

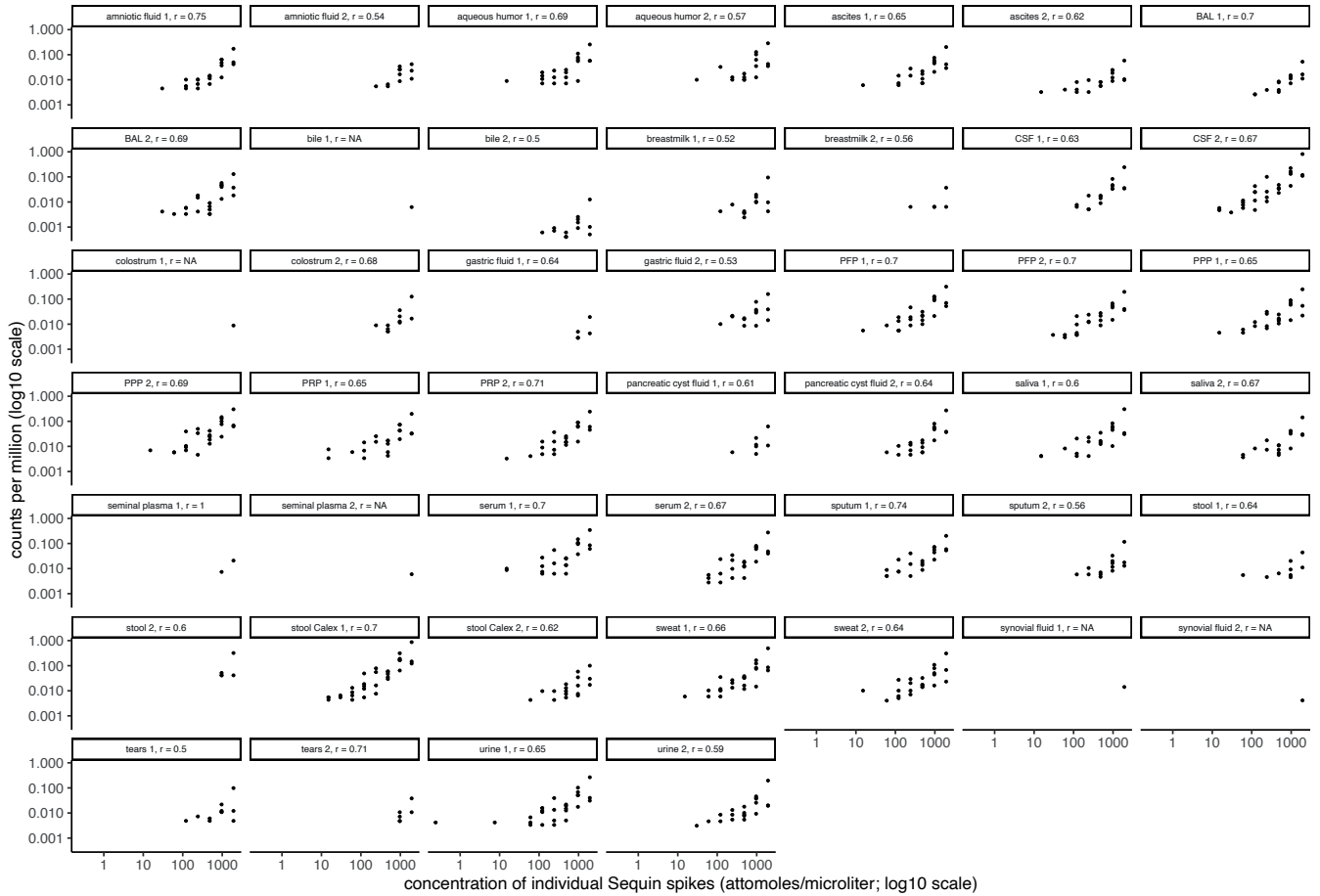
Supplemental Information

Charting Extracellular Transcriptomes

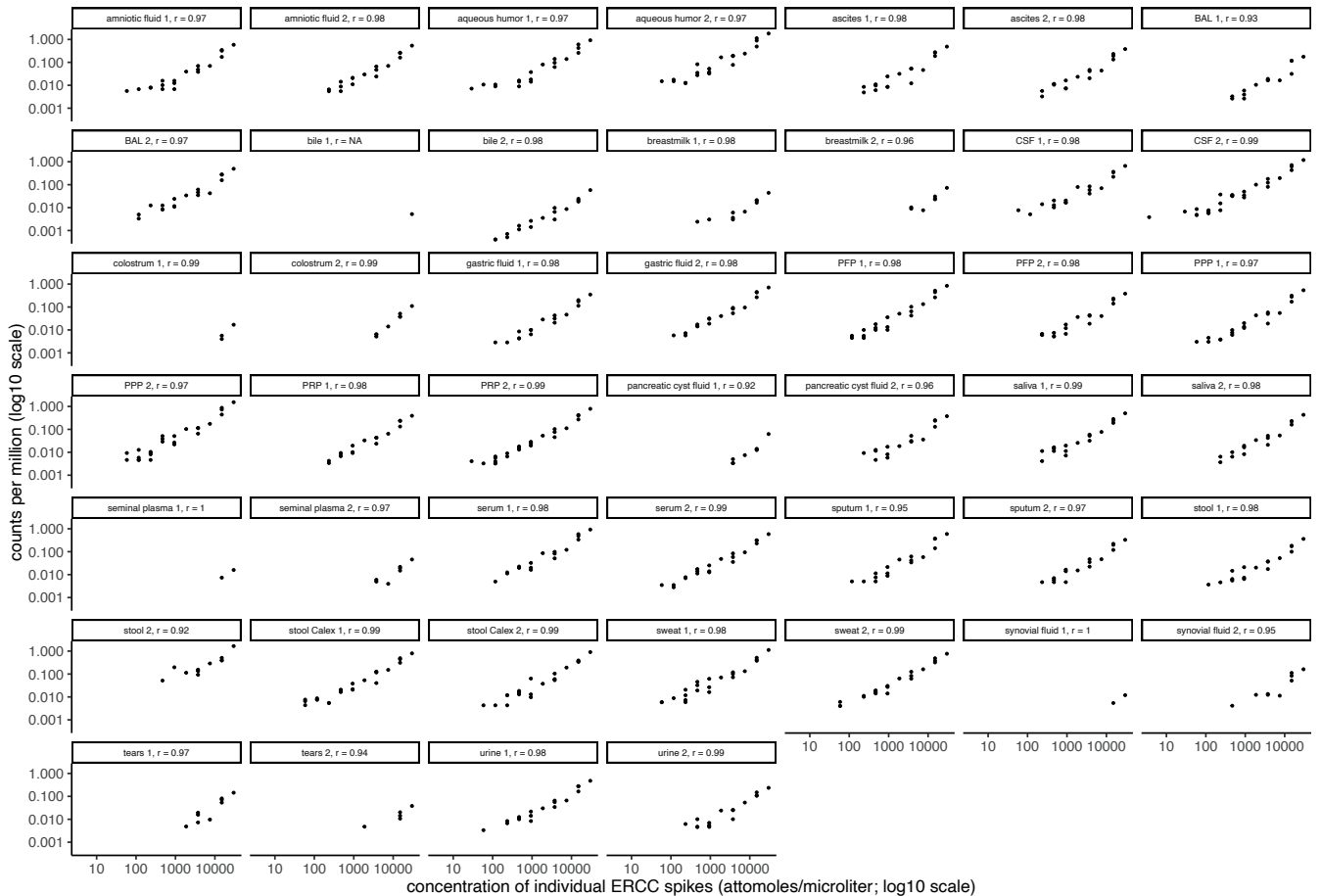
in The Human Biofluid RNA Atlas

Eva Hulstaert, Annelien Morlion, Francisco Avila Cobos, Kimberly Verniers, Justine Nuytens, Eveline Vanden Eynde, Nurten Yigit, Jasper Anckaert, Anja Geerts, Pieter Hindryckx, Peggy Jacques, Guy Brusselle, Ken R. Bracke, Tania Maes, Thomas Malfait, Thierry Derveaux, Virginie Ninclaus, Caroline Van Cauwenbergh, Kristien Roelens, Ellen Roets, Dimitri Hemelsoet, Kelly Tilleman, Lieve Brochez, Scott Kuersten, Lukas M. Simon, Sebastian Karg, Alexandra Kautzky-Willers, Michael Leutner, Christa Nöhammer, Ondrej Slaby, Roméo Willinge Prins, Jan Koster, Steve Lefever, Gary P. Schroth, Jo Vandesompele, and Pieter Mestdagh

Sequin spikes in the discovery cohort – Pearson correlation coefficient is shown per sample

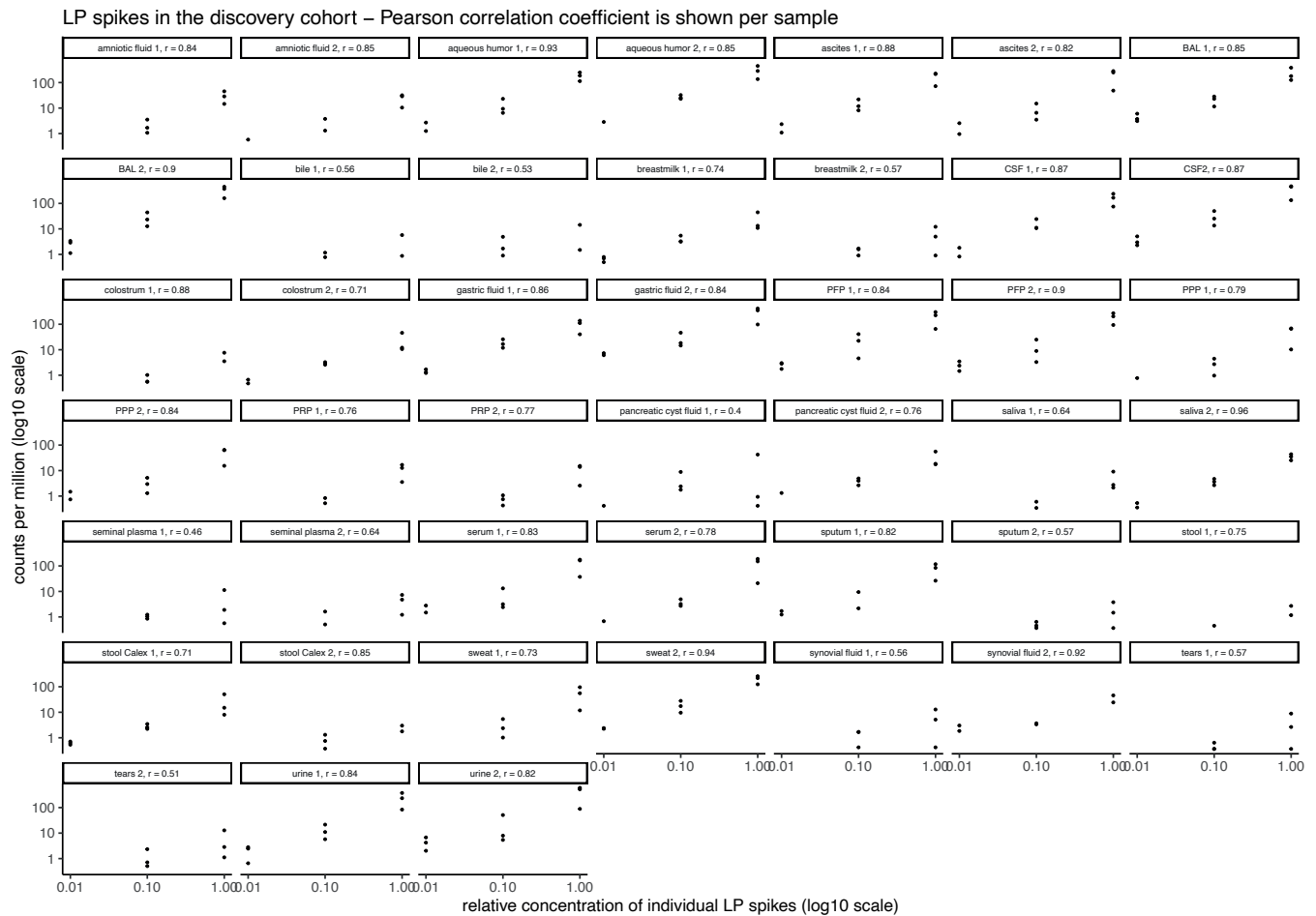
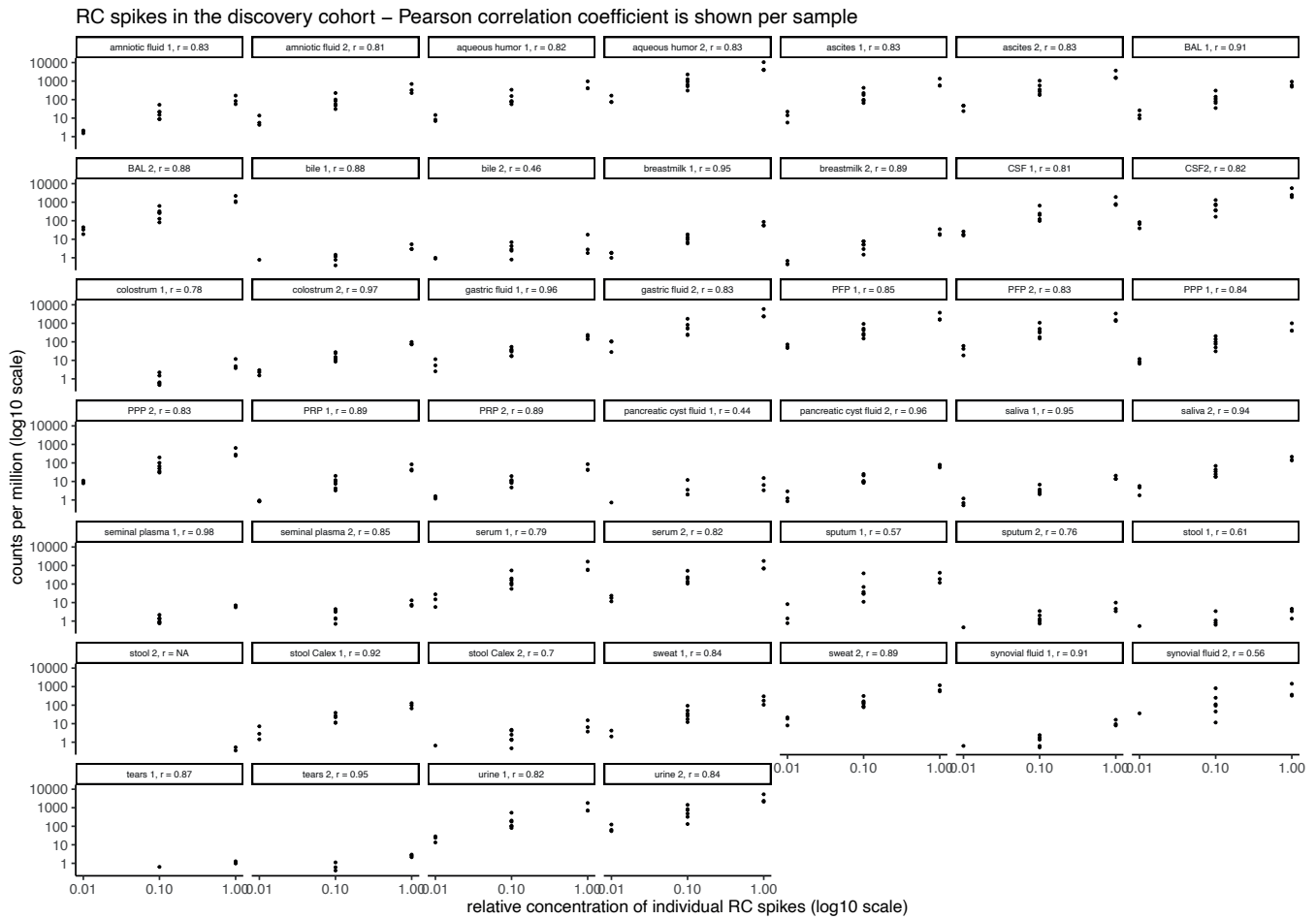


ERCC spikes in the discovery cohort – Pearson correlation coefficient is shown per sample



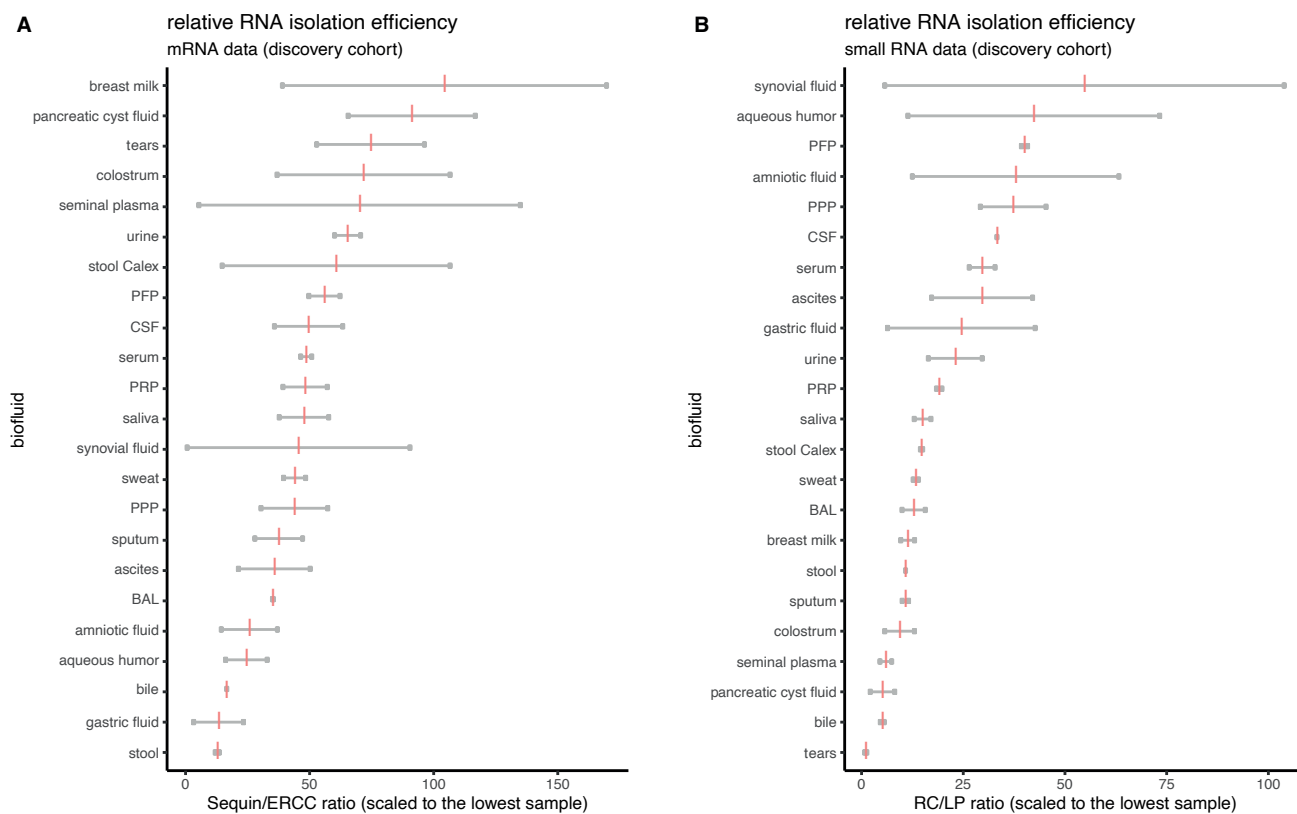
Supplementary Figure 1. In the discovery cohort the expected and the observed relative quantities are correlated within each spike type, Related to Figure 2

Correlation plots per spike type and per sample are shown. The Pearson correlation coefficient is shown per sample.



Supplementary Figure 2. In the discovery cohort the expected and the observed relative quantities are correlated within each spike type, Related to Figure 2

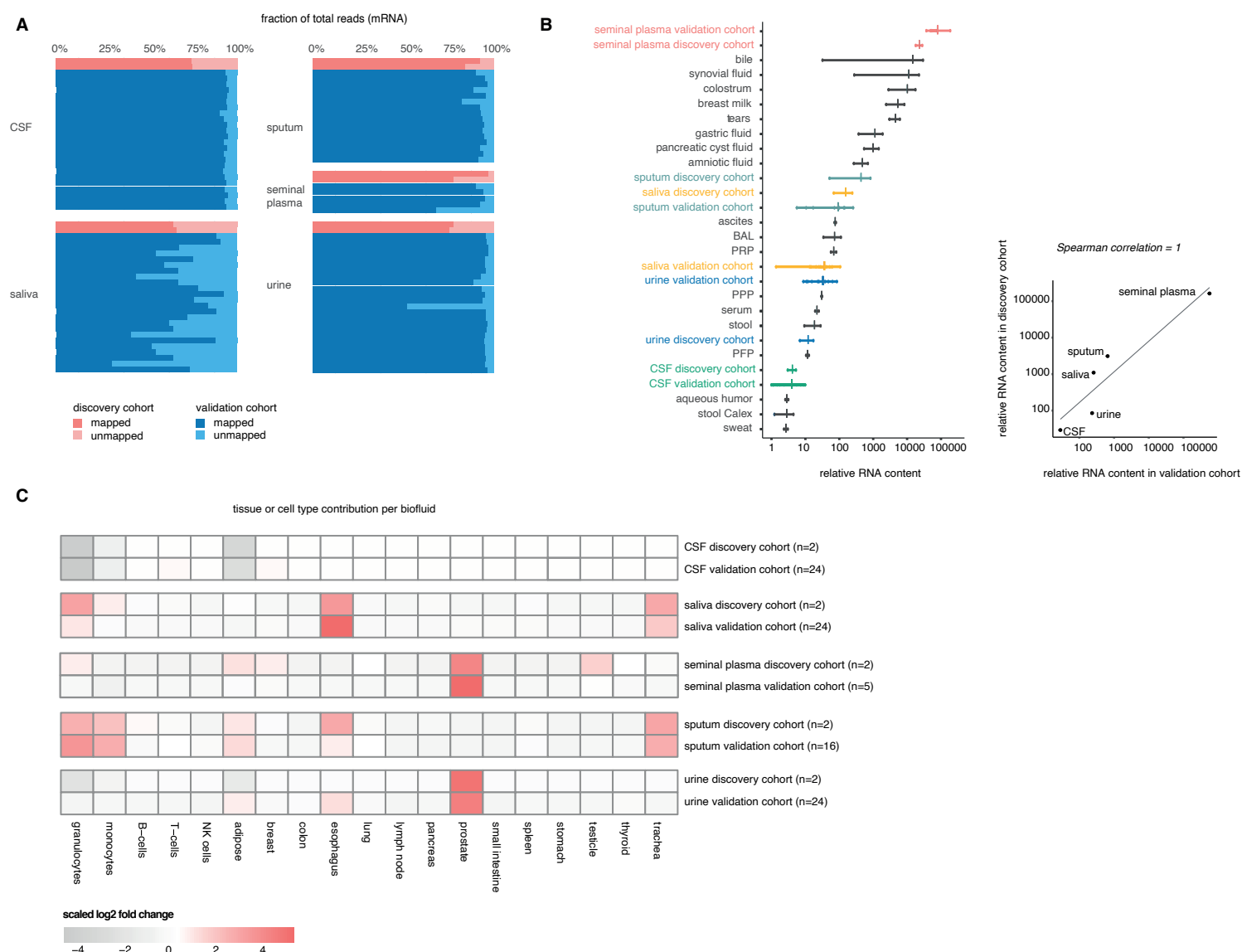
Correlation plots per spike type and per sample are shown. The Pearson correlation coefficient is shown per sample.



Supplementary Figure 3. The relative RNA isolation efficiency in the discovery cohort varies across the different biofluids, Related to Figure 2

(A) The Sequin/ERCC ratio reflects the relative RNA isolation in the mRNA capture sequencing data and is plotted per biofluid. A minimal read count threshold is set on 15 reads per spike type. Each grey dot represents a sample. The pink vertical bar reflects the mean of both samples within a biofluid.

(B) The RC/LP ratio reflects the relative RNA isolation in the mRNA capture sequencing data and is plotted per biofluid. A minimal read count threshold is set on 15 reads per spike type. Each grey dot represents a sample. The pink vertical bar reflects the mean of both samples within a biofluid.



Supplementary Figure 4. Mapping percentage, relative RNA content and tissue of origin analysis in additional CSF, saliva, sputum, seminal plasma and urine samples to illustrate the generalizability of the findings regarding the mRNA content observed in the discovery cohort, Related to Figure 2 and Figure 5

(A) Percentage of the total read count mapping to the human transcriptome for CSF, saliva, sputum, seminal plasma and urine samples. Samples from the discovery cohort (n=2 per biofluid) are shown in pink. Samples from the validation cohort are shown in blue: CSF from 12 patients with hydrocephalus and 12 patients with glioblastoma, saliva from 12 healthy donors and 12 diabetes patients, sputum from 8 healthy donors and 8 COPD patients, seminal plasma from 5 healthy donors, urine from 12 healthy donors and 12 bladder cancer patients.

(B) Relative RNA concentration per biofluid; every dot represents the relative RNA concentration in one sample, every vertical mark indicates the mean per biofluid. Biofluids in black represent samples from the discovery cohort. For biofluids in color additional samples were analyzed: seminal plasma from 5 healthy donors, sputum from 8 healthy donors, saliva from 12 healthy donors, urine from 12 healthy donors and CSF from 12 patients with hydrocephalus. Correlation plot of the relative RNA concentration in the discovery cohort and in the validation cohort for seminal plasma, sputum, saliva, urine and CSF is shown. The Spearman correlation coefficient is 1 (p-value = 0.017).

(C) Assessment of the tissues of origin in CSF, saliva, seminal plasma, sputum and urine. Findings from the discovery cohort are validated in independent, additional samples. Heatmap showing tissues and cell types that contribute more specifically to a certain biofluid compared to the other biofluids. Rows depict the biofluids of the discovery cohort and the columns are the tissues or cell types for which markers were selected based on the RNAAtlas32. For visualization purposes, only tissues and cell types with a z-score transformed log2 fold change $\geq |1|$ in at least one biofluid are shown.

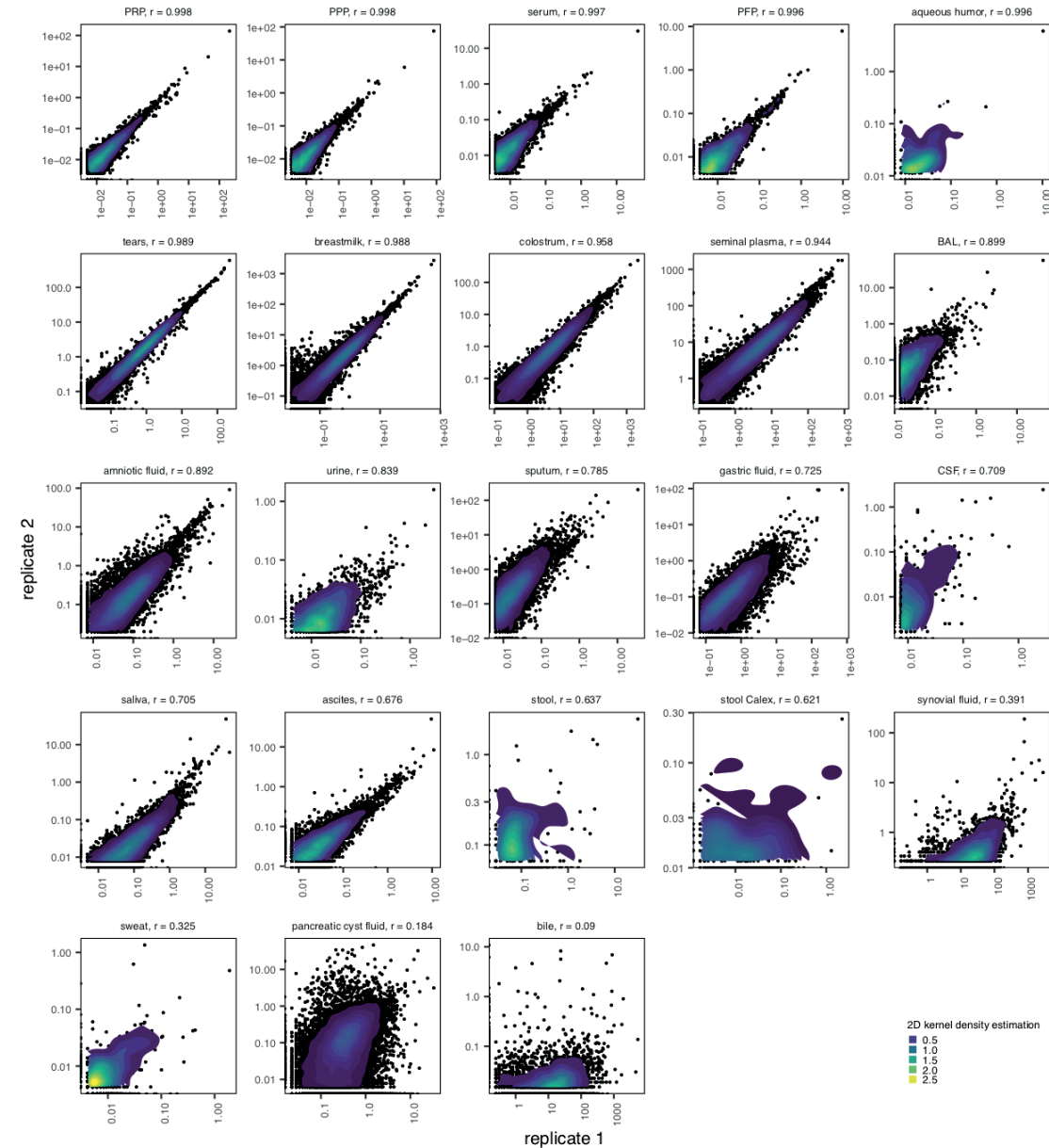
Supplementary Figure 5. Correlation between samples of the same biofluid

(A) Correlation plots of the mRNA data between two samples of the same biofluid are shown. The Pearson correlation coefficient is shown per biofluid.

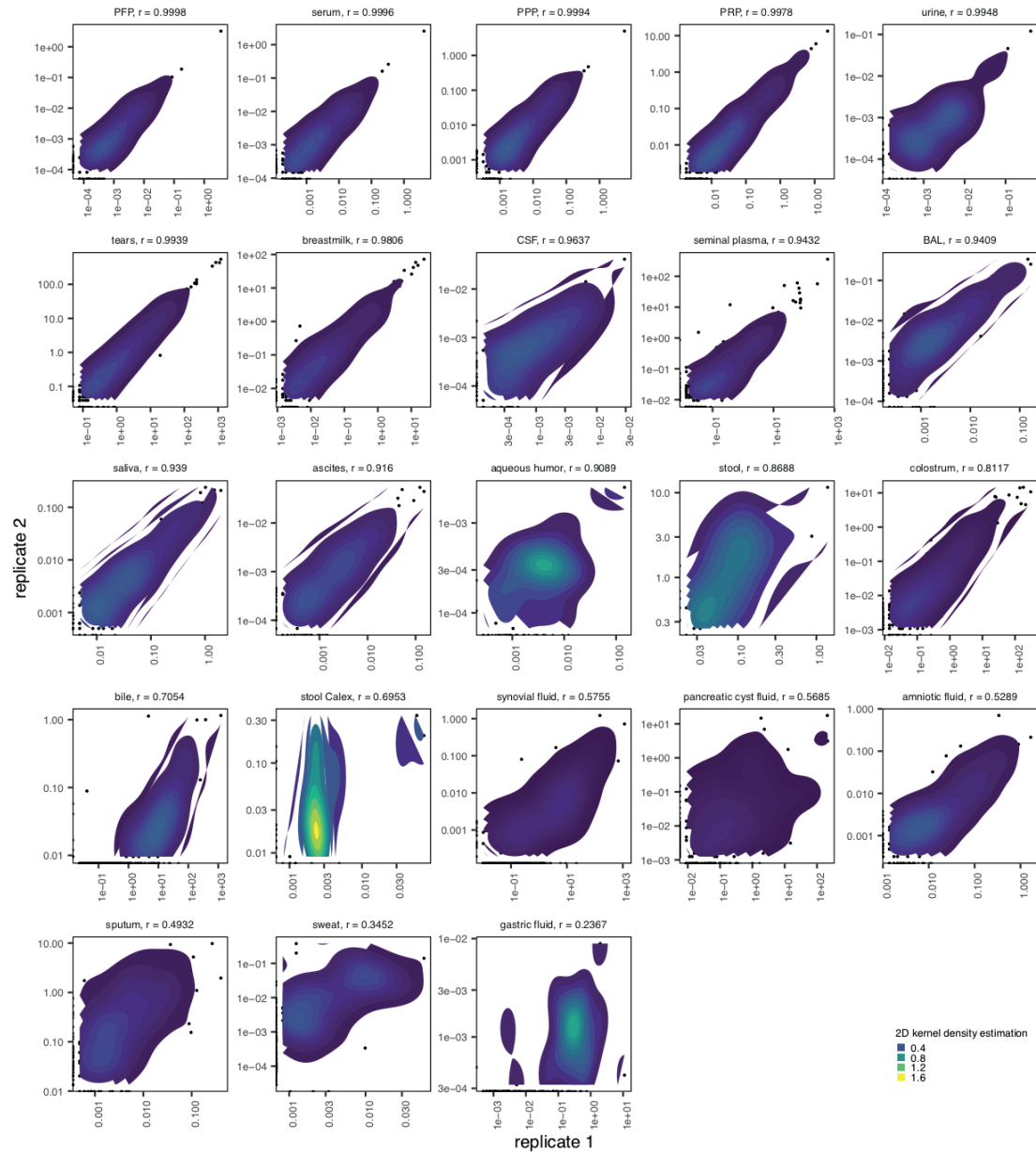
(B) Correlation plots of the miRNA data between two samples of the same biofluid are shown. The Pearson correlation coefficient is plotted per biofluid.

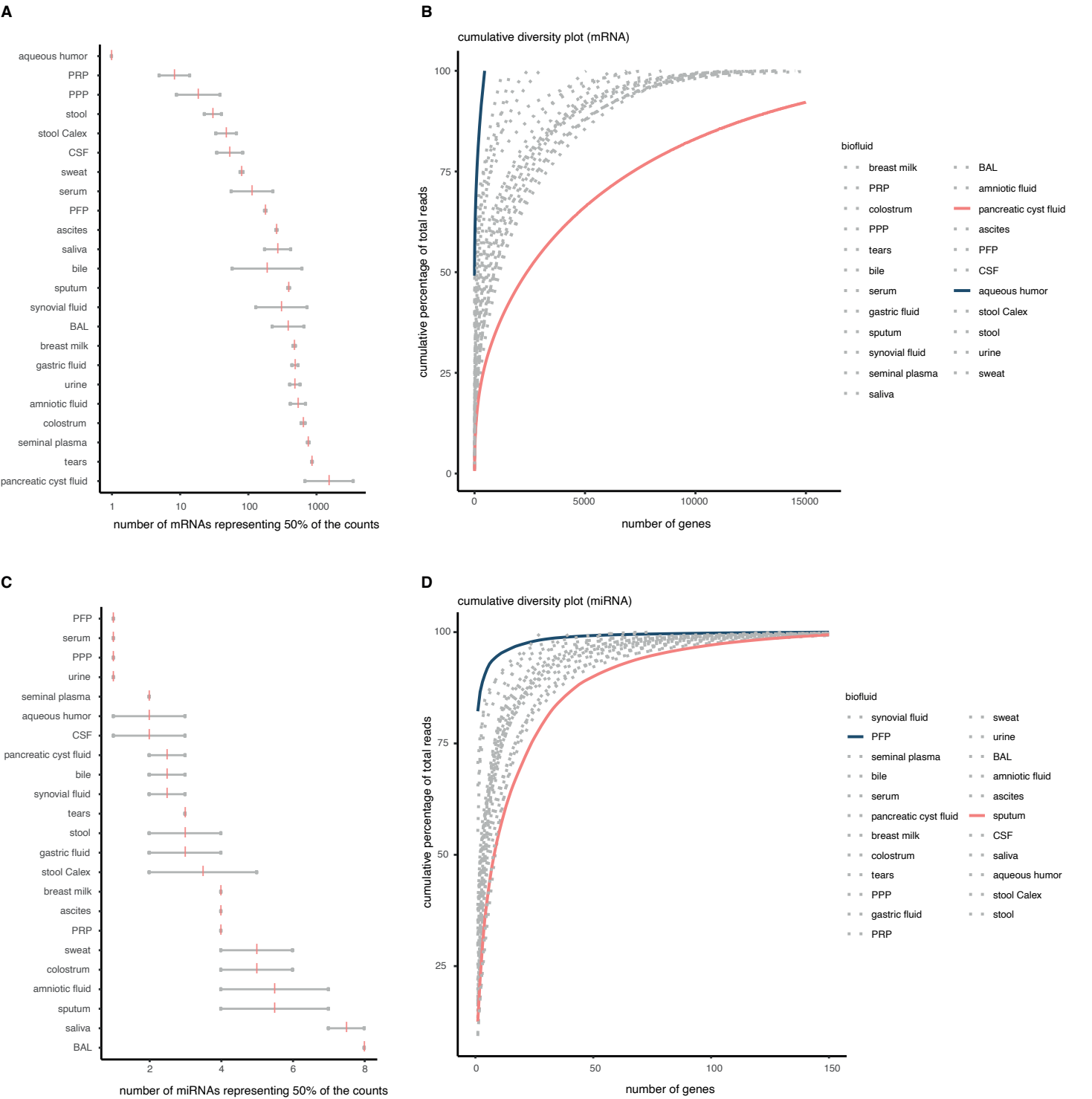
BAL, bronchoalveolar lavage fluid; CSF, cerebrospinal fluid; HNSCC: head and neck squamous-cell carcinoma; PFP, platelet-free plasma; PPP, platelet-poor plasma; PRP, platelet-rich plasma

A Correlation between samples of the same biofluid mRNA data, Pearson correlation coefficient



B Correlation between samples of the same biofluid miRNA data, Pearson correlation coefficient

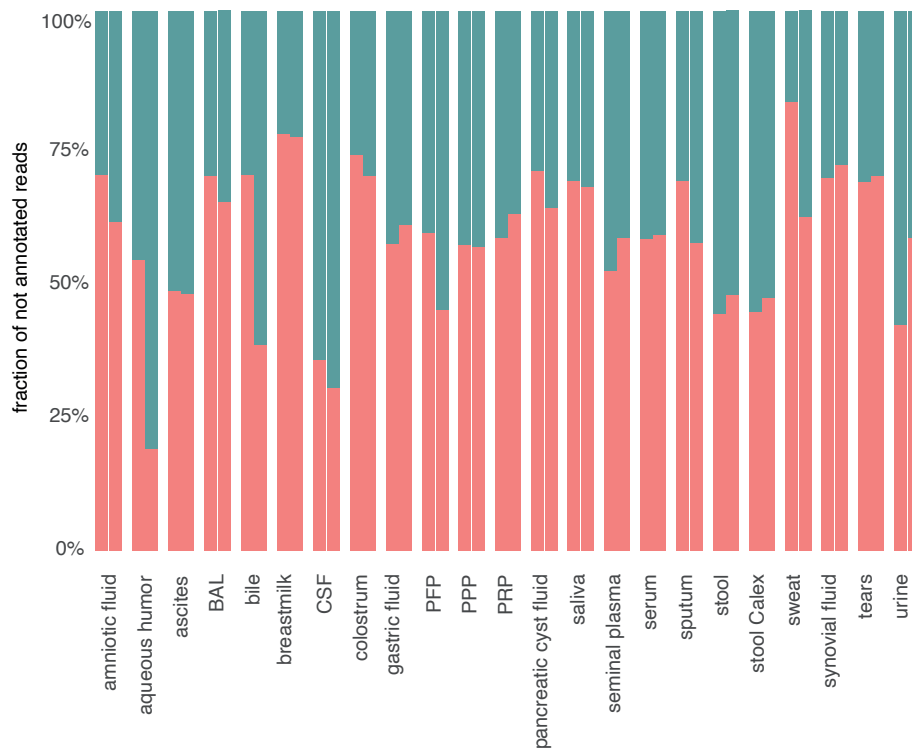




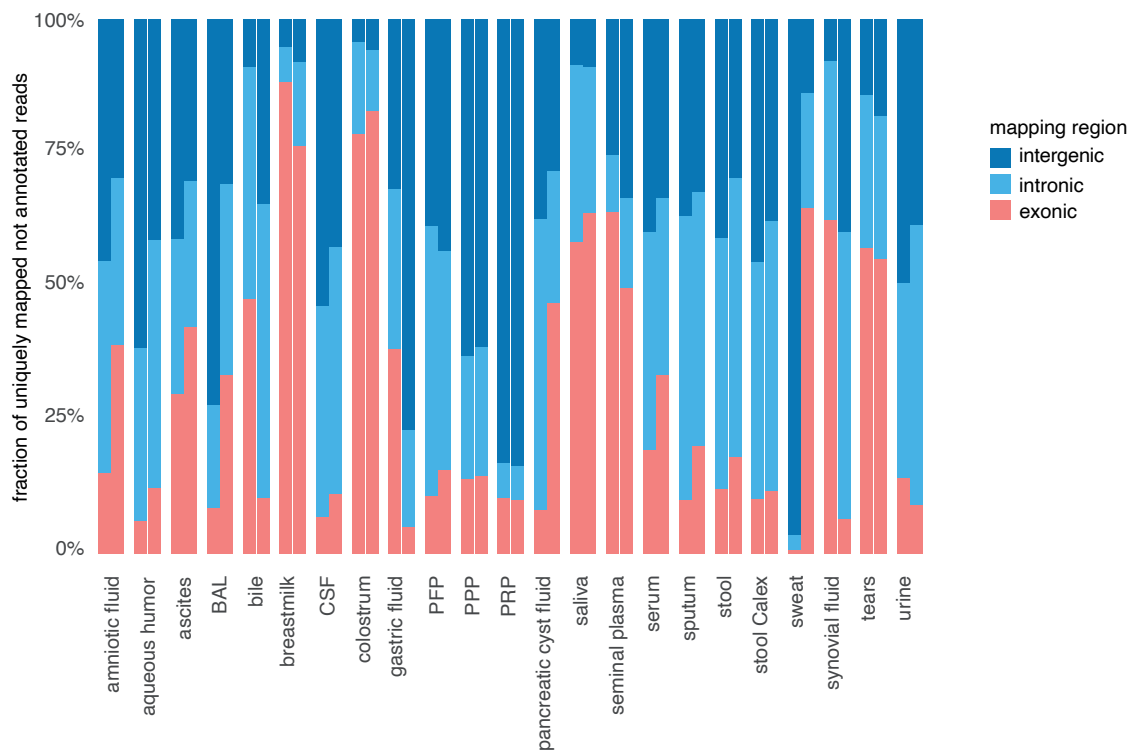
Supplementary Figure 6. Additional RNA diversity metrics, Related to Figure 2C

- (A) The number of mRNAs representing 50% of the counts is shown per biofluid. Each gray dot represents a sample, the pink vertical bar represents the mean value within a biofluid.
- (B) Cumulative diversity plot for the mRNA capture sequencing data.
- (C) The number of miRNAs representing 50% of the counts is shown per biofluid. Each gray dot represents a sample, the pink vertical bar represents the mean value within a biofluid.
- (D) Cumulative diversity plot for the miRNA data.

A



B

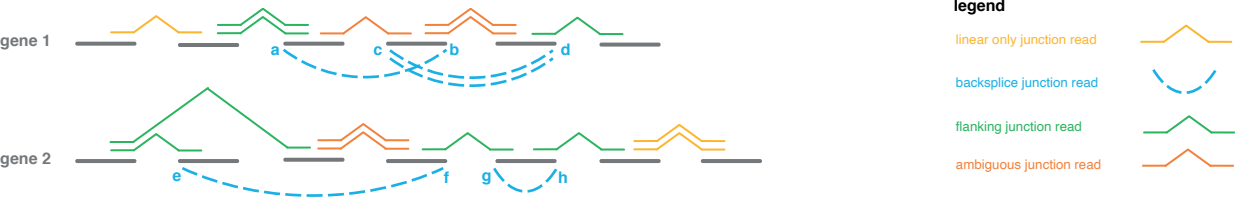


Supplementary Figure 7. Assessment of the origin of the not annotated reads in the small RNA seq data of the discovery cohort, Related to Figure 2E

(A) Fraction of the not annotated reads that uniquely map to the transcriptome is shown per biofluid

(B) Fractions of the uniquely mapped not annotated reads that map to exonic regions, intronic regions and intergenic regions are shown per biofluid

A



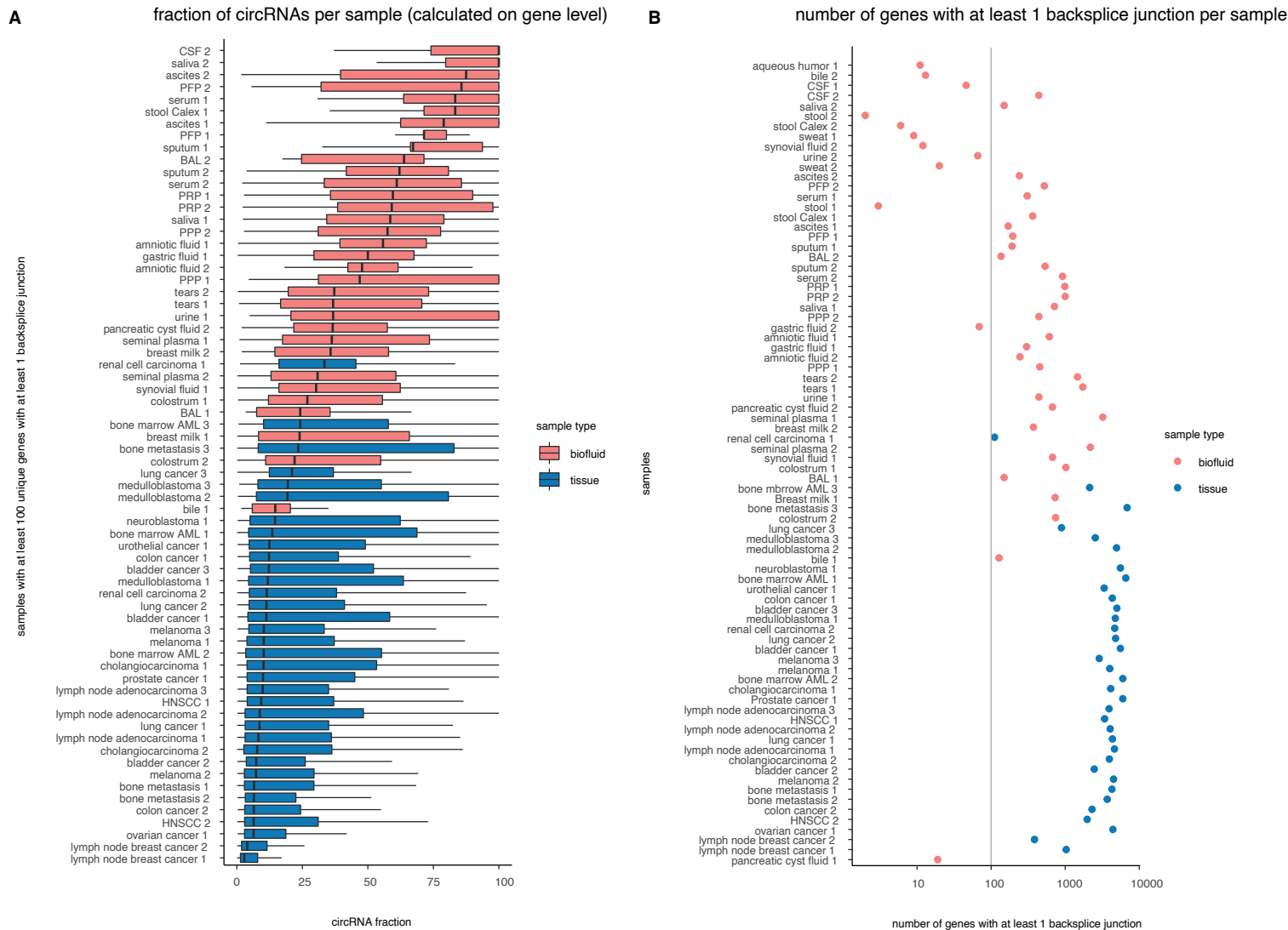
B

	backsplice junction-level method		gene-level method
gene 1 circ_a_b	<div>CIRC = backsplice junction read = 1</div> <div>LIN = average read count of all junctions flanking the backsplice junction of interest = $\frac{2}{2} + \frac{0}{1} = 2$</div> <div>circRNA fraction = 33%</div>	gene 1	<div>CIRC = average number of backsplice junction reads for a given gene = $\frac{3}{2} = 1.5$</div> <div>LIN = average number of linear junction reads for a given gene = $\frac{1}{1} = 1$</div> <div>circRNA fraction = 60%</div>
gene 2 circ_e_f	<div>CIRC = backsplice junction read = 1</div> <div>LIN = average read count of all junctions flanking the backsplice junction of interest = $\frac{2}{2} + \frac{1}{1} = 1$</div> <div>circRNA fraction = 50%</div>	gene 2	<div>CIRC = average number of backsplice junction reads for a given gene = $\frac{2}{2} = 1$</div> <div>LIN = average number of linear junction reads for a given gene = $\frac{2}{1} = 2$</div> <div>circRNA fraction = 33%</div>

Supplementary Figure 8. The circular RNA fraction was assessed using two different methods: the backsplice junction-level method and the gene-level method, Related to Figure 3

(A) Two genes and their different detected junction reads are shown to illustrate the terminology used throughout the manuscript. Junction reads that can only originate from linear transcripts (linear junction reads) are shown in yellow. Backsplice junction reads (shown in blue) can only be part of circular transcripts. Junction reads flanking a backsplice junction read are categorized as flanking junction read (shown in green). Junction reads that can be part of both linear and circular transcripts are referred to as ambiguous junction reads (shown in orange). Ambiguous junction reads are not taken into account for the further analyses.

(B) The circRNA fraction is defined as $100 \times \text{CIRC} / (\text{CIRC} + \text{LIN})$. As an example the calculation of CIRC and LIN according to the backsplice junction-level method and the gene-level method is given below the figure.

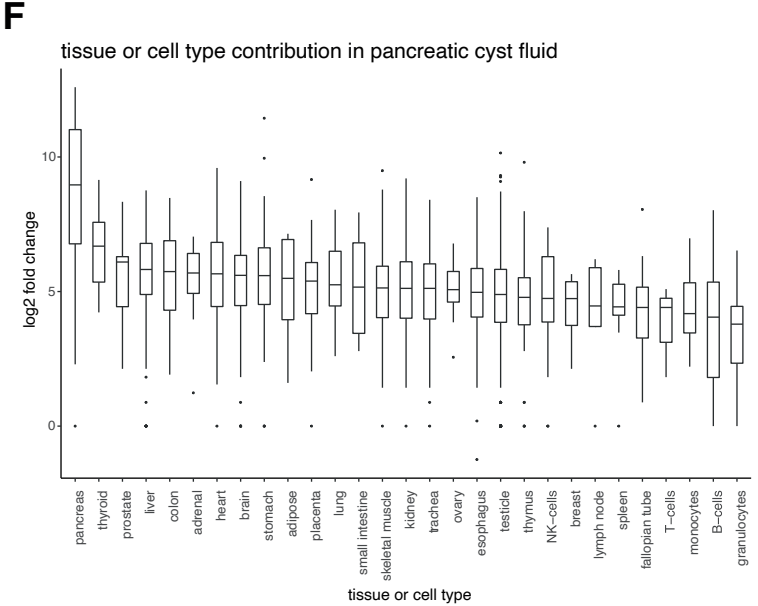
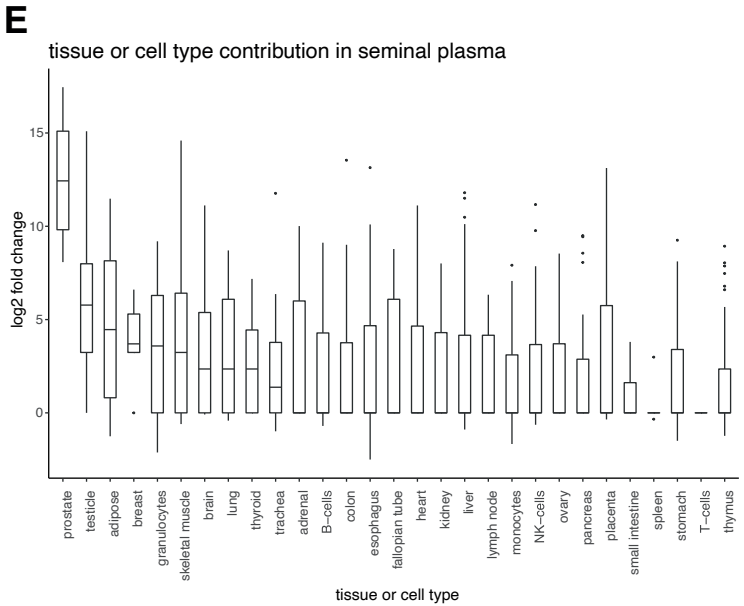
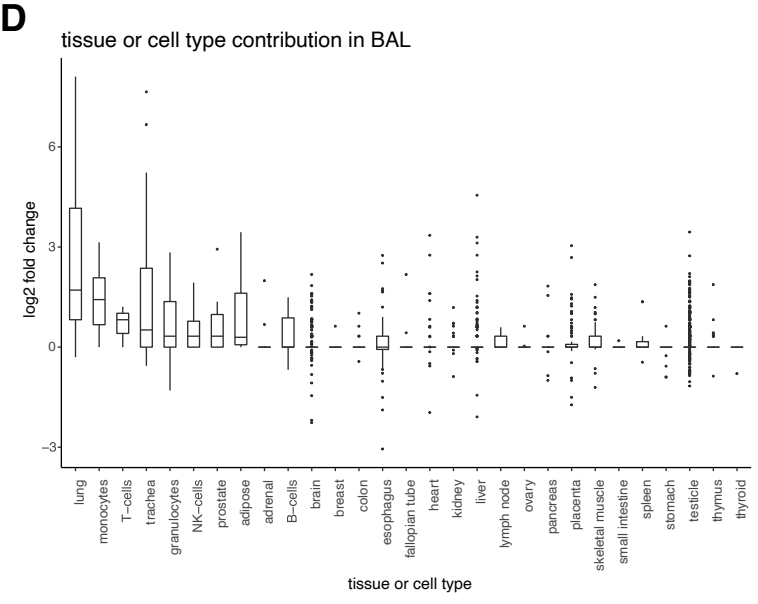
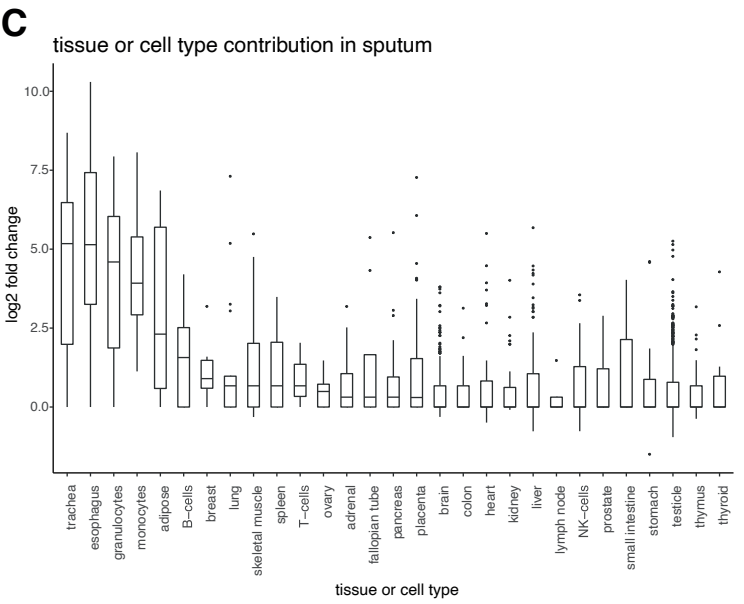
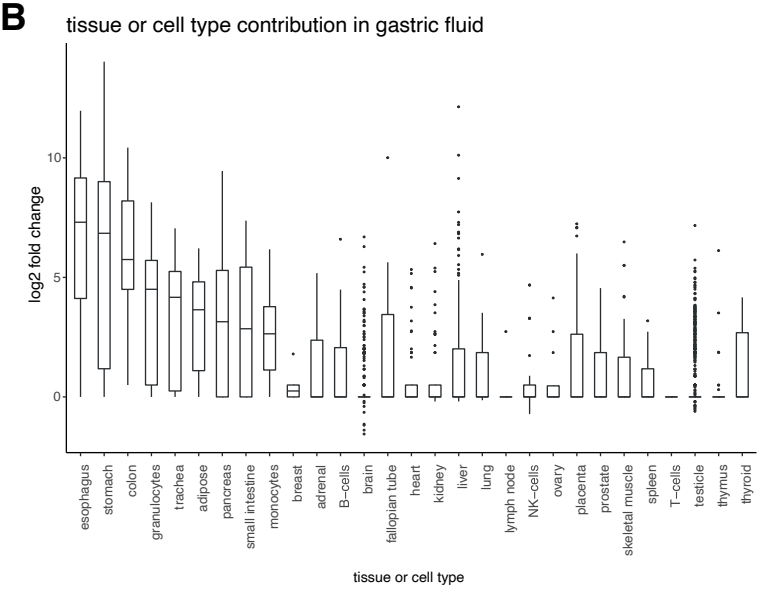
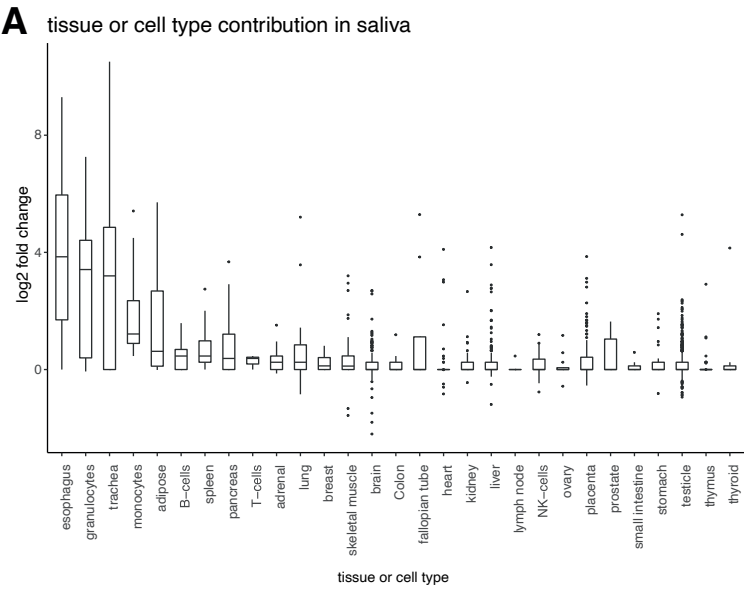


Supplementary Figure 9. CircRNAs are enriched in biofluids compared to tissues, Related to Figure 3

(A) The circRNA fraction, calculated on gene level, is plotted per sample and is higher in cell-free biofluid samples than in tissue samples. Only samples with at least 100 backsplice junctions are plotted.

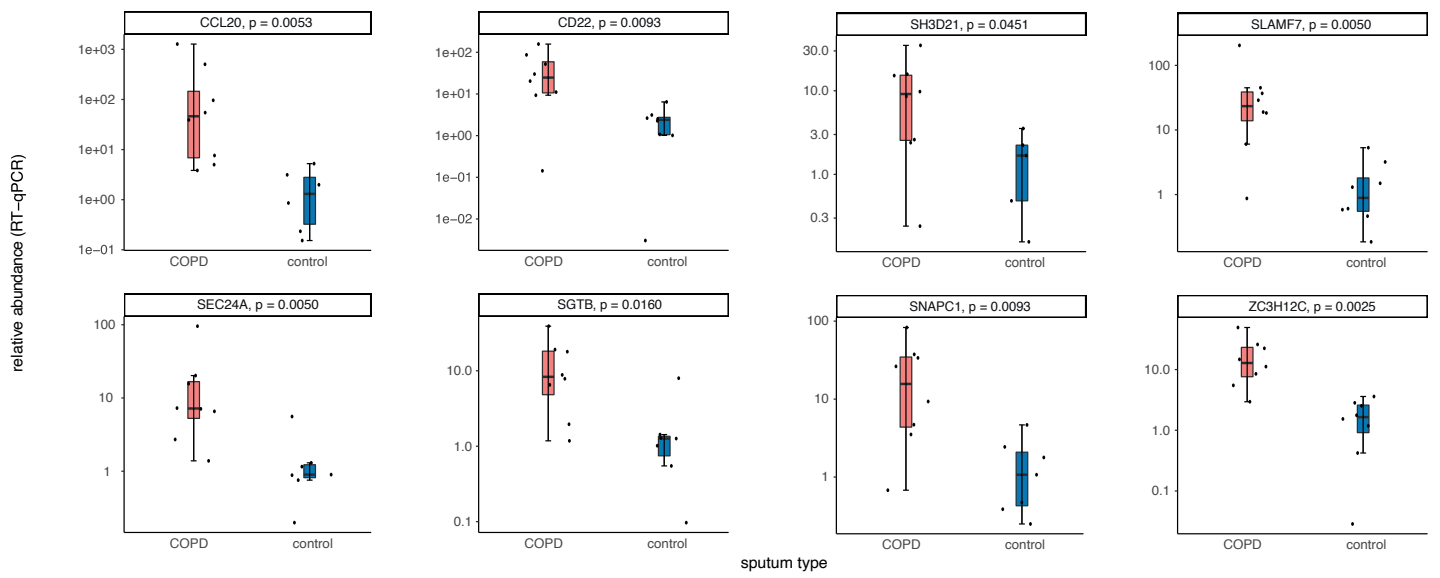
(B) The number of genes with at least 1 backsplice junction per sample is higher in tissue samples compared to biofluid samples, in line with the higher input concentration of RNA.

AML, acute myeloid leukemia; BAL, bronchoalveolar lavage fluid; CSF, cerebrospinal fluid; HNSCC: head and neck squamous-cell carcinoma; PFP, platelet-free plasma; PPP, platelet-poor plasma; PRP, platelet-rich plasma

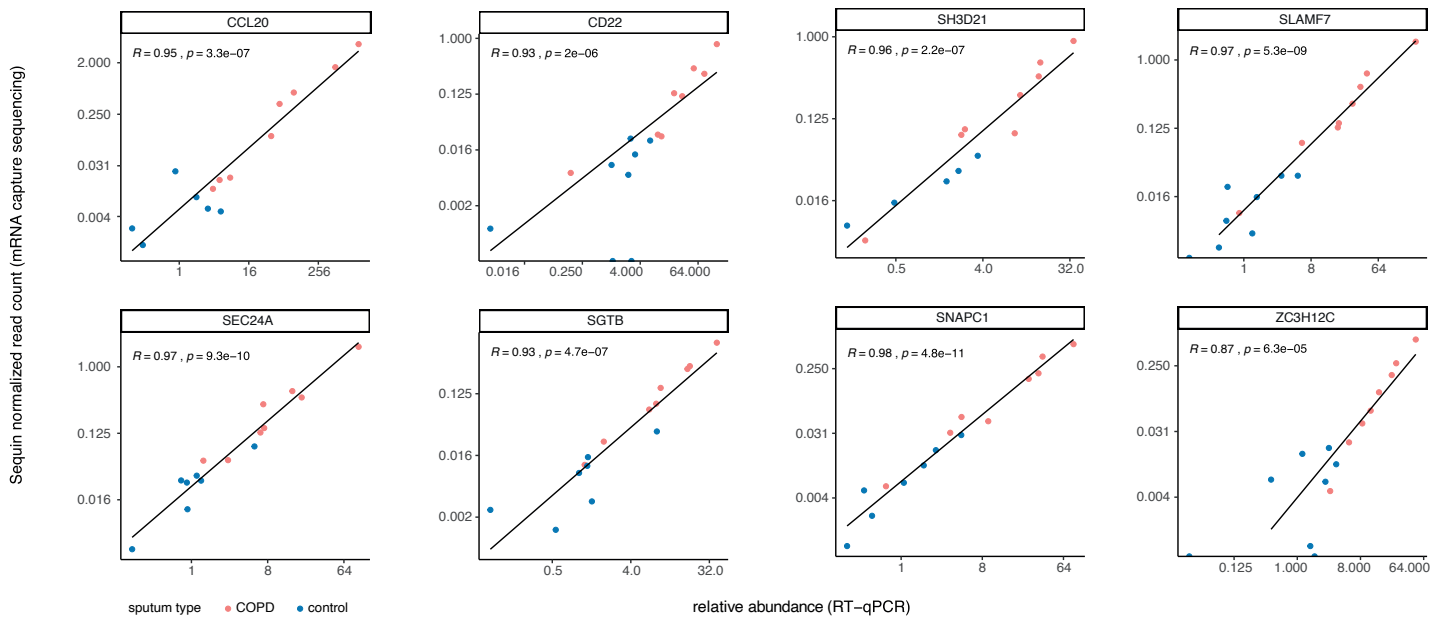


Supplementary Figure 10. Tissue or cell type contribution in selected biofluids from the discovery cohort, Related to Figure 5
Boxplots showing the log2 fold change for a gene set with markers specific for a certain tissue or cell type. The log2 fold change is calculated between the median read count for both samples of 1 biofluid (indicated in the title of each plot) and the median read count of all other biofluids. The tissues or cell types for which markers were selected based on the RNA Atlas Project are shown on the x-axis.
(A) Tissue or cell type contribution in saliva
(B) Tissue or cell type contribution in gastric fluid
(C) Tissue or cell type contribution in sputum
(D) Tissue or cell type contribution in bronchoalveolar lavage fluid (BAL)
(E) Tissue or cell type contribution in seminal plasma
(F) Tissue or cell type contribution in pancreatic cyst fluid

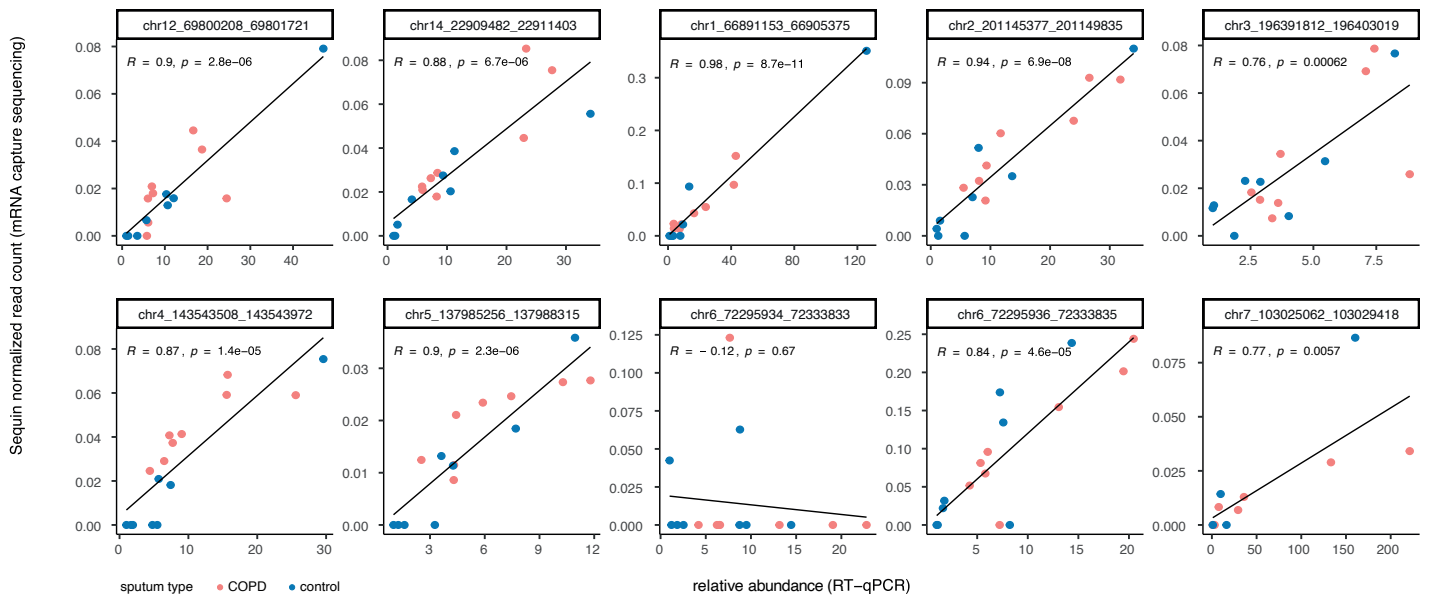
A relative abundance (RT-qPCR) for the most differentially expressed mRNAs in sputum



B orthogonal validation of the most differentially expressed mRNAs in sputum by RT-qPCR



C orthogonal validation of the most abundant circRNAs in sputum by RT-qPCR

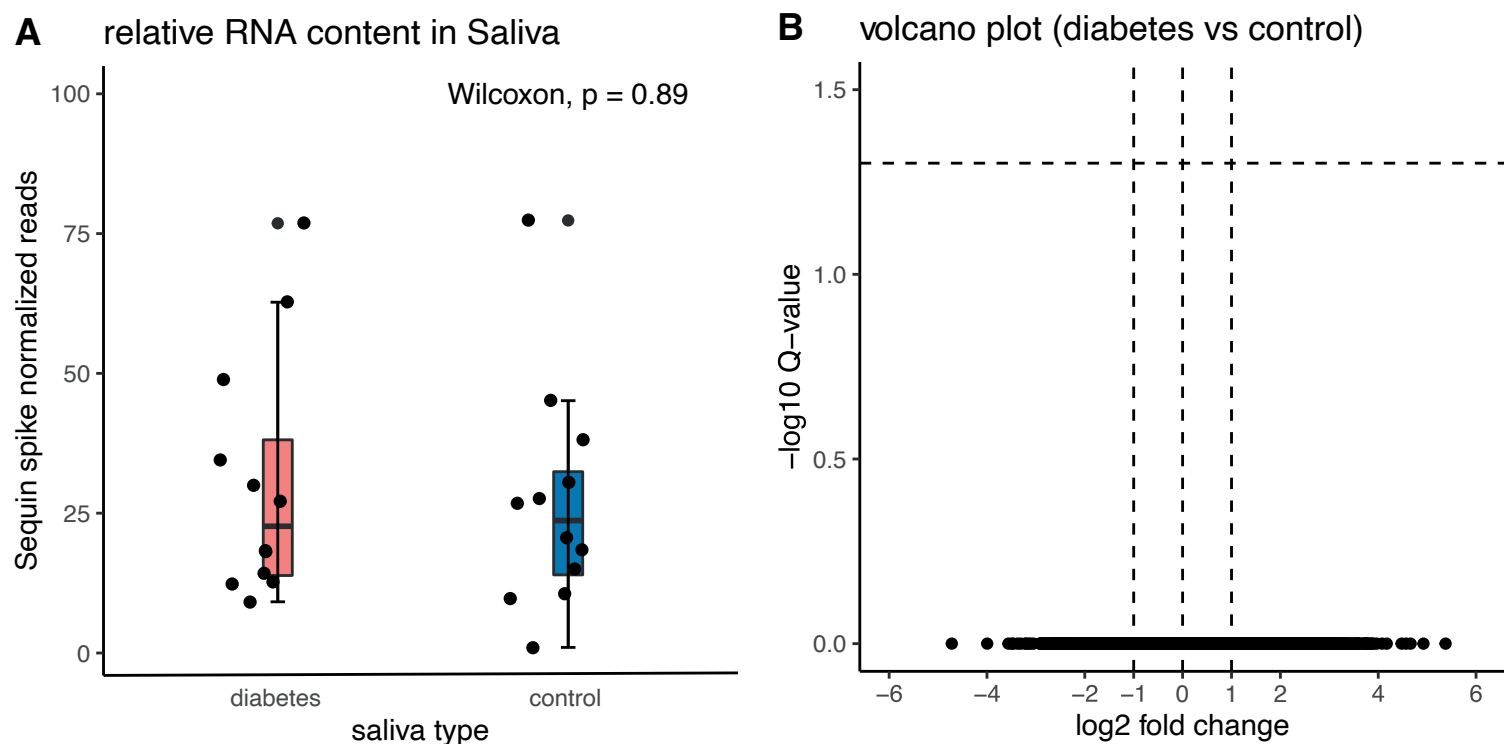


Supplementary Figure 11. Validation of the expression of the most differentially expressed mRNAs and the most abundant circRNAs in sputum by RT-qPCR, Related to Figure 6A-B

(A) Relative RNA expression of 8 mRNAs based on RT-qPCR in sputum samples of the case- control cohort. Based on RNA sequencing data, all 8 mRNAs are upregulated in sputum from COPD patients compared to sputum from healthy donors. These observations could be validated with RT-qPCR. For each mRNA the p-value for a two-sided Mann-Whitney U test with multiple testing correction using the Benjamini Hochberg method is shown.

(B) Correlation plots of the Sequin normalized read counts and the relative expression based on RT-qPCR for 8 differentially expressed mRNAs in sputum samples of the case-control cohort. The Pearson correlation coefficient and the p-value are shown.

(C) Correlation plots of the Sequin normalized read counts and the relative expression based on RT-qPCR for 10 abundant circRNAs in sputum samples of the case-control cohort. The Pearson correlation coefficient and the p-value are shown.



Supplementary Figure 12. Relative RNA content and volcano plot in the saliva case/control cohort, Related to Figure 6

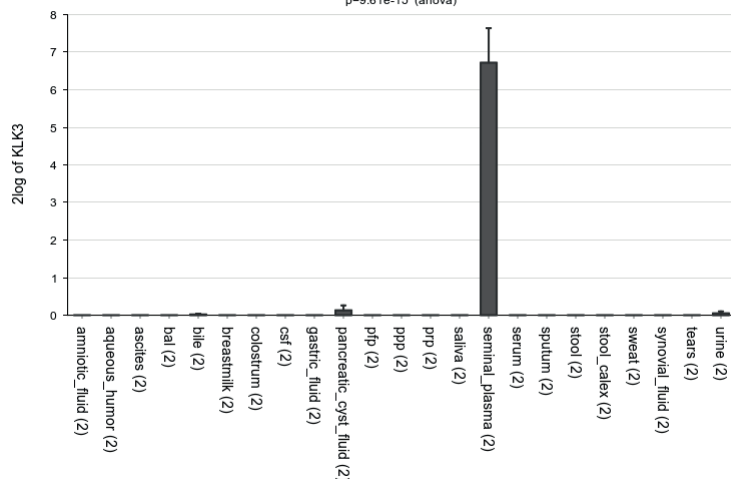
(A) The relative RNA content (Sequin normalize reads) is comparable in saliva from diabetes patients ($n = 12$) and saliva from healthy donors ($n = 12$; Wilcoxon rank test, two-sided, $p = 0.89$).

(B) Volcano plot of differentially expressed mRNAs between saliva samples of diabetes patients and saliva samples of healthy volunteers at $q < 0.05$. No differentially expressed genes were identified.

A

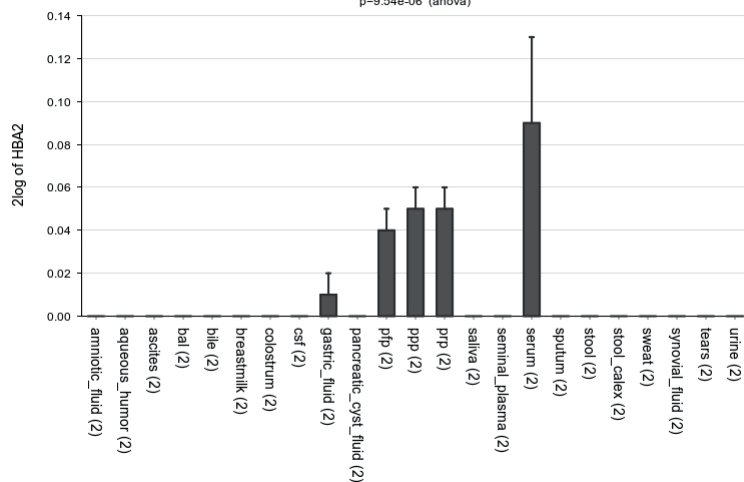
Mixed Human Biofluid RNA Atlas discovery cohort (mRNA) - Mestdagh - 46

seqnorm - hsens91
NG_biofluid vs KLK3 (ENSG00000142515)
p=9.61e-15 (anova)



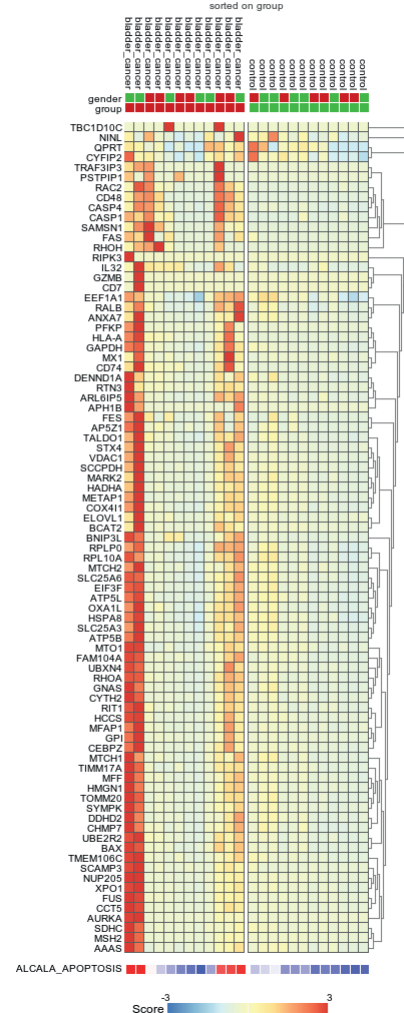
Mixed Human Biofluid RNA Atlas discovery cohort (mRNA) - Mestdagh - 46

seqnorm - hsens91
NG_biofluid vs HBA2 (ENSG00000188536)
p=9.54e-06 (anova)



B

seqnorm - hsens91
sample_slave: db:geneset_broad_2019_c2_curated
ALCALA_APOPTOSIS
sorted on group



Supplementary Figure 13. Examples illustrating the applicability of the interactive R2 platform, Related to STAR Methods

Normalized sequencing data can readily be explored in the interactive R2 platform. R2 allows the analysis and visualization of mRNA, circRNA and miRNA abundance. All samples can be used for correlations, principal component analyses, gene set enrichment analysis and many more. Two examples are shown in this panel.

(A) Abundance of KLK3 and HBA2 in the discovery cohort. Prostate-specific antigen is encoded by KLK3. KLK3 is highly abundant in seminal plasma. HBA2 mRNA, coding for the alpha globin chain of hemoglobin, is highly abundant in serum, platelet-rich plasma, platelet-poor plasma and platelet-free plasma.

(B) A heatmap can be generated with a geneset of interest in a specific sample cohort, e.g. the abundance of genes of the ALCALA_APOPTOSIS geneset are plotted per sample of the urine case/control cohort.

Methods S1: Details on the RNA spike-in controls and details on the sample collection protocol per biofluid, Related to STAR Methods

Sequin and external RNA Control Consortium (ERCC) spike-in controls

In the messenger and circular RNA sequencing workflow, Sequin spike-in controls (Garvan Institute of Medical Research, mix A) are added to the biofluid lysate prior to further RNA purification. An overview of the Sequin spikes with their stock concentration is provided in Table S1. A volume of 2 µL Sequin spike-in controls at a dilution of 1/1 300 000 was added to 200 µL of all fluid samples, except tears for which an extra pooling step is introduced during the RNA isolation. To enable a fair comparison over all fluids, 2 µL Sequin spike-in controls are added at a dilution of 1/2 600 000 in the tear samples before RNA isolation so that the final pooled RNA samples from tears had the same Sequin amount as the other samples. ERCC spike-in controls (ThermoFisher Scientific, 4456740, mix 1) are added to the RNA eluate before gDNA removal. In all fluid samples, 2 µL ERCC spike-in controls at a dilution of 1/1 000 000 are added to 12 µL RNA eluate. An overview of the ERCC spikes with their stock concentration is provided in Table S1.

RNA extraction Control (RC) and Library Prep control (LP) spike-in controls

In the small RNA sequencing workflow, RNA extraction Control (RC) spike-ins are added to the biofluid lysate prior to RNA isolation, and Library Prep control (LP) spike-in controls to the RNA eluate prior to library prep. All spike-in RNA molecules were ordered at Integrated DNA Technologies as 5'-monophosphate and 3'OH RNA oligonucleotides, and diluted in 500 nM carrier DNA oligo (TCGAAGTATTC). Prior to their use, individual spike-ins are balanced for differences in small RNA sequencing efficiency.

RC spike-in controls are a pool of RNA extraction dynamic range controls (25-mer oligoribonucleotides; RC1-01 - RC1-12; 3-points 10-fold-dilution calibration curve with 3

spikes per dilution) and RNA extraction size controls (25-, 28- or 34-mer oligoribonucleotide; RC2-25 - RC2-34; used at mid-point concentration of the calibration curve) selected from literature(1). RC spike-in IDs and sequences are listed in Table S1. A volume of 2 μ L from a 190 fM pool (i.e. concentration of highest concentrated spikes that are at 1x in the pool) of RC spikes was added to 200 μ L of the lysate before RNA isolation in all fluids, except tears. In tears, 2 μ L from a 96.4 fM pool of RC spikes was added before RNA isolation.

LP spike-ins are a pool of small RNA sequencing library prep controls (22-mer oligoribonucleotides; LP1-01 - LP1-12; 3-points 10-fold-dilution calibration curve with 3 spikes per dilution). LP spike-in IDs and sequences are listed in Table S1. A volume of 2 μ L from a 33.8 fM pool (i.e. concentration of highest concentrated spikes that are at 1x in the pool) of LP spikes was added to 12 μ L RNA eluate in all fluids.

Amniotic fluid

Amniotic fluid was collected in two donors undergoing a clinically-indicated amniocentesis.

In one of the donors, the amniocentesis was performed at 22 weeks pregnancy because of CMV seroconversion. No CMV was detected in the amniotic fluid. In the other donor, an amniocentesis was performed at 15 weeks pregnancy because of a *de novo* deletion in 16q22.3 in the firstborn of the parents.

In both donors, 15 mL amniotic fluid was collected in a 50 mL conical centrifuge tube (Falcon, Fisher Scientific). A volume of 1 mL of the amniotic fluid was transferred in a 15 mL conical centrifuge tube. These samples were centrifuged to remove cells (2000 x g, 10 min) and the supernatant was stored at -80 °C within 70 minutes after collection.

Aqueous humor

Aqueous humor was collected in 9 patients undergoing cataract surgery. Topical anesthesia with oxybuprocaine hydrochloride 0.4% was applied to the cornea and conjunctivae.

Following instillation of povidone-iodine 5% solution, aqueous humor was obtained through the introduction of a 30-gauge needle on a 1-cc tuberculin syringe onto the corneoscleral junction into the anterior chamber. A volume of 100-200 μL aqueous humor was collected until a shallow anterior chamber was observed. The aqueous humor of donor 1-4 was pooled together obtaining a total volume of 870 μL . The aqueous humor of donor 5-9 was pooled together obtaining a total volume of 850 μL . The two pooled samples were centrifuged to remove cells (2000 x g, 10 min) and the supernatant was stored at -80 °C within 120 minutes after collection.

Ascites (peritoneal fluid) collection

Ascites was collected using paracentesis in two patients with decompensated cirrhosis caused by alcohol (HBV and HCV negative). In each donor, 12 mL ascites was collected with a 20-gauge needle on a 20 mL syringe and transferred to a 15 mL conical centrifuge tube (Fisher Scientific). Both samples were centrifuged to remove cells (2000 x g, 10 min) and the supernatant was stored at -80 °C within 120 minutes after collection.

Bronchoalveolar lavage fluid

Bronchoalveolar lavage fluid (BAL) was collected from two patients during flexible bronchoscopy. In one of the donors bronchoalveolar lavage was performed to obtain microbiological samples for persisting bronchiolitis. In the other donor bronchoalveolar lavage was performed to investigate interstitial lung disease. No infectious focus was found in both samples. Normal saline (100-300 mL divided into three-five aliquots) at room temperature was instilled through the bronchoscope. After the instillation of each aliquot, instilled saline was retrieved using a negative suction pressure less than 100 mm Hg. All aliquots of the retrieved BAL fluid were pooled and 4-5 mL BAL was transferred to a 15 mL conical centrifuge tube (Fisher Scientific). Both samples were centrifuged to remove cells

(2000 x g, 10 min) and the supernatant was stored at -80 °C within 120 minutes after collection.

Bile

Bile was collected in two patients during an endoscopic retrograde cholangiopancreatography (ERCP) procedure. In each donor, 1 mL bile was collected in a sterile recipient before injection of contrast. In the first donor the ERCP was performed for stent exchange in the context of an anastomotic bile duct stricture after liver transplantation. In the second donor, the ERCP was performed for a hilar ischemic cholangiopathy after liver transplantation. All samples were centrifuged to remove cells (2000 x g, 10 min) and the cell-free supernatant was stored at -80 °C within 2 hours after collection.

Cerebrospinal fluid (discovery cohort)

CSF samples were obtained in two donors to rule out cerebral vasculitis. In both CSF samples normal results were obtained after routine clinical analysis, including the measurement of glucose, total protein content, lactate, lactate dehydrogenase, albumin, IgG, IgM, IgA, leukocytes (with differentiation), erythrocytes and isoelectric focusing. Microbiology cultures were negative in both samples. CSF was aspirated by lumbar puncture using a 25-gauge atraumatic (pencil-point) needle, and transferred to a 15 mL conical centrifuge tube (Fisher Scientific). All samples were centrifuged to remove cells (2000 x g, 10 min) and the cell-free supernatant was stored at -80 °C within 2 hours after collection.

Cerebrospinal fluid (case/control cohort)

Briefly, 4–6 mL of CSF was obtained during a lumbar puncture between the L3 and L5 vertebrae before surgical intervention in glioblastoma patients or during standard therapeutic management of patients with normal-pressure hydrocephalus (non-tumor

patients). CSF samples containing blood-derived cells were excluded. Subsequently, CSF samples were centrifuged at 500 g for 10 min at 4 °C, and the cell-free supernatant were aliquoted to 1 mL tubes and stored at –80 °C within one hour after collection.

Gastric fluid

Gastric fluid was collected during oesophagogastrosocopy in two patients with complaints of gastroesophageal reflux disease. The procedure was performed after a 12-hour overnight fast. A volume of 4 mL gastric fluid was aspirated using a 10 mL syringe immediately upon entering the stomach. Both samples were centrifuged to remove cells (2000 x g, 10 min) and the cell-free supernatant was stored at -80 °C within 2 hours after collection.

Mature breast milk and colostrum

Colostrum and breast milk were collected in 4 healthy donors using a sterile milk collection unit and an electric breast pump. All samples were transferred to a 15 mL conical centrifuge tube (Fisher Scientific) and centrifuged to remove cells (2000 x g, 10 min). The cream layer was not removed when aspirating the underlying fraction. Aliquots of the cell-free supernatant (including the skim milk fraction and a small fraction of the cream layer) were stored at -80 °C within 50 minutes after collection. Both colostrum samples were collected 3 days after birth. One of the mature breast milk samples was collected 37 days after giving birth, the other sample was collected 50 days after giving birth.

Pancreatic cyst fluid

Pancreatic cyst fluid was collected in two patients during an endoscopic ultrasound procedure. A volume of 1 mL cyst fluid was aspirated with a 10 mL syringe. Both patients underwent the procedure to investigate a cystic lesion in the pancreas. The first patient was diagnosed with a walled off necrosis collection after necrotizing pancreatitis, the second patient was diagnosed with a side-branch intra papillary mucinous neoplasia. The samples

were centrifuged to remove cells (2000 x g, 10 min) and stored at -80°C within 2 hours after collection.

Blood draw for plasma and serum

Venous blood was collected in 2 healthy donors from an elbow vein after disinfection with 2% chlorhexidine in 70% alcohol. All blood draws were performed with a butterfly needle of 21 gauge (BD Vacutainer, Push Button Blood Collection Set, Catalogue number 367326, Becton Dickinson and Company, NJ, USA). Tubes were filled to the volume recommended by the manufacturer. Plasma was prepared from whole blood collected in a 10 mL BD Vacutainer K2-EDTA tube (Catalogue number 367525, Becton Dickinson and Company, NJ, USA). Serum was collected in a 6 mL BD Vacutainer SST II Advance tube (Catalogue number 366444, Becton Dickinson and Company, NJ, USA). Immediately after the blood draw, plasma tubes were inverted five times and transported upright in an isolated transportation box at room temperature to the laboratory for further processing. The serum tubes were left on the bench upright for 30-60 minutes to allow clotting, followed by centrifugation at 1200 x g for 15 minutes. The supernatant was subsequently aliquoted per 200 µL into 2 mL LoBind tubes (Eppendorf LoBind microcentrifuge tubes, Z666556-250EA), snap frozen in liquid nitrogen and stored at -80 °C.

Before initiating the first centrifugation step, all plasma tubes were inverted five times. All centrifugation steps were performed in a centrifuge with swinging bucket rotor, without brake and at room temperature.

Plasma preparation procedure

Platelet-rich plasma (PRP) was obtained after centrifugation of blood tubes at 400 x g for 20 minutes. Using a 1 mL pipet, plasma was carefully transferred to a fresh 15 mL conical tube (CELLSTAR polypropylene tube, catalogue number 188271, Greiner Bio-One International,

Kremsmünster, Austria) leaving approximately 0.5 cm above the buffy coat. If PRP was needed for further experiments, PRP was aliquoted into LoBind tubes (Eppendorf LoBind microcentrifuge tubes, Z666556-250EA), snap frozen in liquid nitrogen and subsequently stored at -80 °C. The remaining volume of PRP was subsequently used for a second spin at 800 x g for 10 minutes. Platelet-poor plasma (PPP) was carefully transferred to a fresh 15 mL tube leaving approximately 0.5 cm above the platelet pellet. If necessary for further experiments, PPP was aliquoted into LoBind tubes (Eppendorf LoBind microcentrifuge tubes, Z666556-250EA), snap frozen in liquid nitrogen and subsequently stored at -80 °C. The remaining volume of PPP was used for a third spin at 2500 x g for 15 minutes. Platelet-free plasma (PFP) was carefully transferred to a fresh 15 mL tube leaving approximately 0.5 cm above the platelet pellet. All PFP was aliquoted in LoBind tubes (Eppendorf LoBind microcentrifuge tubes, Z666556-250EA), snap frozen in liquid nitrogen and stored at -80 °C. All plasma and serum samples were stored at -80 °C within 2 hours after blood collection.

Saliva (discovery cohort)

Spit samples from two healthy donors were collected in a 15 mL conical centrifuge tube. The donors refrained from eating, drinking, smoking and oral hygiene procedures for at least 1 h before sample collection. The samples centrifuged to remove cells (2000 x g, 10 min) and stored at -80 °C within 30 minutes after collection.

Saliva (case/control cohort)

Spit samples from 12 healthy donors and 12 type II diabetes patients were collected in a sterile 15 mL conical centrifuge tube. The donors refrained from eating, drinking other fluids than water, smoking and oral hygiene procedures for at least 1h before sample collection. Five minutes before the collection each donor rinsed his/her mouth twice with water without swallowing it. After a minimum volume of 10 mL saliva per donor was collected,

samples were centrifuged (3000 x g, 20 min, 4 °C) and the cell-free supernatant was removed carefully. Aliquots of 1.5 mL cell-free saliva were stored at -80 °C.

Seminal plasma

Semen samples were produced by masturbation into a sterile container and were allowed to liquefy for 30 min at 37 °C. Samples were centrifuged in a 15 mL conical centrifuge tube to remove cells (2000 x g, 10 min) and stored at -80 °C within 2 hours after collection.

Sputum (discovery cohort)

Spit samples from two donors with a common cold were collected in a 15 mL conical centrifuge tube. The donors refrained from eating, drinking, smoking and oral hygiene procedures for at least 1h before sample collection. The samples were centrifuged to remove contaminating cells (2000 x g, 10 min) and stored at -80 °C within 30 minutes after collection.

Sputum (case/control cohort)

Sputum induction was performed on 16 subjects who were recruited from the outpatient pulmonary clinic of the Ghent University Hospital or by advertising. Subjects inhaled sterile, pyrogen-free, hypertonic, nebulised saline at increasing concentrations of NaCl (3%, 4% and 5%) over a 5-minute period after inhalation of salbutamol (2 x 200 µg). Subsequently, subjects were encouraged to cough and expectorate an adequate sample. Sputum plugs were selected, transferred in a polystyrene tube and mixed with dithiotreitol (DTT; 10% Sputalysin with an amount of four times the weight of the sputum plugs, Boehringer-Calbiochem Corp, San Diego, CA, USA) for 30 seconds by vortex and for 15 minutes by tube rocker. Next, an amount of PBS equal to the volume of DTT was added. The sample was incubated for 5 minutes, filtered and centrifuged for separation between cell fraction and cell-free supernatant fraction. The supernatant was aspirated, aliquoted, and stored at -80 °C until further analysis.

Stool

Stool was collected from two healthy donors in a sterile recipient. Two different preparation protocols were followed.

Protocol without extraction buffer (referred to as “stool” in the manuscript)

Immediately after collection 0.5 mL stool was transferred from the sterile recipient into a 15 mL conical centrifuge tube. To homogenize the samples 1 mL RNase free Sigma water was added and the samples were vortexed for 30 seconds. The samples were centrifuged to remove cells (2000 x g, 10 min) and stored at -80 °C within 30 minutes after collection.

Protocol with extraction buffer (referred to as “Calex stool” in the manuscript)

Immediately after collection stool from the sterile recipient was added to the Calex Cap extraction device (Calex, Bühlmann Laboratories AG, Schönenbuch, Switzerland) according to manufacturer’s instructions, resulting in a diluted stool sample (1:500). The diluted sample was subsequently transferred to a 15 mL conical centrifuge tube. The samples were centrifuged to remove cells (2000 x g, 10 min) and stored at -80 °C within 30 minutes after collection.

Sweat

Sweat was collected in two healthy volunteers during exercise on bicycle rollers. Prior to the collection the forehead of the volunteers was disinfected with 2% chlorhexidine in 70% alcohol. Sweat dripping from the forehead was collected in a 15 mL conical centrifuge tube. The samples were centrifuged to remove cells (2000 x g, 10 min) and stored at -80 °C within 30 minutes after collection.

Synovial fluid

Aspiration of synovial fluid from the swollen knee joint was performed in two patients through a sterile puncture with a 21-gauge needle on a 20 mL syringe. The synovial fluid was

subsequently transferred to a 15 mL conical centrifuge tube (Fisher Scientific). Both samples were centrifuged to remove cells (2000 x g, 10 min) and stored at -80 °C within 50 minutes after collection. One of the patients was previously diagnosed with spondyloarthropathy, the second patient was diagnosed with rheumatoid arthritis. In both patients, septic arthritis was excluded.

Tear collection

Tears were collected in 8 healthy, non-contact lens wearing volunteers, using Schirmer strips. Collection was performed under sterile conditions. The closed eyes were first cleaned with a sterile swab. Subsequently 2 Schirmer strips (TearFlo Sterile Strips, HUB Pharmaceuticals, LLC) per eye were folded into the inferior palpebral fold of the right and the left eye simultaneously. Tears were absorbed into the Schirmer strips while the eyes were closed for maximum 5 minutes. The length of the moistened area was measured using the millimeter scale on the strip. Per donor the total length of the moistened area for ranges from 55 mm to 140 mm (median 127.5 mm). Four Schirmer strips of one donor were collected in one 2 mL Eppendorf tube. RNA isolation was performed within 2 hours after collection.

Urine (discovery cohort)

The mid-stream portion of second morning urine was collected in a sterile recipient. Urine was obtained from one male and one female healthy donor. The samples were centrifuged to remove cells (2000 x g, 10 min) and stored at -80 °C within 30 minutes after collection.

Urine (case/control cohort)

Samples of first morning voided urine were collected in 15 ml tubes with EDTA used for nucleic acid preservation. Samples were centrifuged at 4 °C to remove cells (2000 x g, 15 min). Supernatant was collected and stored at -80 °C until analyzed. Urine was collected in

12 (6 male and 6 female) patients with muscle invasive bladder cancer and in 12 healthy volunteers (6 male and 6 female patients).