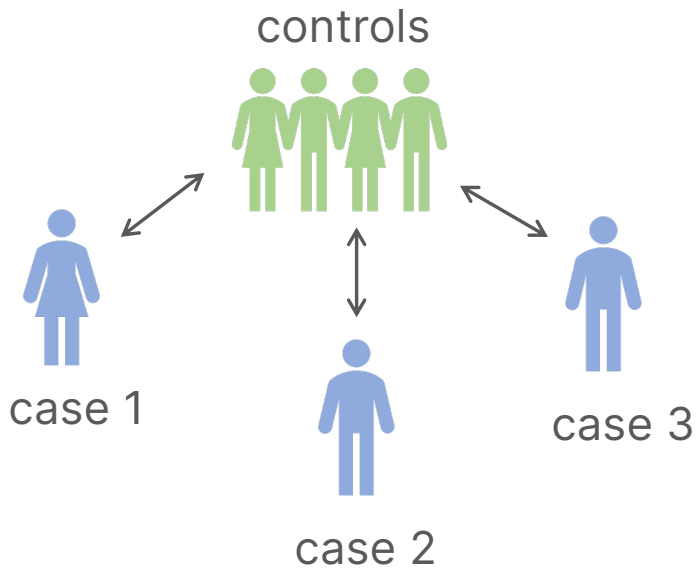
Lack of overlap because of heterogeneous nature of exRNA profiles (and tumor themselves)?

Morlion et al., medRxiv, 2024

OncoRNALab

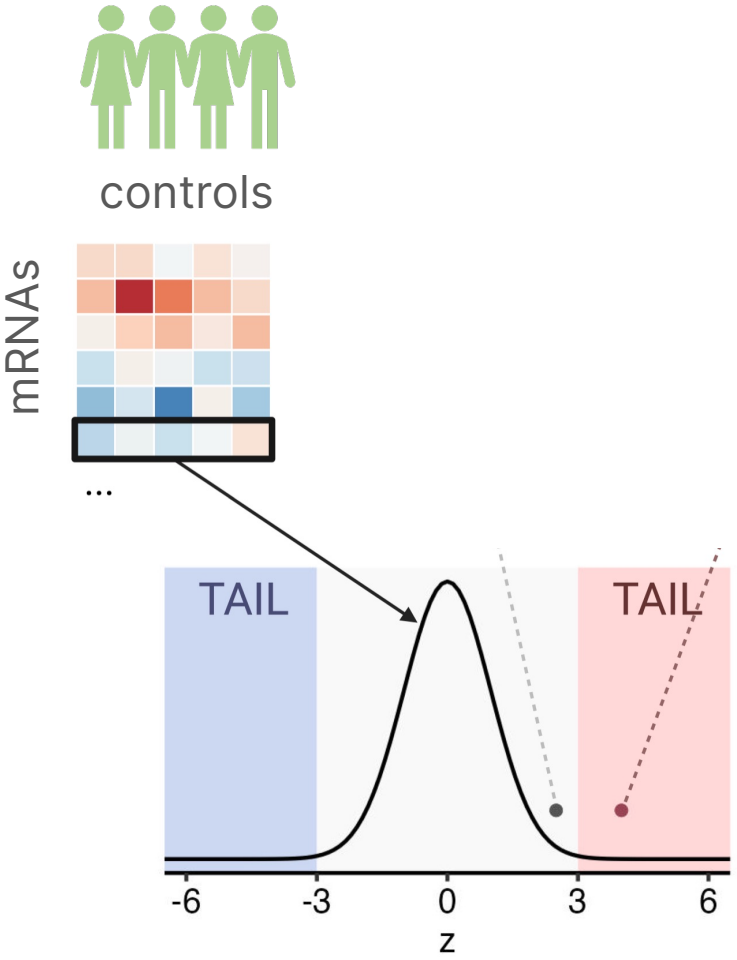
An alternative approach to find biomarkers in noisy or heterogeneous data

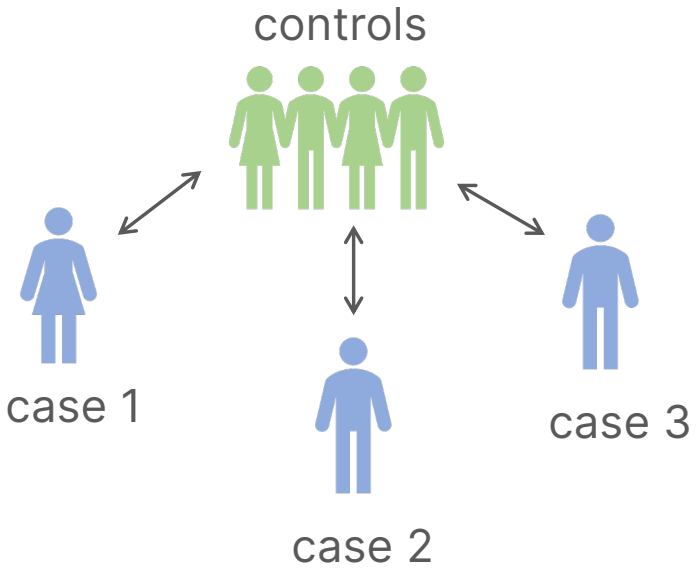




Morlion et al., medRxiv, 2024

An alternative approach to find biomarkers in noisy or heterogeneous data

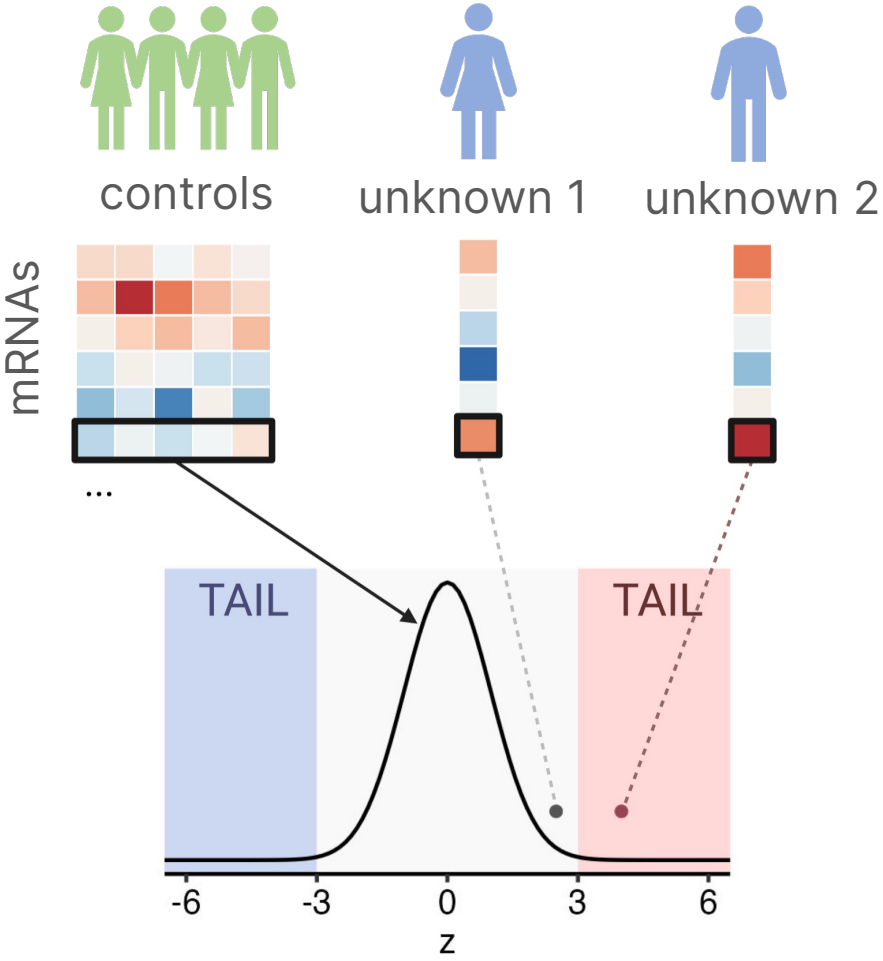


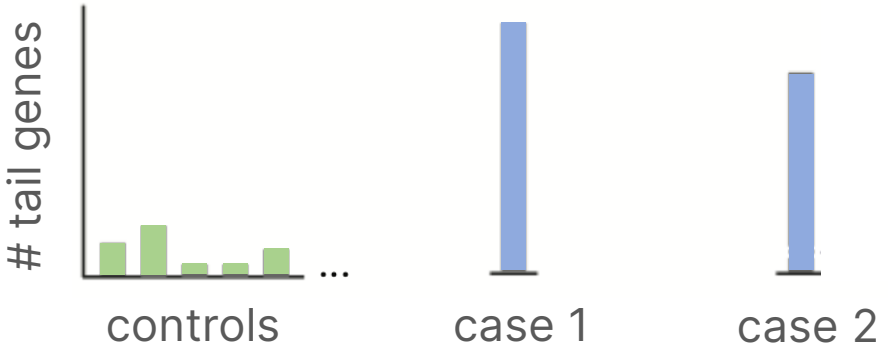


Morlion et al., medRxiv, 2024

OncoRNALab

An alternative approach to find biomarkers in noisy or heterogeneous data



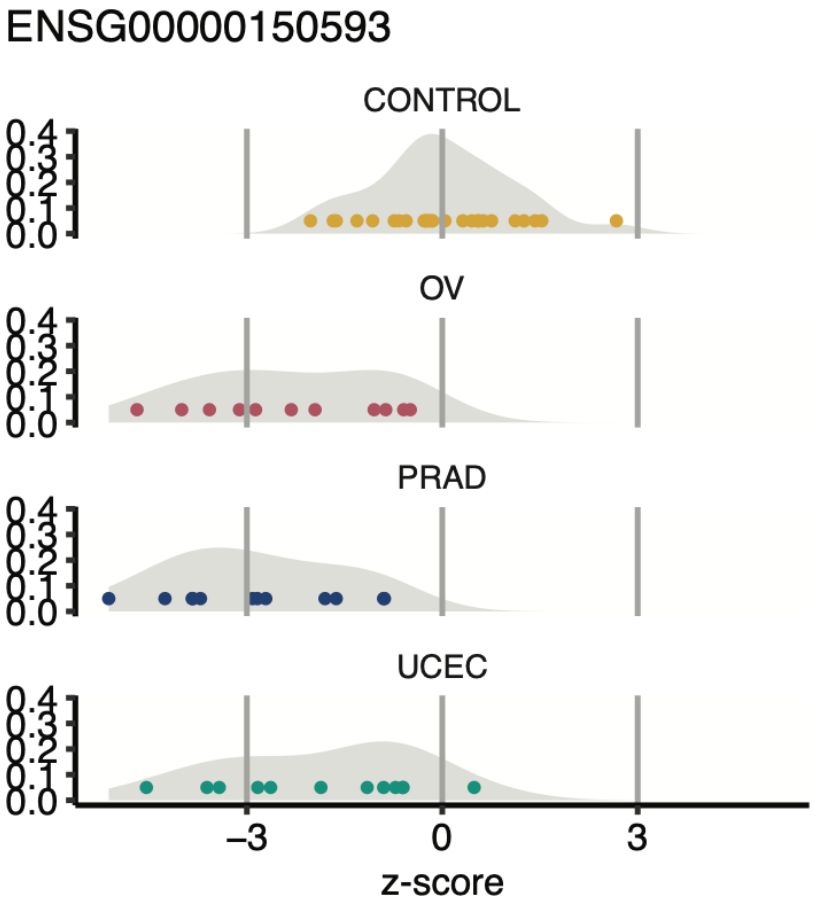


- Fisher's exact test > biomarker tail genes
- 10× 5-fold cross validation
- independent cohort validation

Morlion et al., medRxiv, 2024

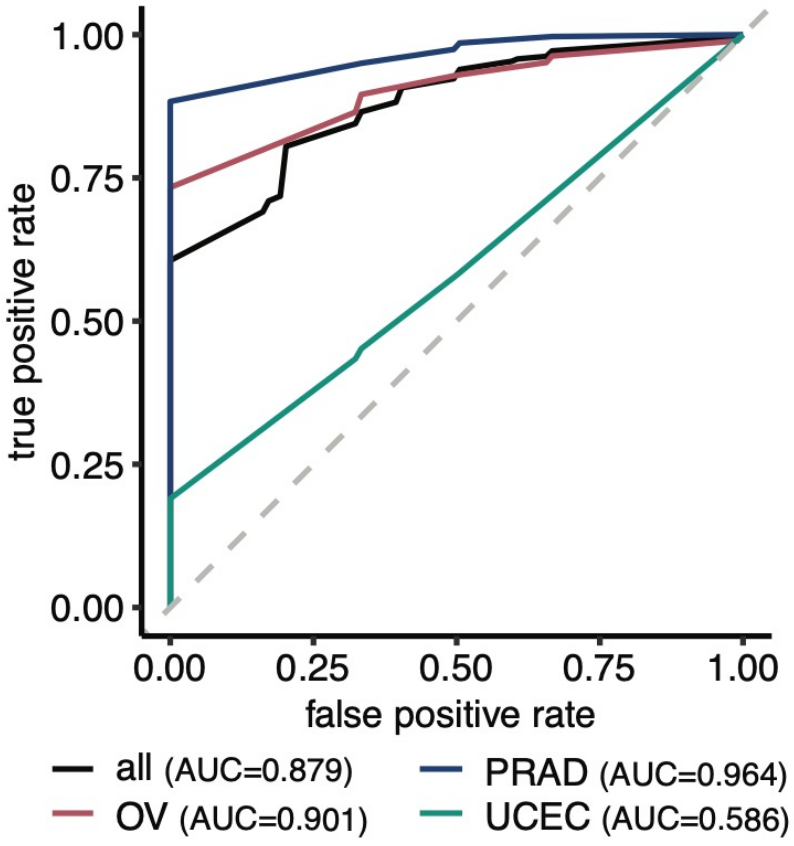
OncoRNALab

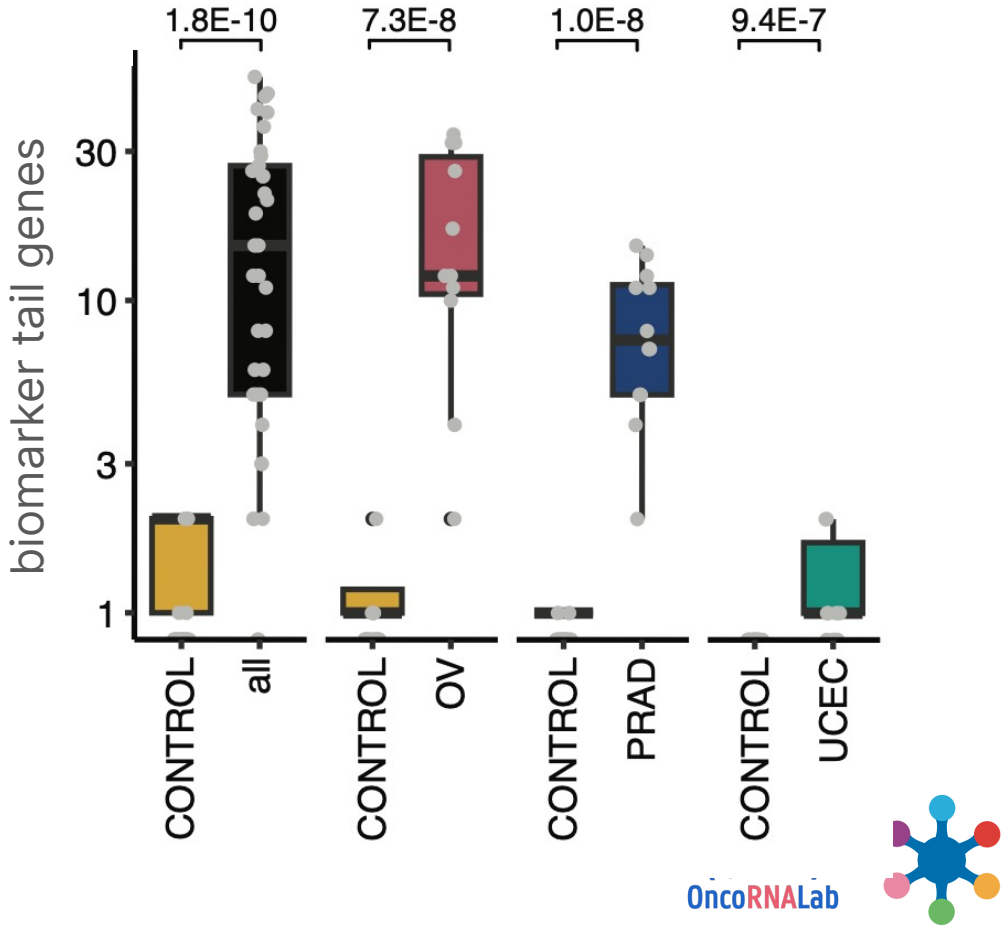
Example tail gene



OncoRNALab

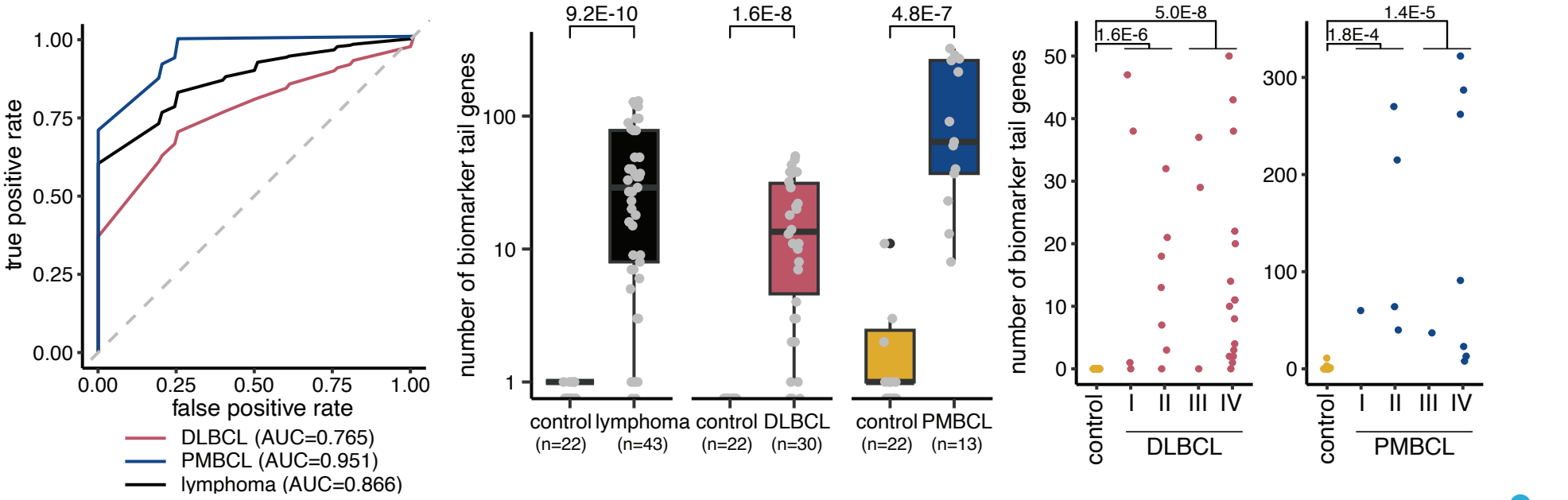
Classification of cancer and controls based on biomarker tail genes





Biomarker tail gene concept validated in lymphoma

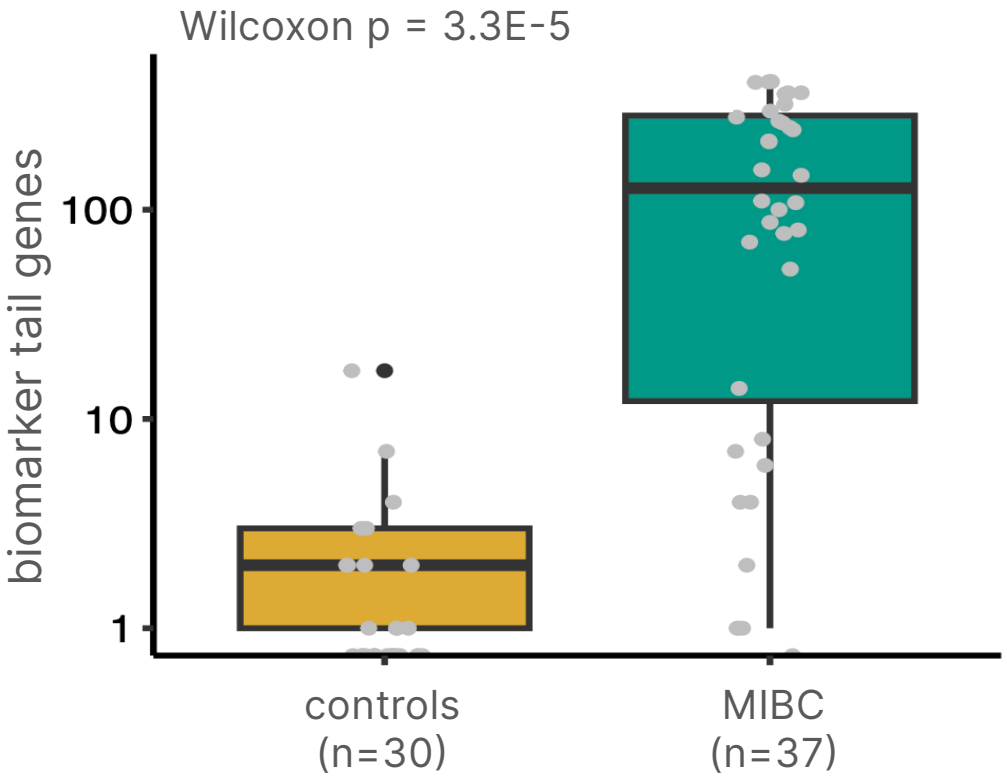
- different blood collection tube – different lib prep (total RNA sequencing)
- across all stages, both DLBCL and PMBCL

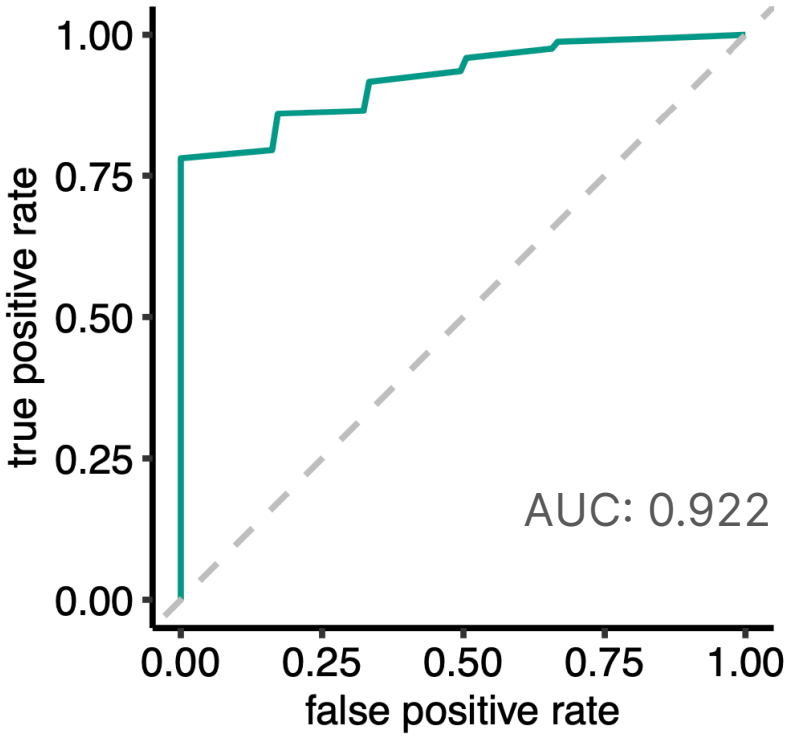


Decruyenaere et al., Frontiers in Oncology, 2023

OncoRNALab

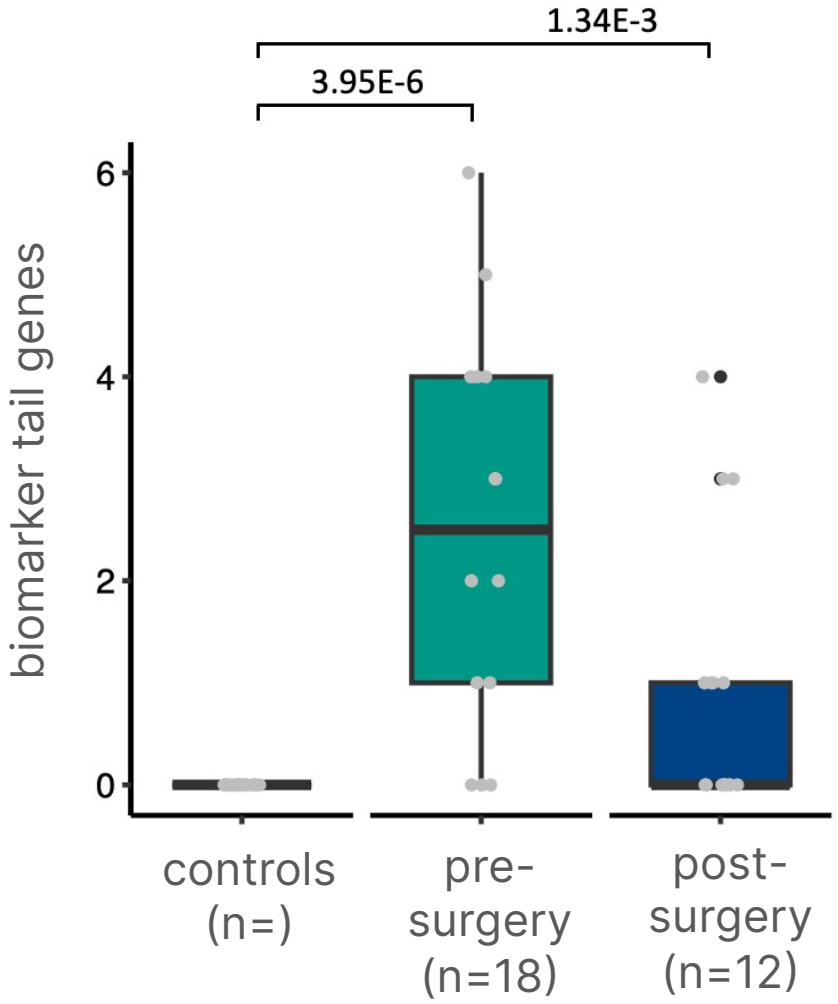
Tail genes in urine from muscle-invasive bladder cancer



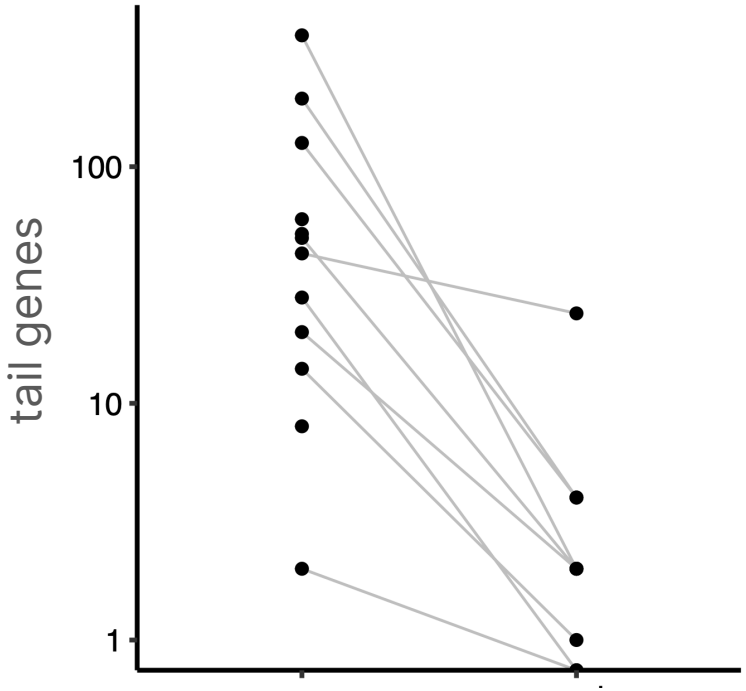


OncoRNALab

Tail genes in blood plasma from glioblastoma patients



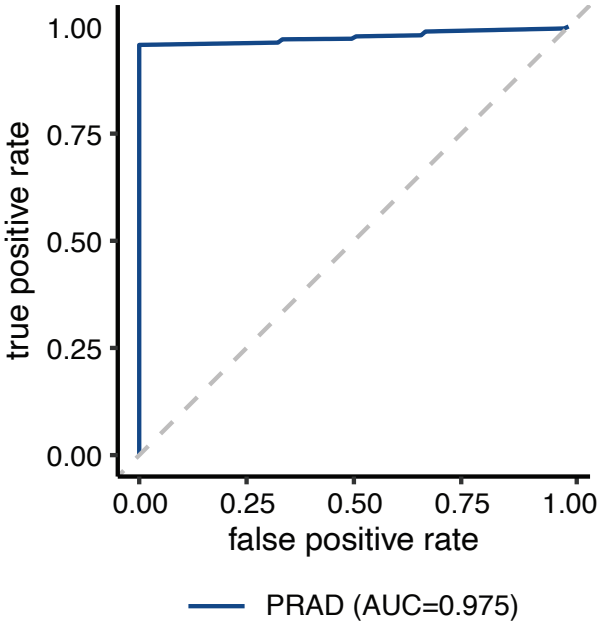
8/12 cases no more tail genes after surgery

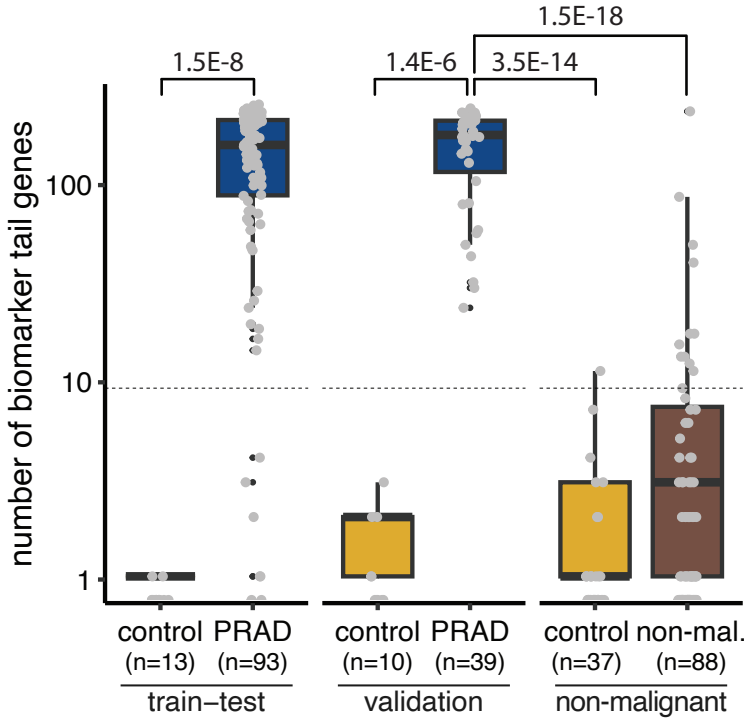


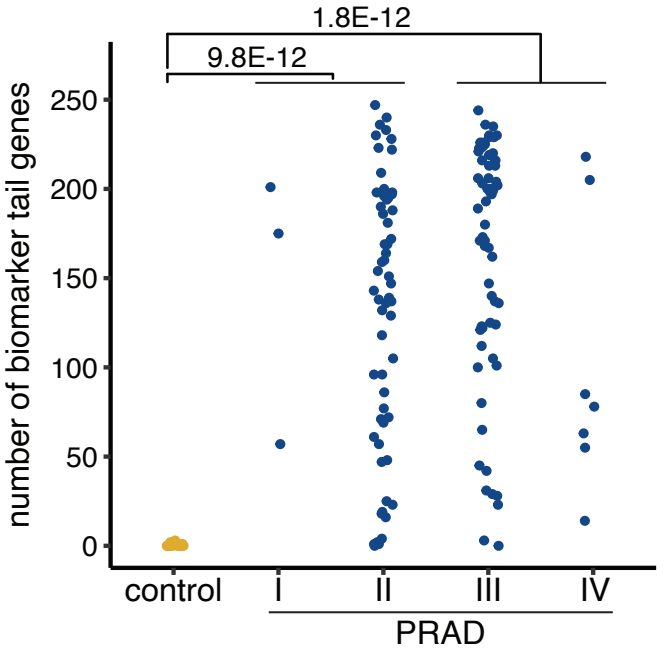
OncoRNALab

Large independent validation in prostate cancer and non-malignant patients

- 132 prostate cancer + 23 controls > 70% train-test + 30% validation
- 88 non-malignant patients + 37 controls







no link with grade or PSA level

Take home messages

- rich repertoire of RNA in human biofluids
- exRNA is a powerful biomarker source
- platelet-depleted blood plasma is most informative for oncology
- (pre-)analytical method standardization + transparent reporting is key
- tail genes are a novel and robust biomarker concept

jo.vandesompele@ugent.be

OncoRNALab