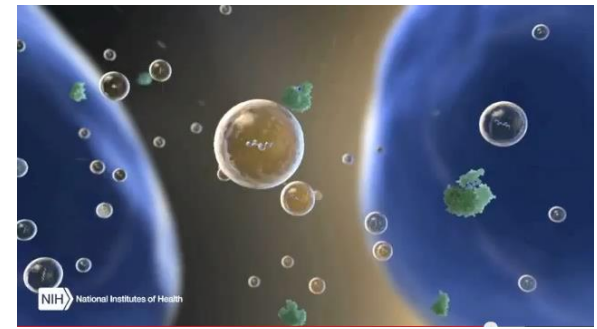


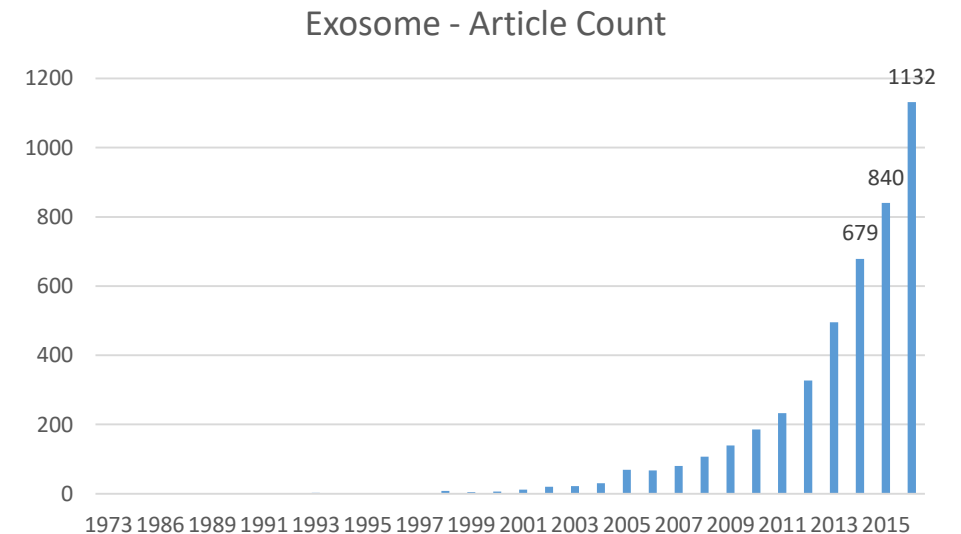
exRNA

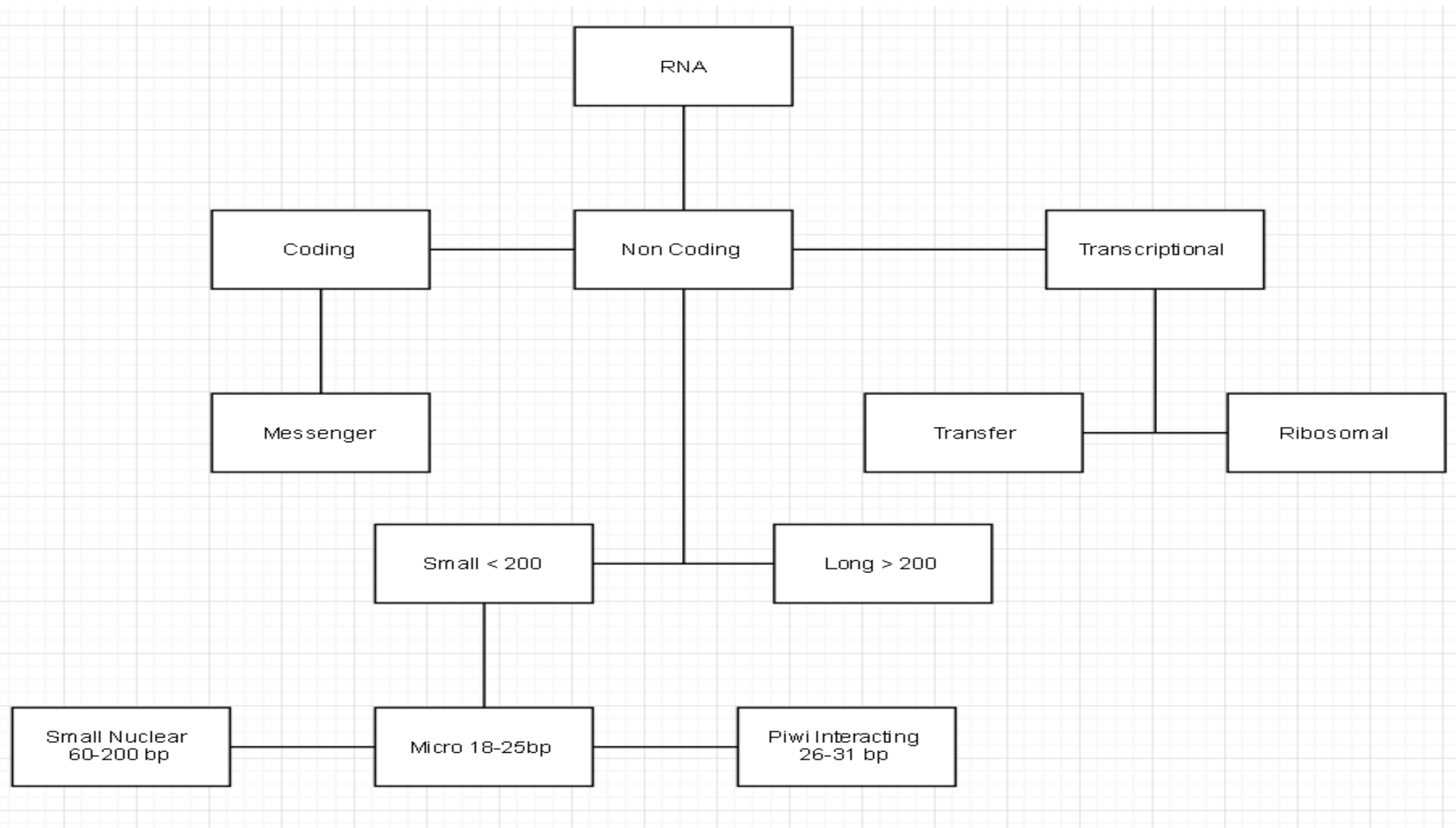




# Intro

- Biogenesis, Cargo and Uptake of Extracellular Vesicles
  - Protein content
  - **RNA** Content
  - Lipid Content
- Collection and Processing of Cell Culture Media and Body Fluids Prior to Isolation of Extracellular Vesicles
  - Cell Culture Media
  - Blood **Plasma** and **Serum**
  - Breast Milk
  - Urine
  - **CSF**
- Techniques for Characterization and Quantification of Extracellular Vesicles
  - EM
  - Flow Cytometry
  - Nanoparticle Tracking Analysis
  - **Sequencing**





# exRNA Atlas

- **1,369** exRNA profiles

exRNA Atlas: Data, Tools & Computable Knowledge

[☰ Datasets](#)
[📊 Analysis Tools](#)
[🆘 Help](#)
[More >](#)
[📄 Cite Us](#)
[✉ Contact Us](#)

Bringing exRNA Data and Analysis Tools Together

The exRNA Atlas is the data repository of the Extracellular RNA Communication Consortium (ERCC), which includes small RNA sequencing and RT-qPCR-derived exRNA profiles from human and mouse biofluids. All RNA-seq datasets are processed using *version 4* of the *excerpt small RNA-seq pipeline* and ERCC-developed quality metrics are uniformly applied to these datasets.

[Getting Started](#)

Select exRNA Profiles: (0 selected)

Select, view and download Atlas data by clicking one or more slices from one or more charts. Then, click the icon in the floating menubar to apply filters and view the results (grid opens in a new tab). Click the icon to select all slices from all charts (i.e. all exRNA profiles in the Atlas) or click the icon to clear selections. Please note that the size of each slice (representing a profile count) has been log-transformed.

**Condition**

- Healthy Control (226)
- Liver Disease (8)
- Gestationally Small Fetus (4)
- Glioblastoma (4)
- Preeclampsia (15)
- Cardiovascular Disorder (19)
- Parkinson's (107)
- Ulcerative Colitis (3)
- Lupus (3)
- Nephrotic Syndrome (78)
- Hemorrhage (523)
- Carcinoma (163)
- Alzheimer's (213)

**Biofluid**

- Bile (5)
- Urine (79)
- Cerebrospinal Fluid (537)
- Plasma (511)
- Conditioned Media (22)
- Serum (215)

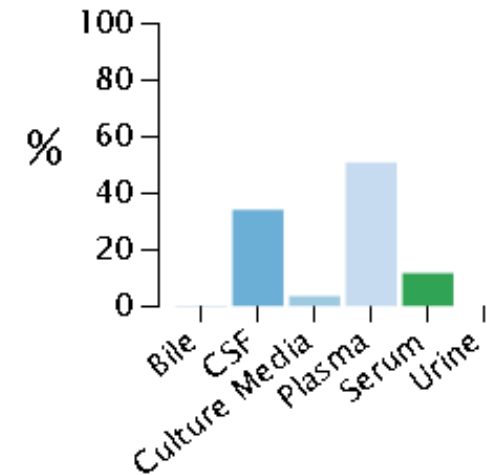
**RNA Source**

- Extracellular exosome (22)
- Extracellular vesicle (8)
- HDL complex (8)
- Total cell-free biofluid RNA (1333)

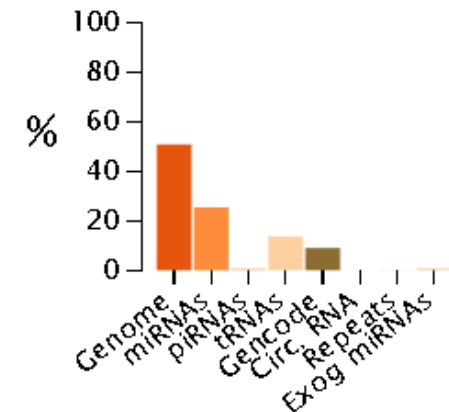
**RNA Isolation Kit**

- miRvana Paris (967)
- SeraMir (4)
- miRNeasy (274)
- Other (79)
- miRcuy Biofluids (41)
- miRcuy Cell & Plant (4)

Transcriptome Mapped Reads / Biofluid

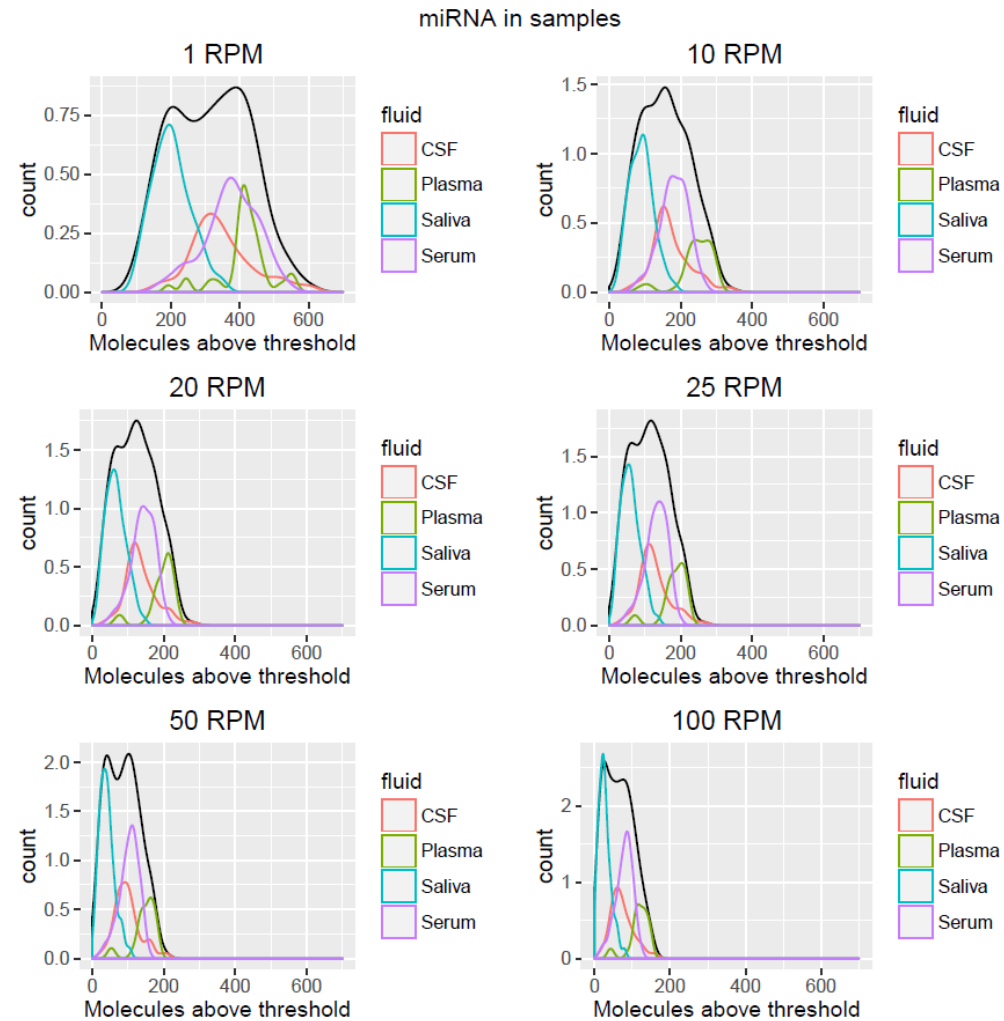


Read Mappings / RNA Type

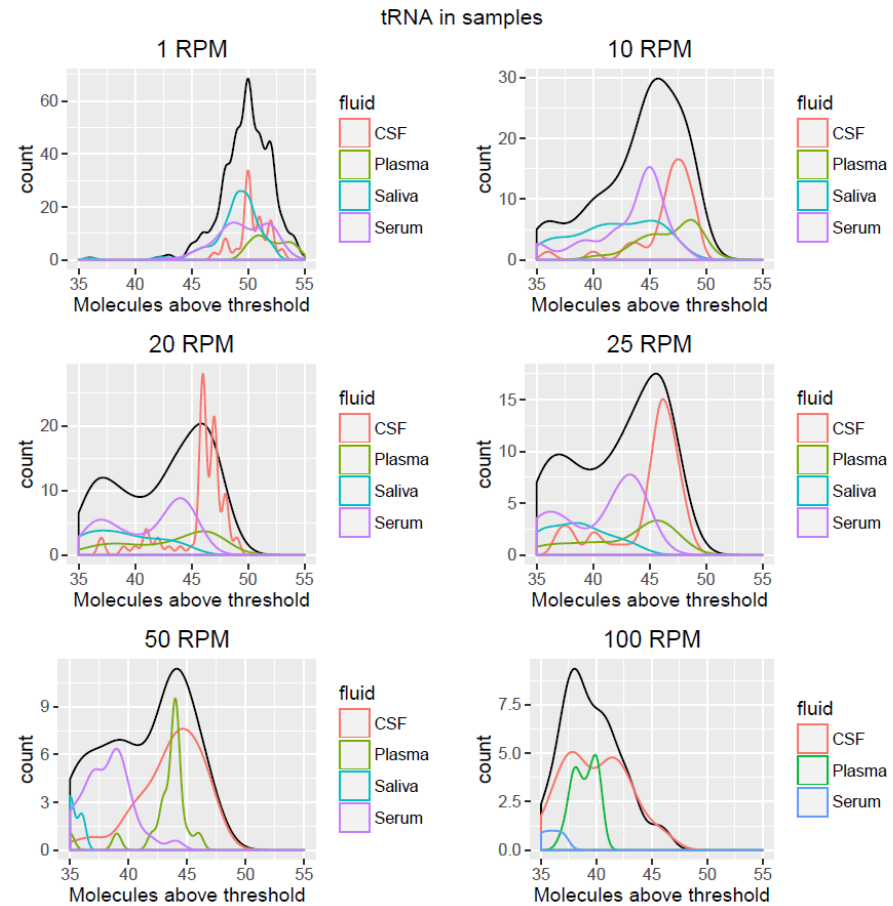


How many small RNA are expressed in different biofluids

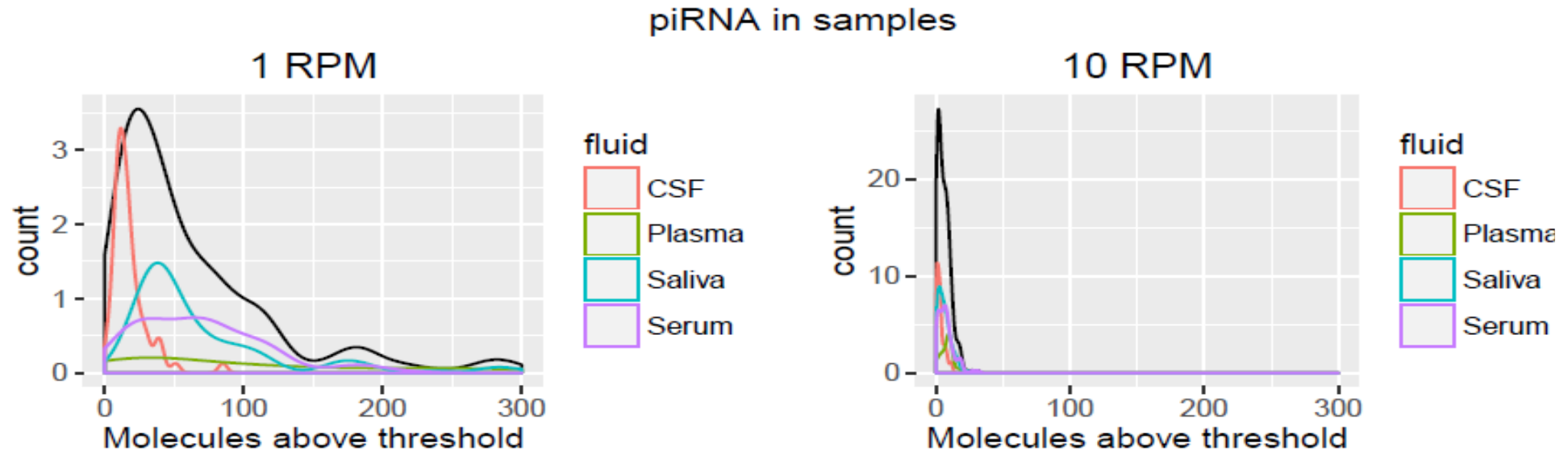
# No. of expressed miRNA split by biofluid and threshold



# No. of expressed small tRNA split by biofluid and threshold



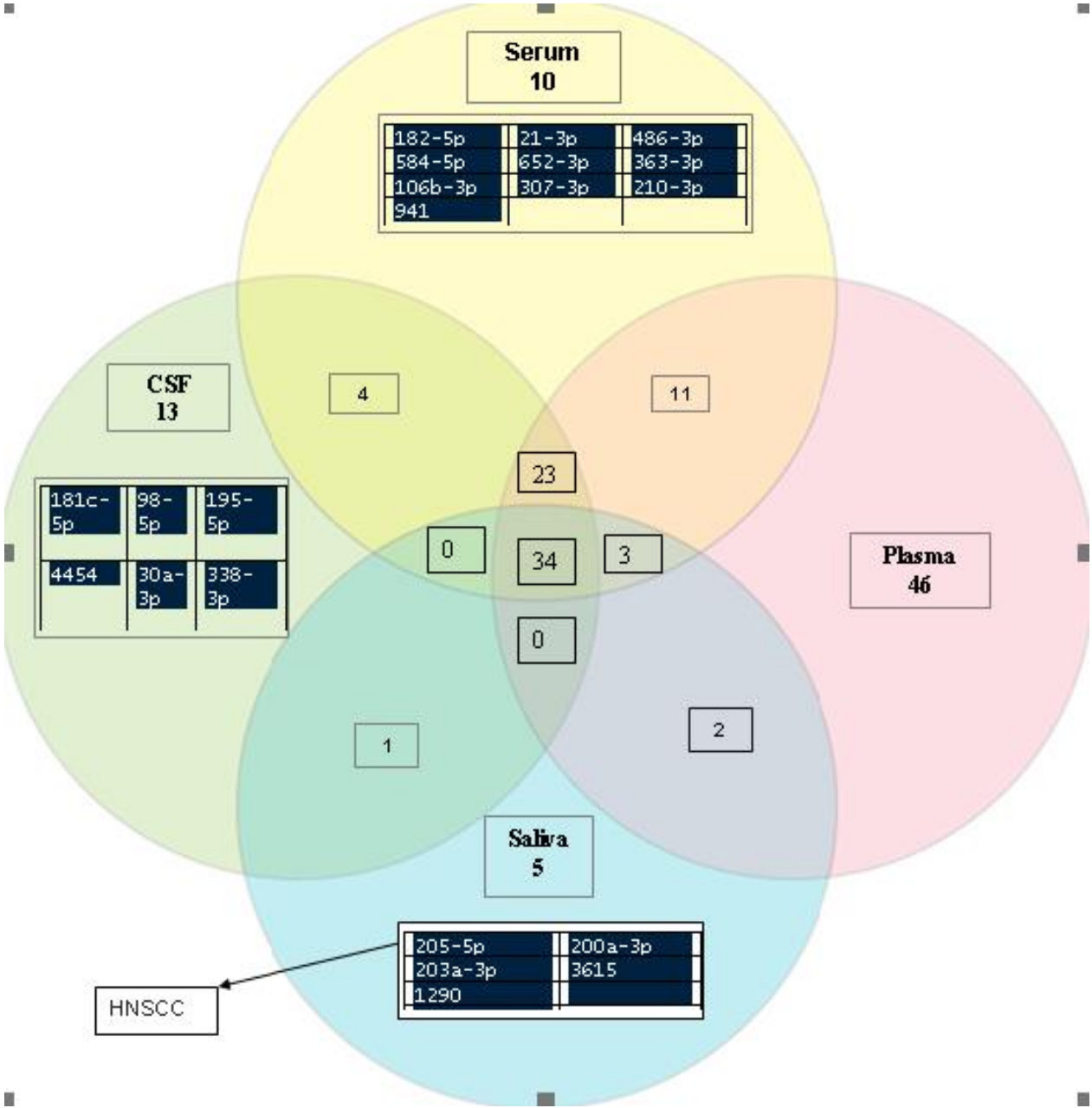
# No. of expressed small piRNA split by biofluid and threshold



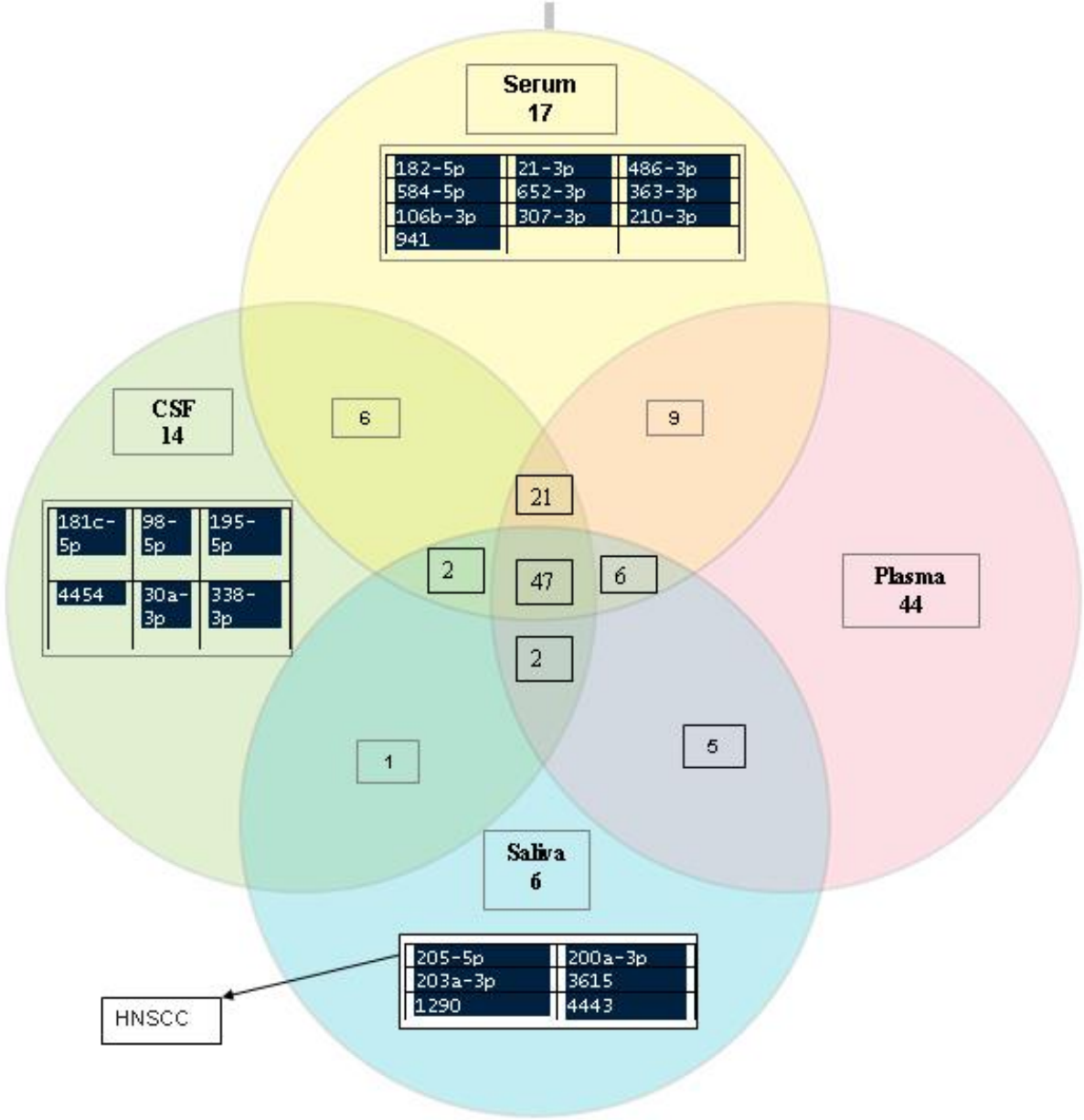


What are those MIRs?  
Are they shared between biofluids?

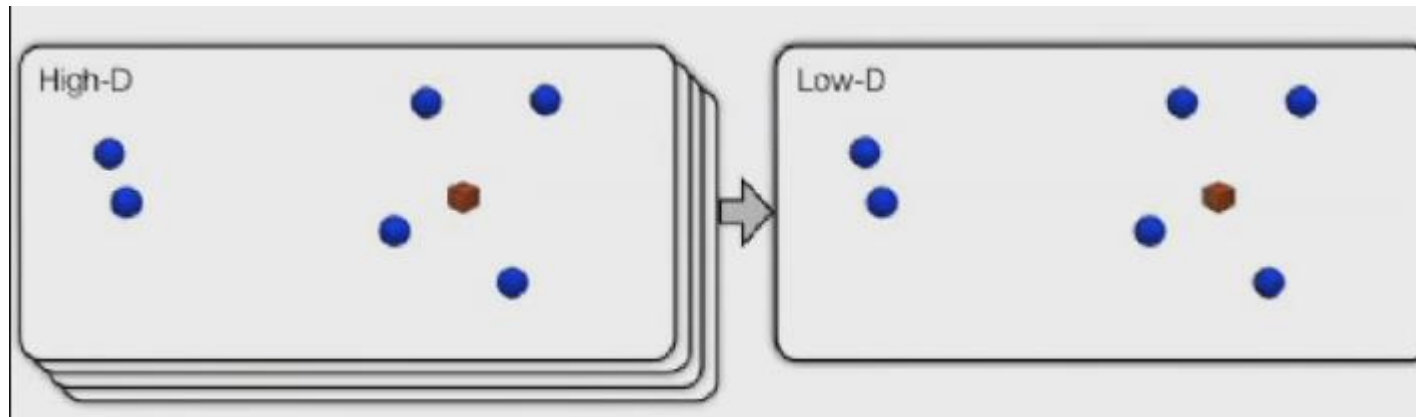
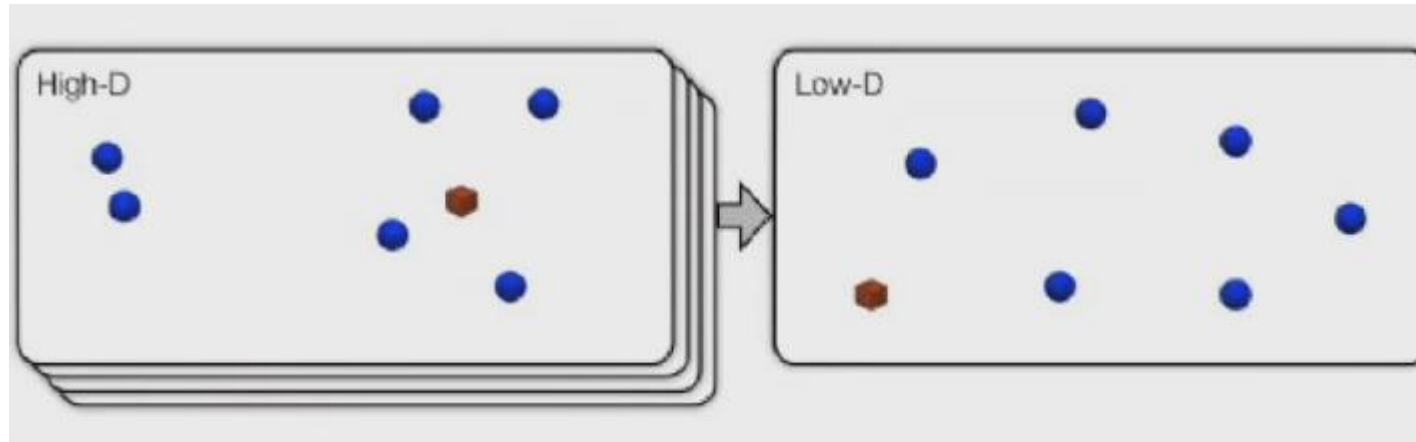
# Threshold 25



# Threshold 100

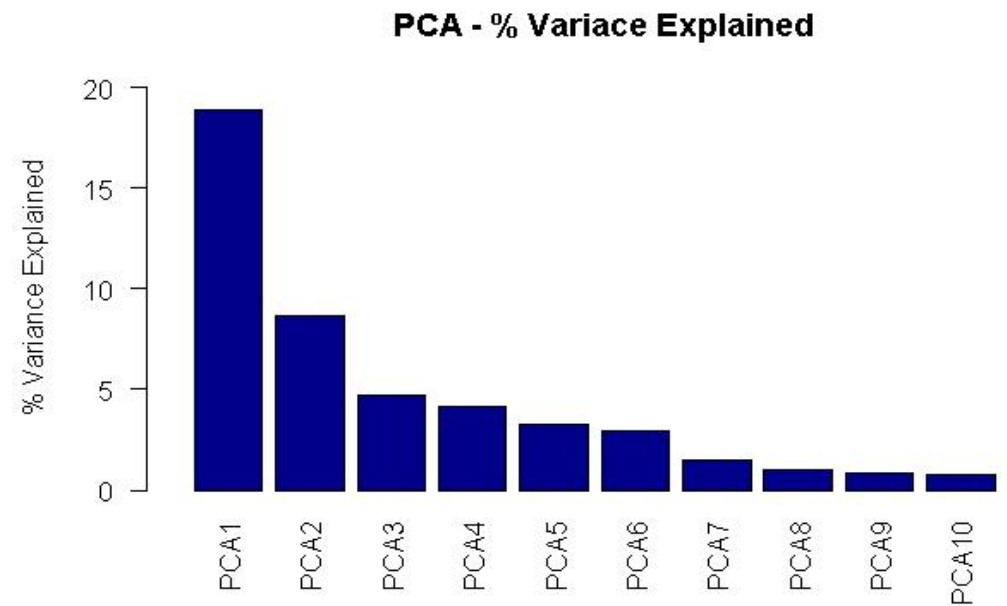
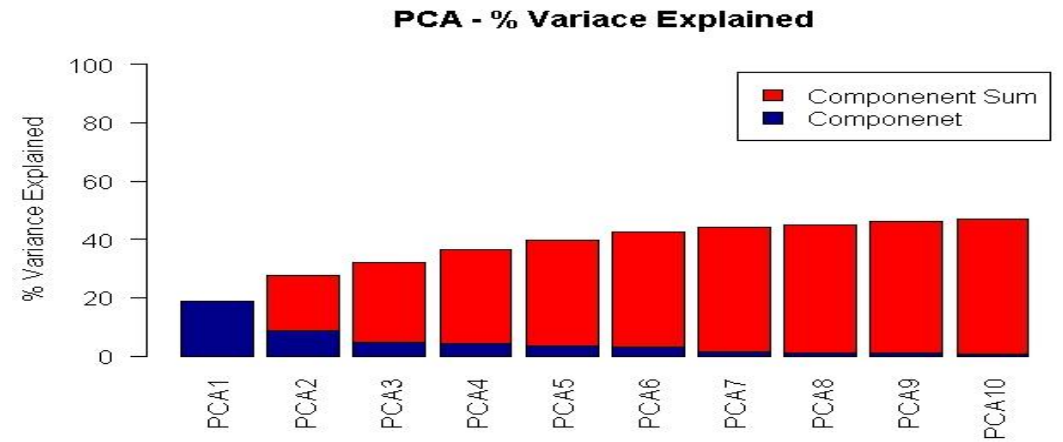


# Visualization - Dimensionality reduction

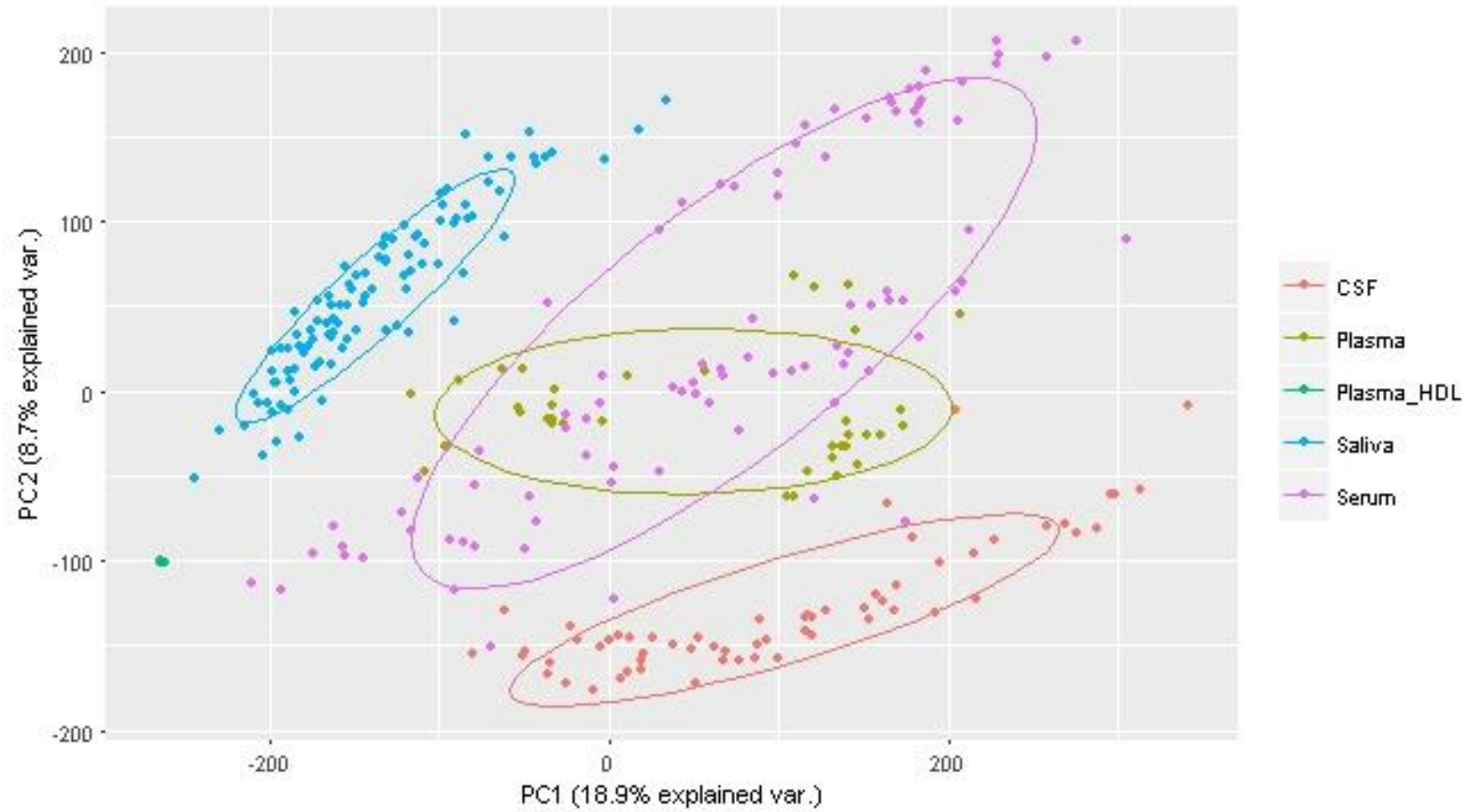


# PCA

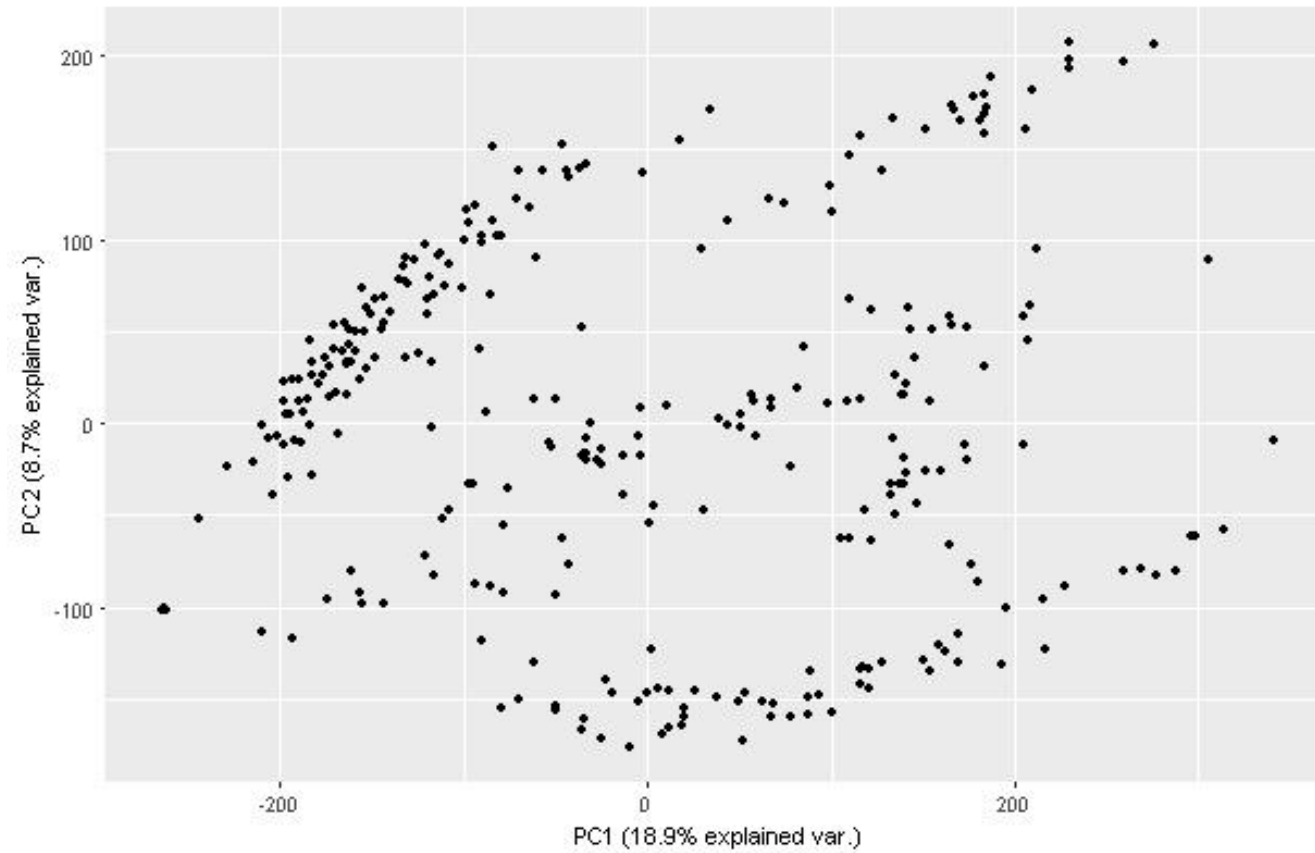
- Search for low dimensional subspace that minimize a **squared error** between distances in the original data and distances in the map
- Maximize the variance of the data
- Linear projection

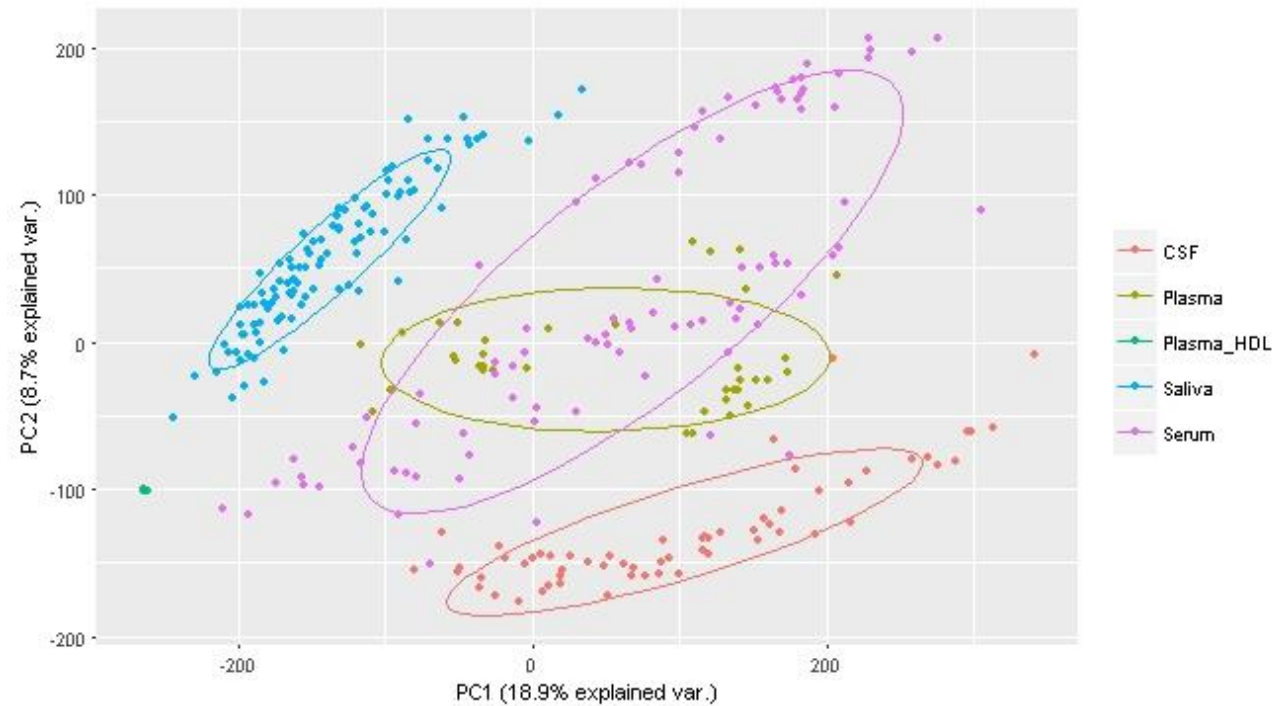
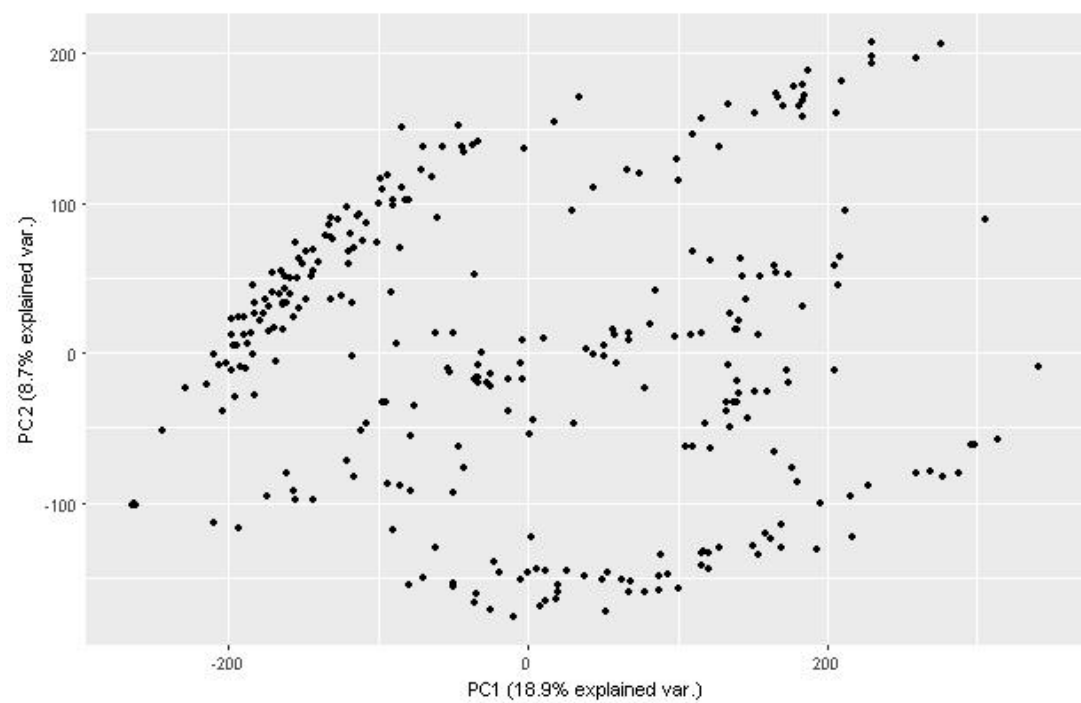


# PC2 vs. PC1



# PC2 vs. PC1



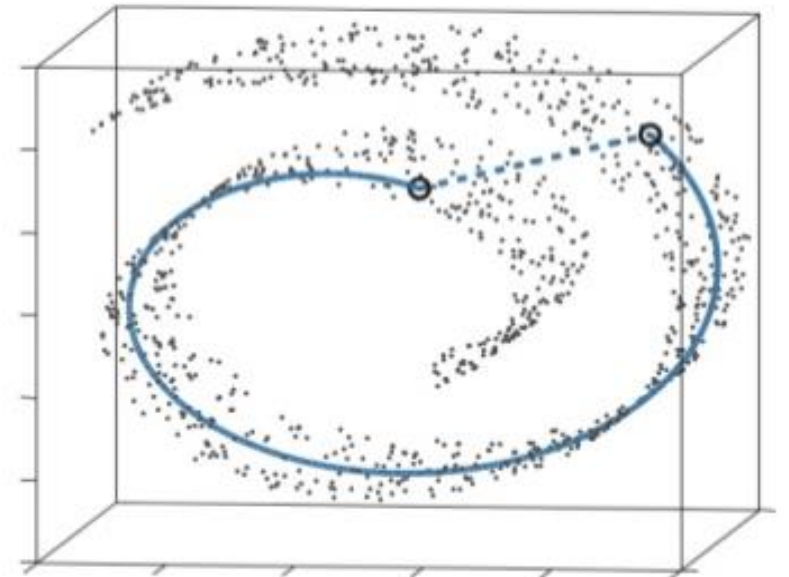




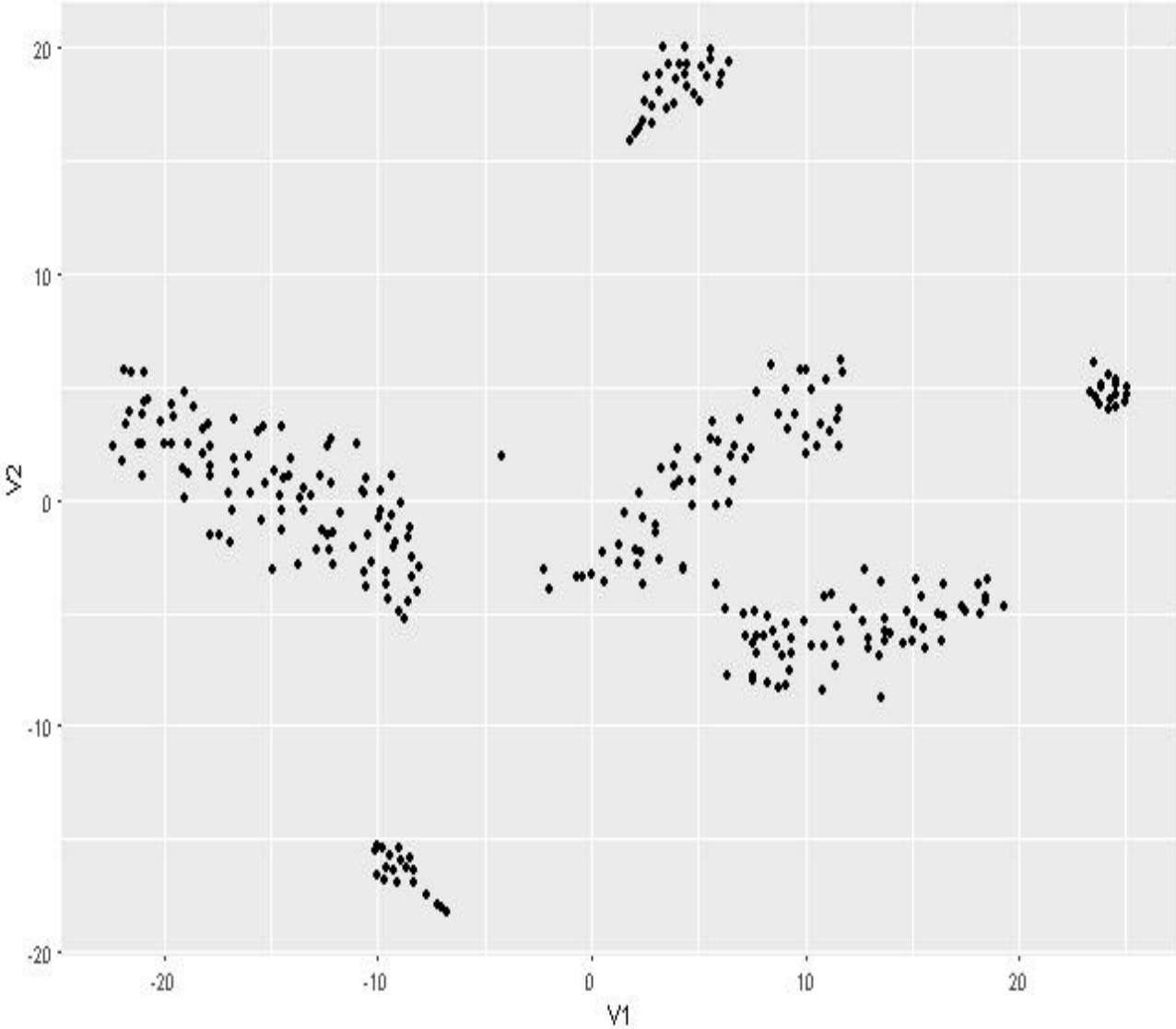
# Can we do better?

## Is PCA minimizing the right objective function?

- Linear projection
- Trying to minimize a squared error between distances in the original data and distances in the map -> Trying to **maximize variance**
- PCA is mainly concerned in preserving **large pairwise** distances in the map

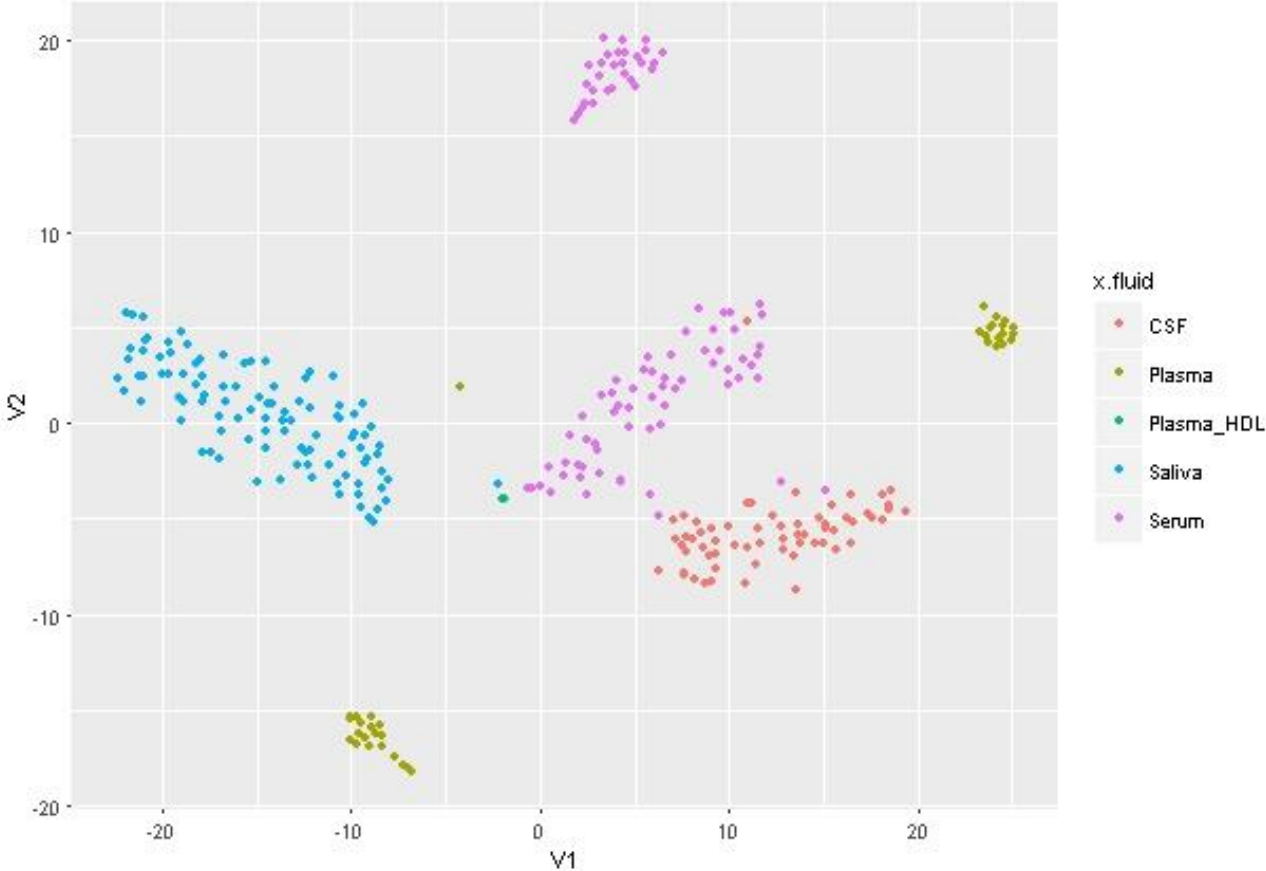


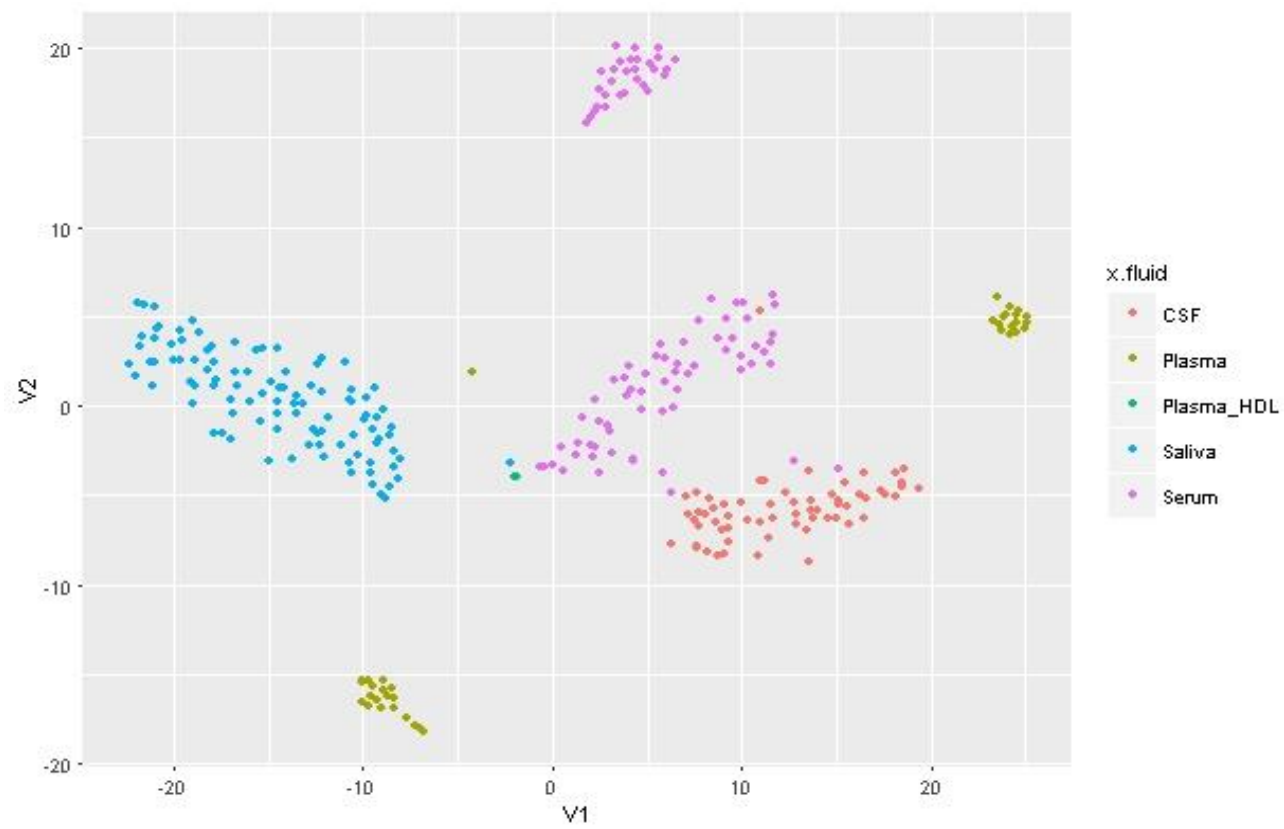
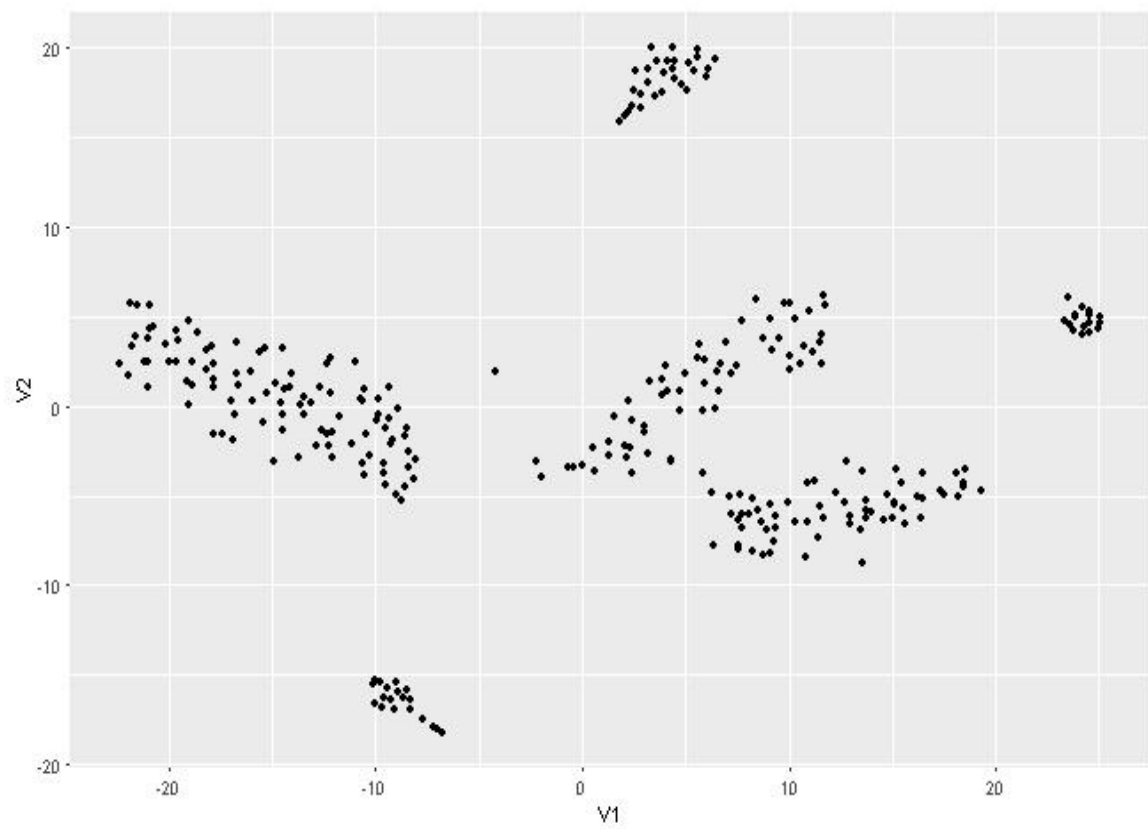
TSNE output



# TSNE output

With biofluids labels

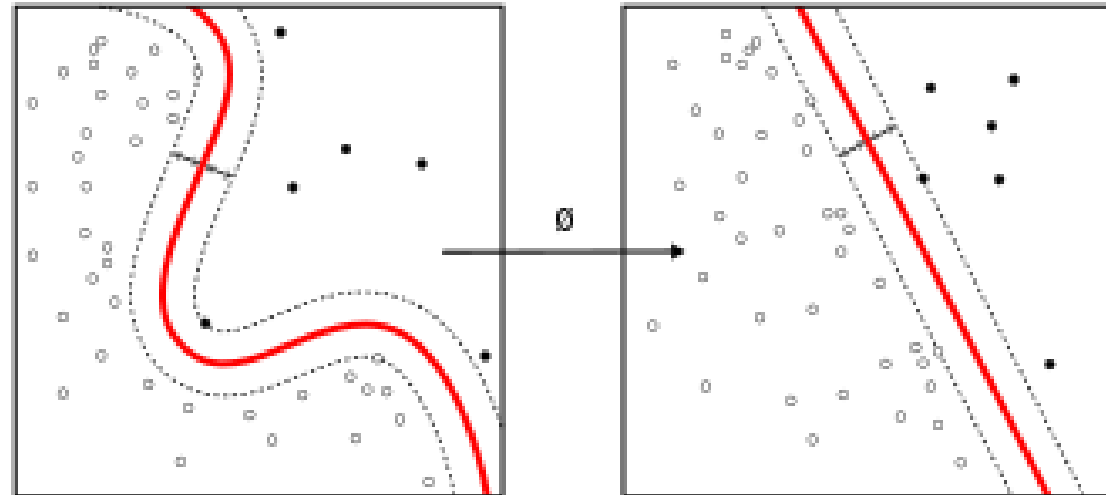




# T-SNE

## T-distributed Stochastic Neighbor Embedding

- Developed by Geoffrey Hinton & Laurens van der Maaten
- **Nonlinear** dimensionality reduction technique
- 2 Stages:
  - Constructs a [probability distribution](#) over pairs of high-dimensional objects
  - Defines a similar probability distribution over the points in the low-dimensional map



# t-SNE – step 1

- What is reliable? Very small Euclidean distances (point & nn)
- Measure similarities between points that looks only on local similarities (to nearby points)

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)},$$

# t-SNE – step 2

- Look at the low dimensional space
- Lay points in that map
- Represent each high dimensional object by a point on the map
- A probability that measures similarity of 2 points in the low dimensional map
- We want  $Q_{ij}$  to reflect the similarities in  $P_{ij}$  as well as possible

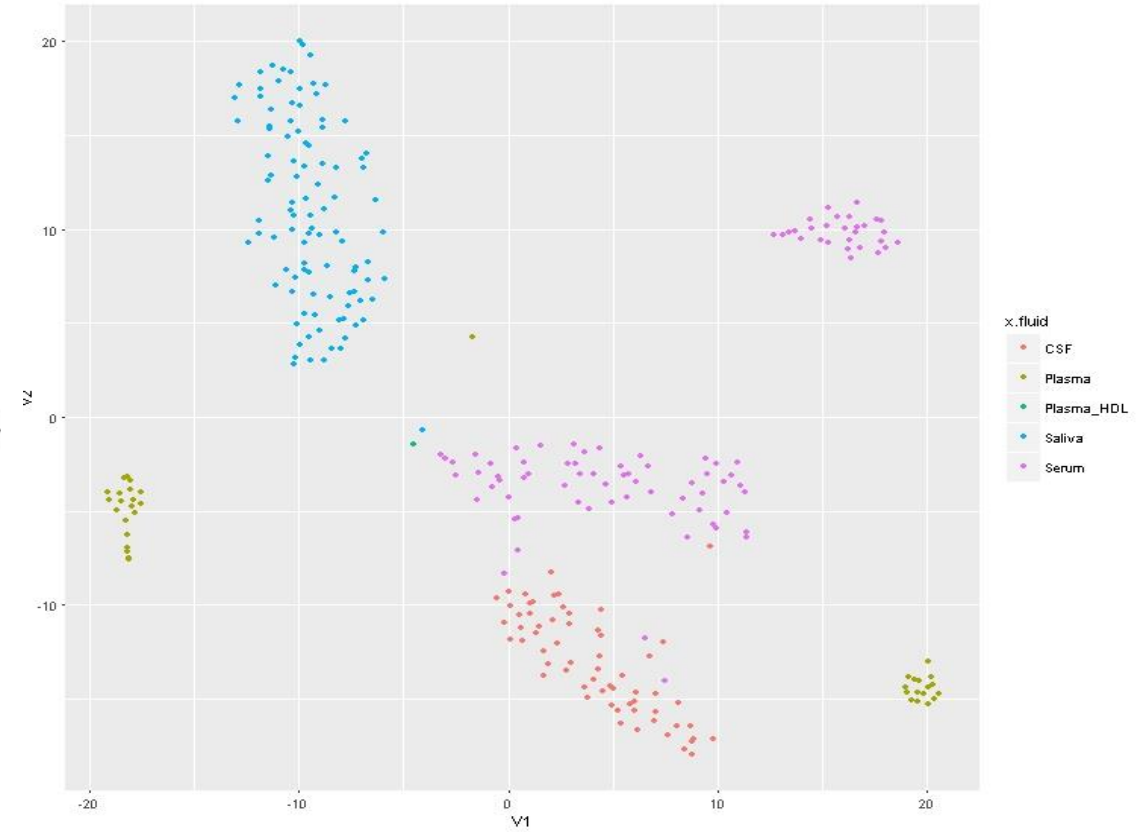
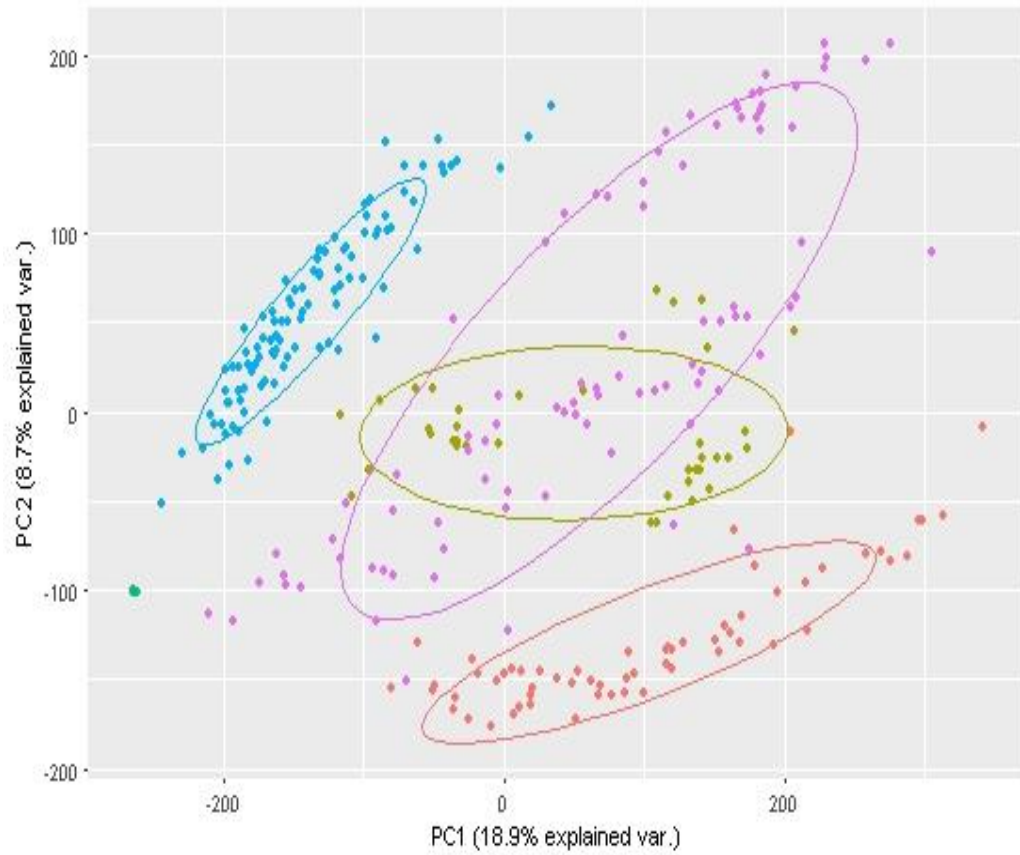
$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|\mathbf{y}_k - \mathbf{y}_i\|^2)^{-1}}$$

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

# Comparison

PCA

TSNE





# Conclusion & Potential directions

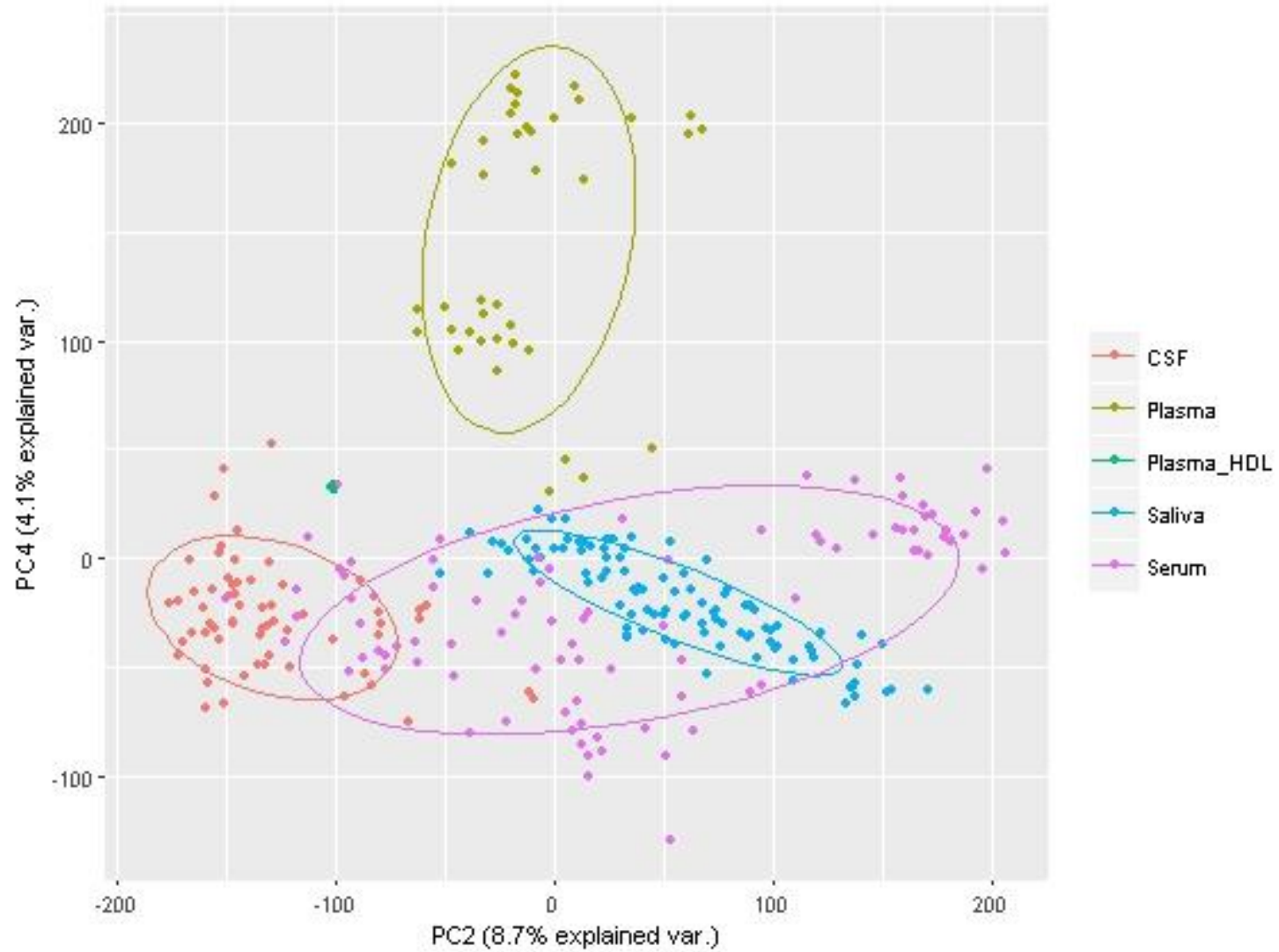
- Difference between cell content & EV cargo – comparison between profile of EV and originating Cell
- Dimensionality reduction for other small RNA
- Prediction – use to classify new samples



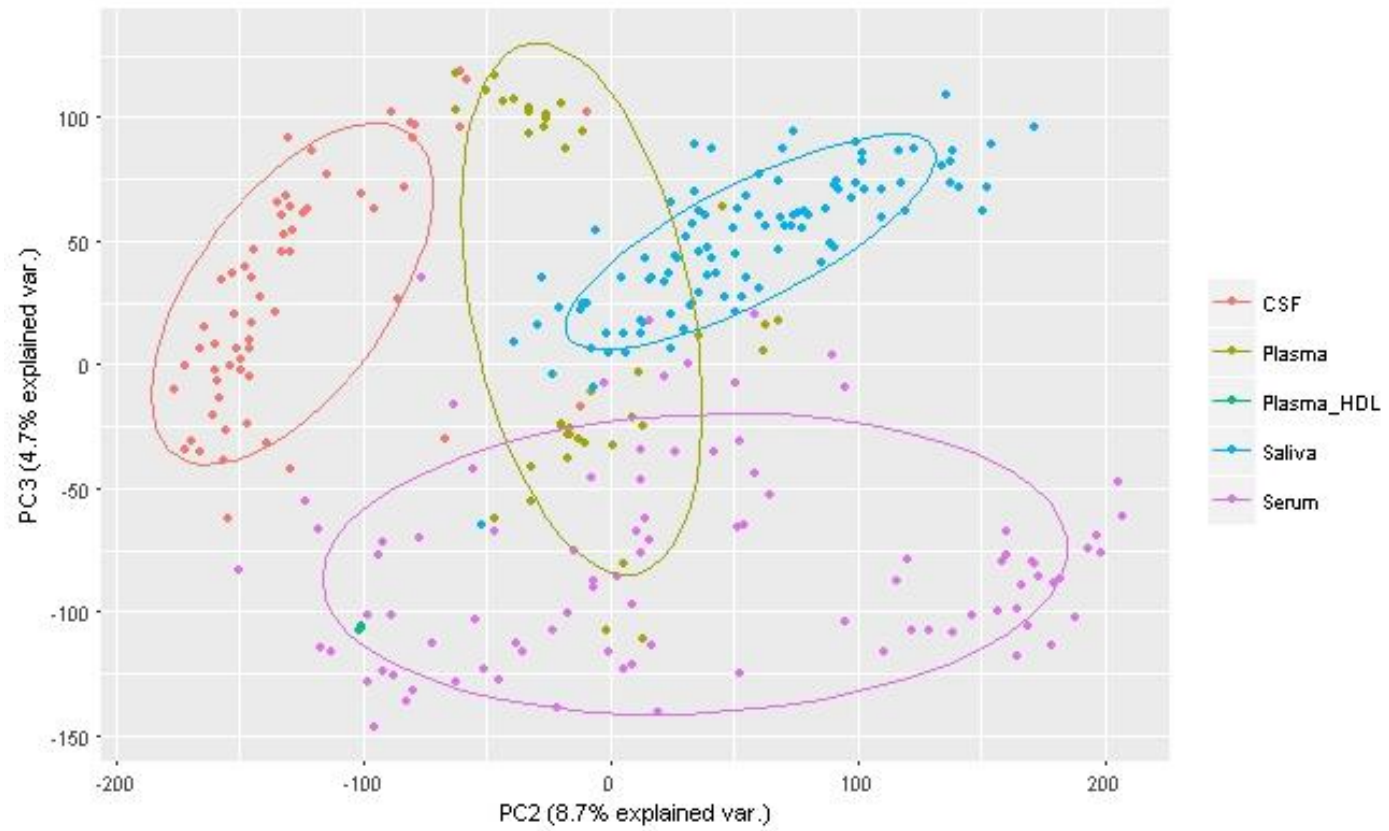
# Thank you

- Joel Rozowsky
- Mark Gerstein

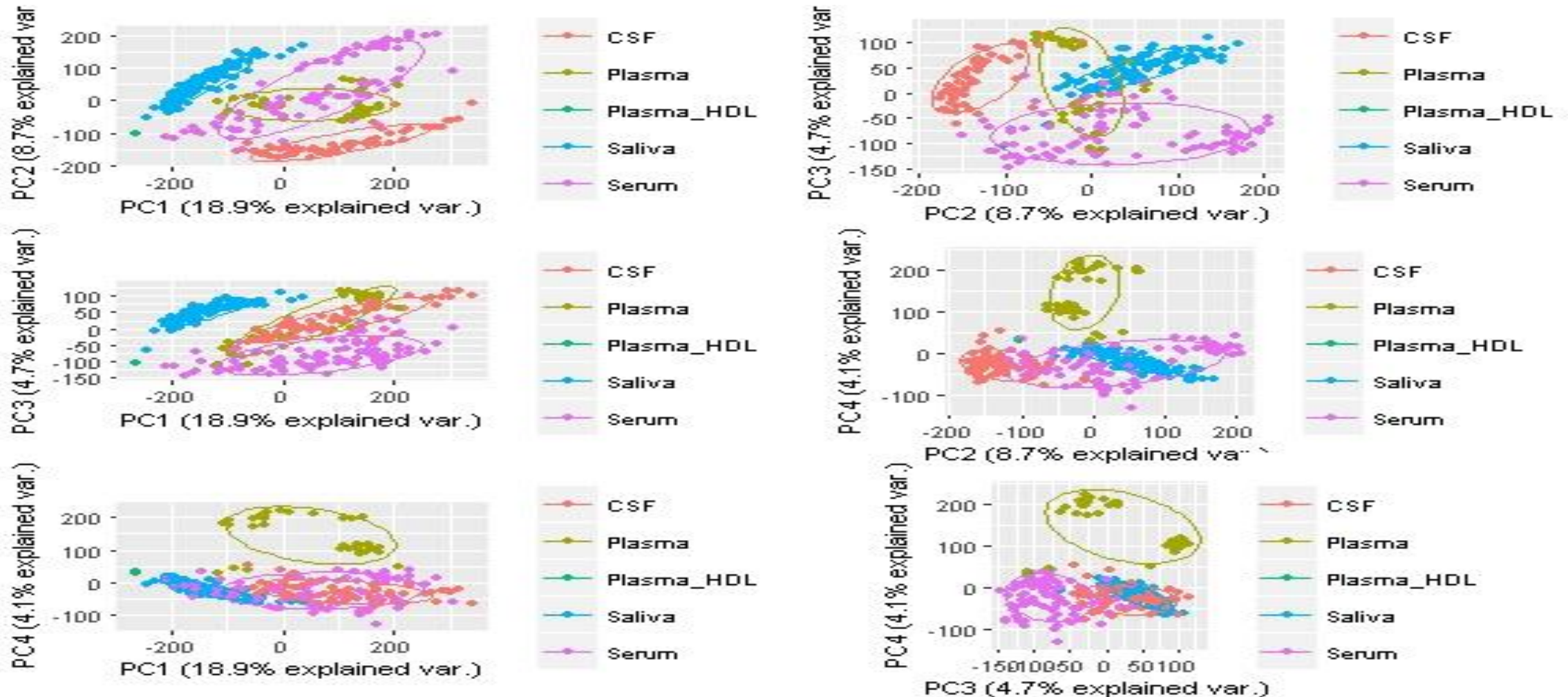
# PC4 Vs. PC2



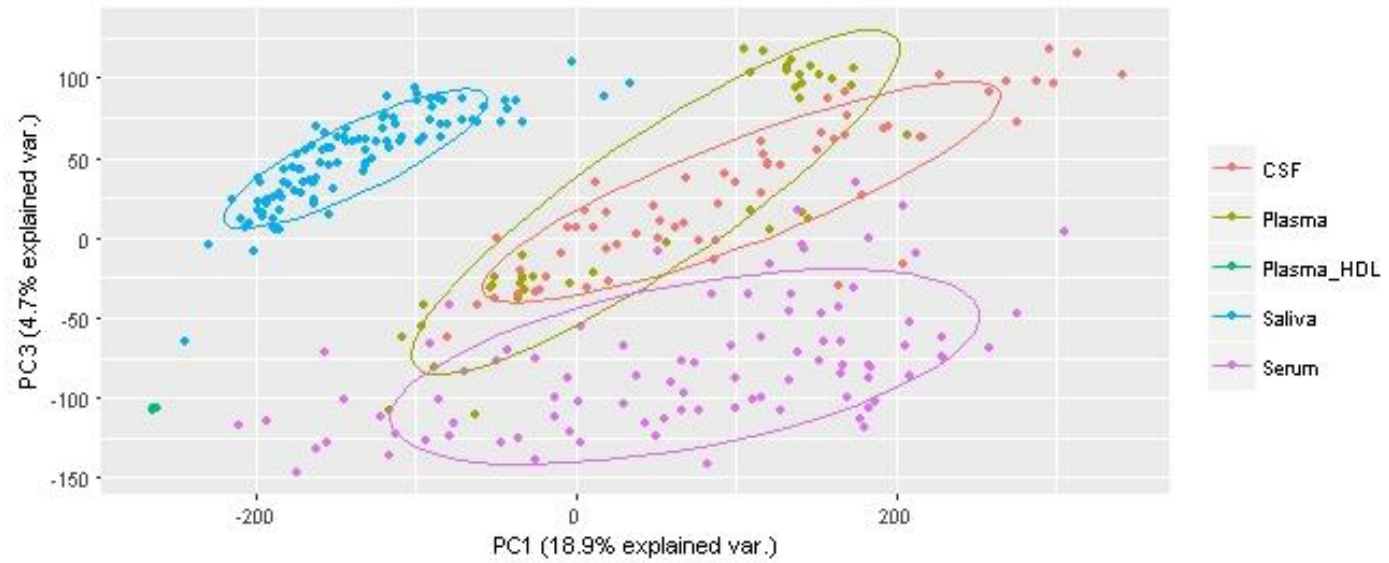
# PC3 Vs. PC2



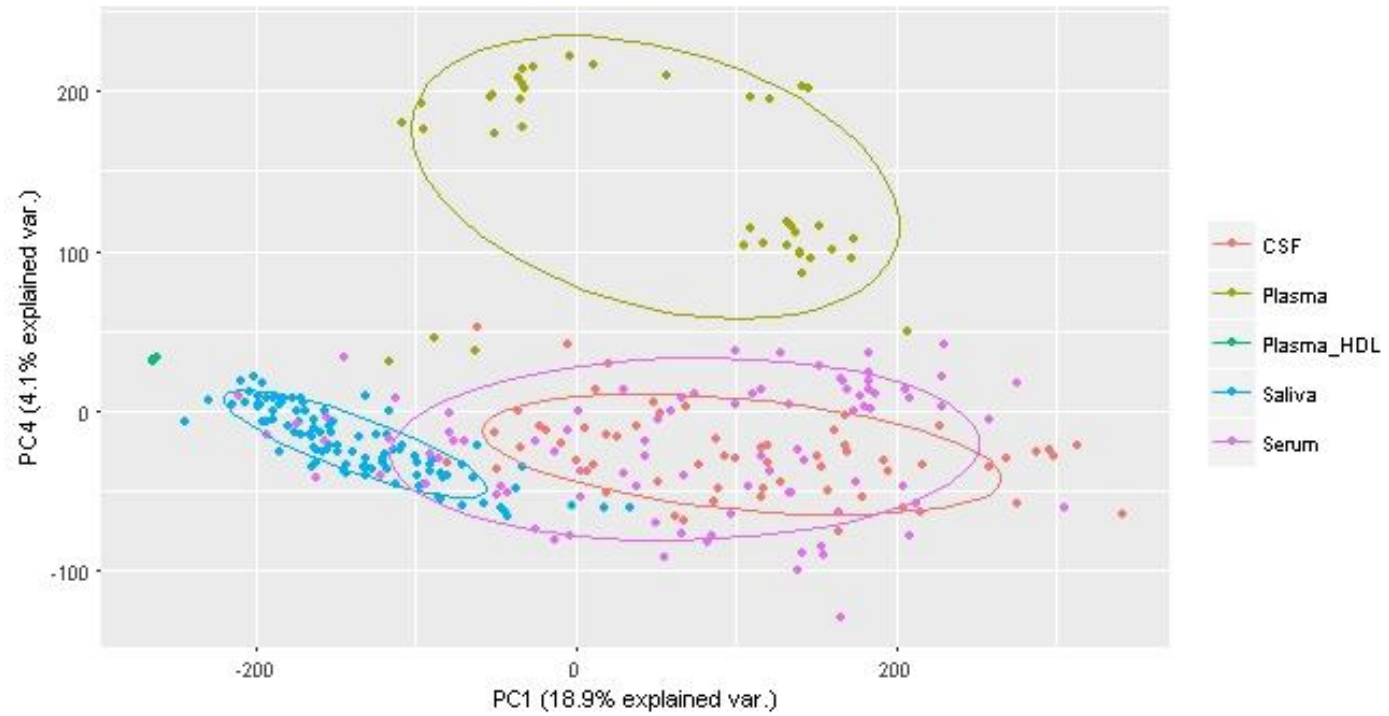
# PCA – Center = True; Scale = False



# PC3 Vs. PC1



# PC4 Vs. PC1





# PC3 Vs. PC4

