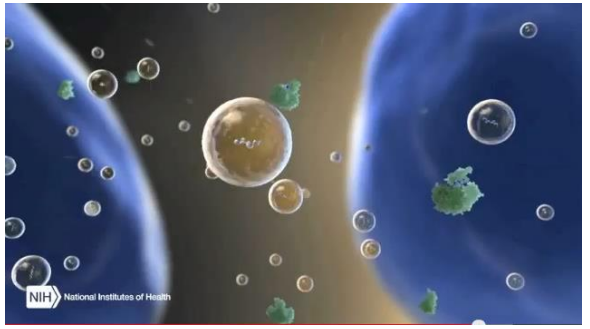
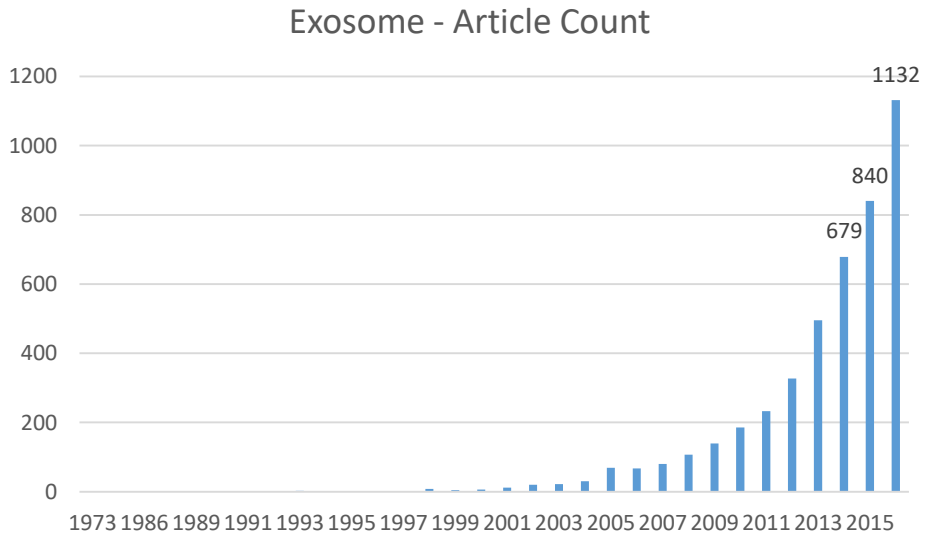
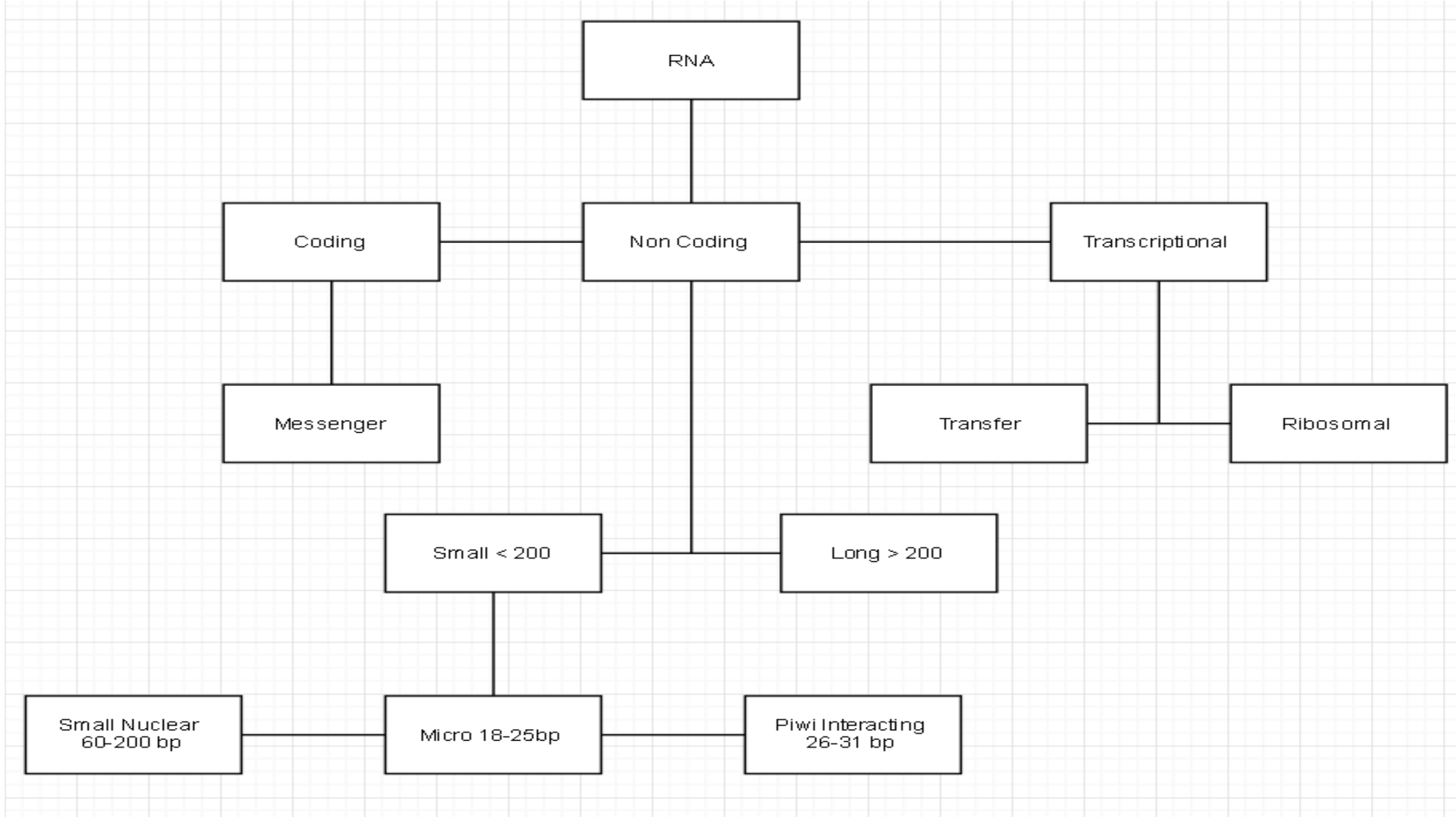
exRNA

Intro



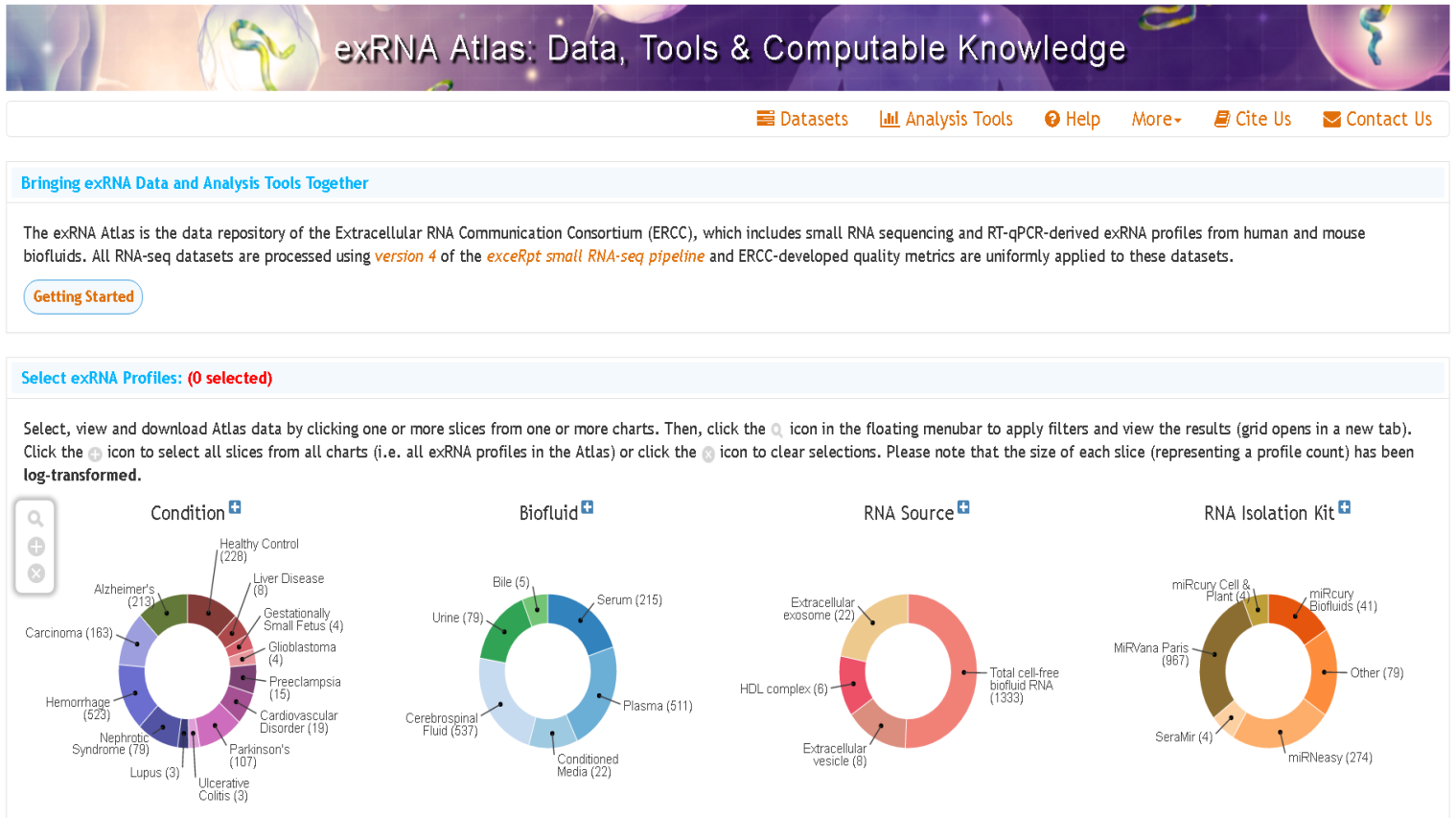
- Biogenesis, Cargo and Uptake of Extracellular Vesicles
  - Protein content
  - **RNA** Content
  - Lipid Content
- Collection and Processing of Cell Culture Media and Body Fluids Prior to Isolation of Extracellular Vesicles
  - Cell Culture Media
  - Blood **Plasma** and **Serum**
  - Breast Milk
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

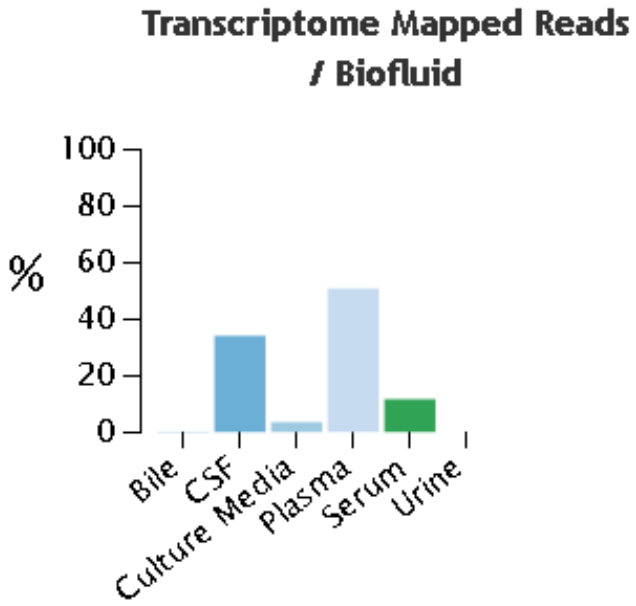


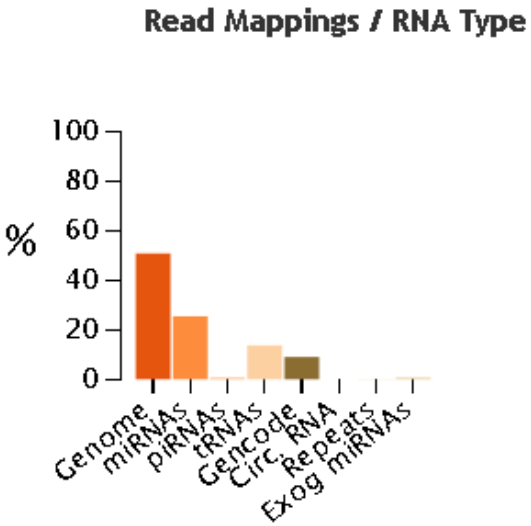

RNA
Coding
Non Coding
Transcriptional
Messenger
Transfer
Ribosomal
Small < 200
Long > 200
Small Nuclear
60-200 bp
Micro 18-25bp
Piwi Interacting
26-31 bp

exRNA Atlas

- **1,369** exRNA profiles



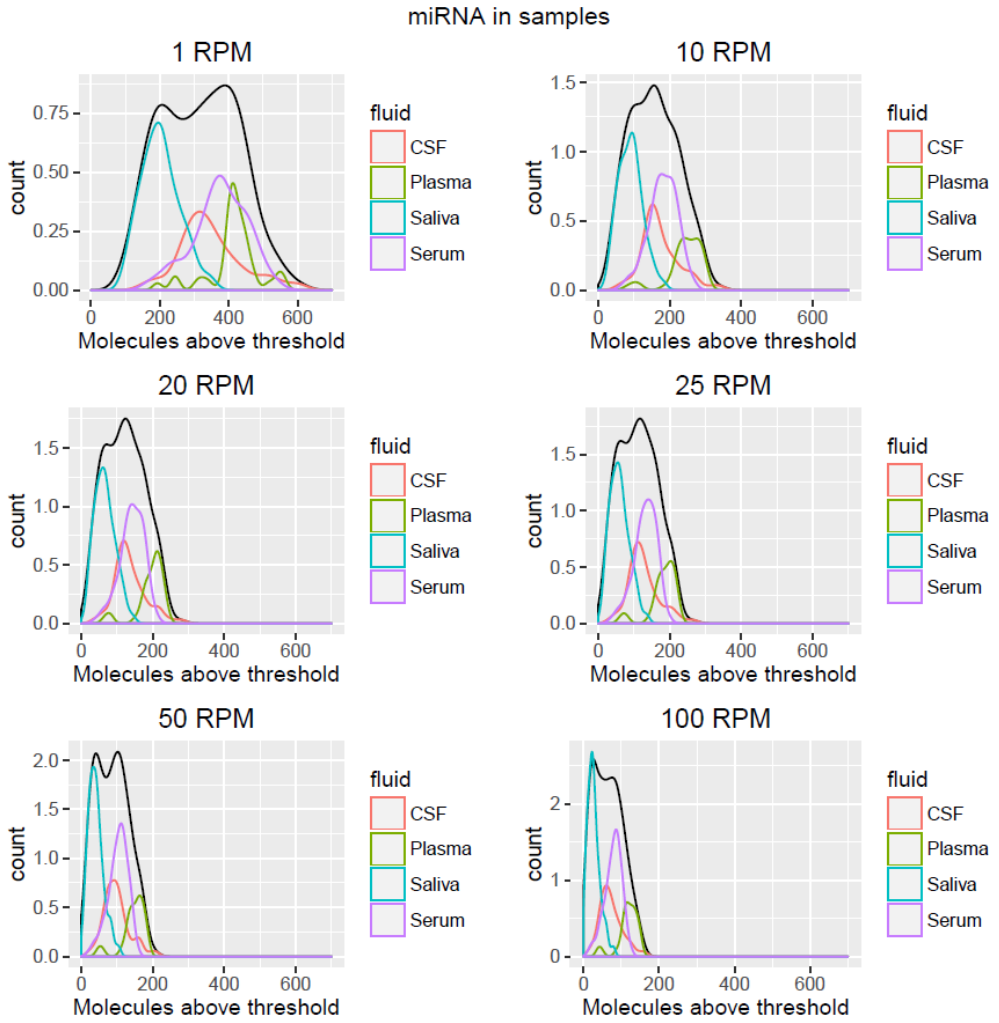




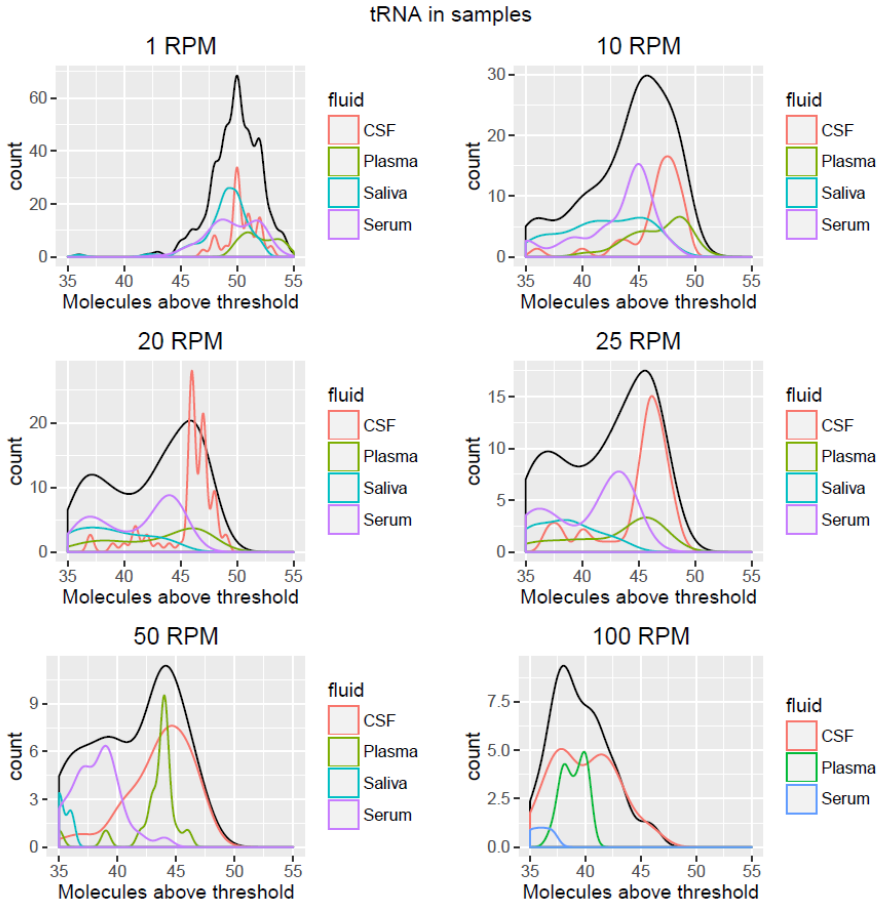
How many small RNA are expressed in different biofluids

No. of expressed miRNA

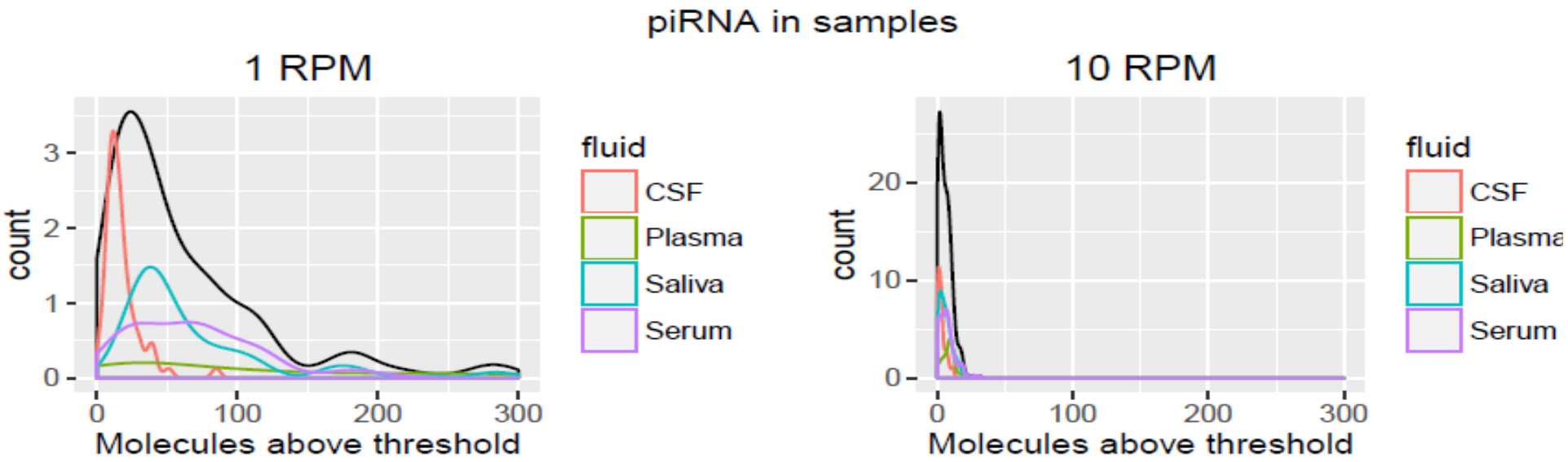
split by biofluid and threshold



No. of expressed small tRNA split by biofluid and threshold



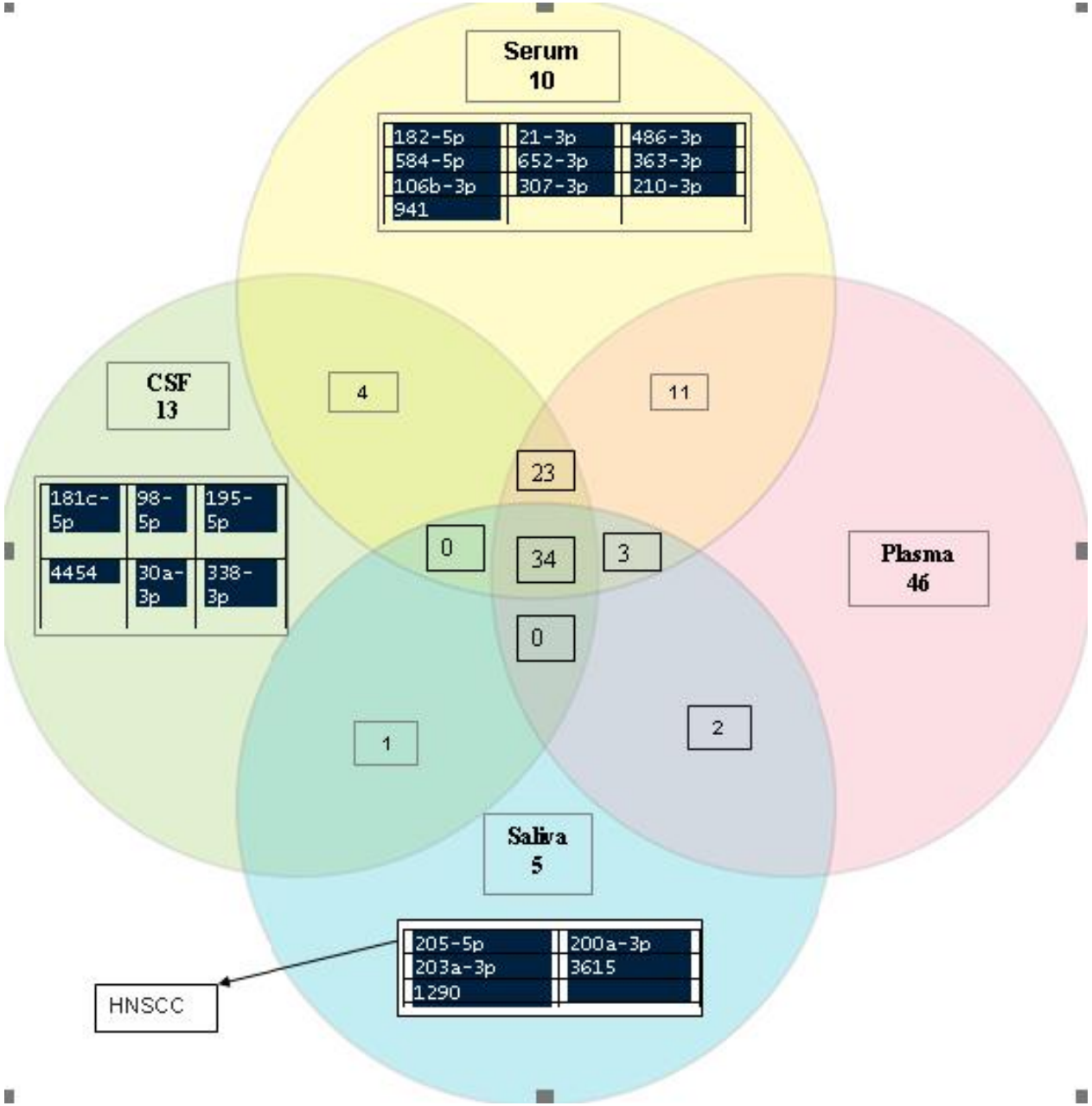
No. of expressed small piRNA split by biofluid and threshold



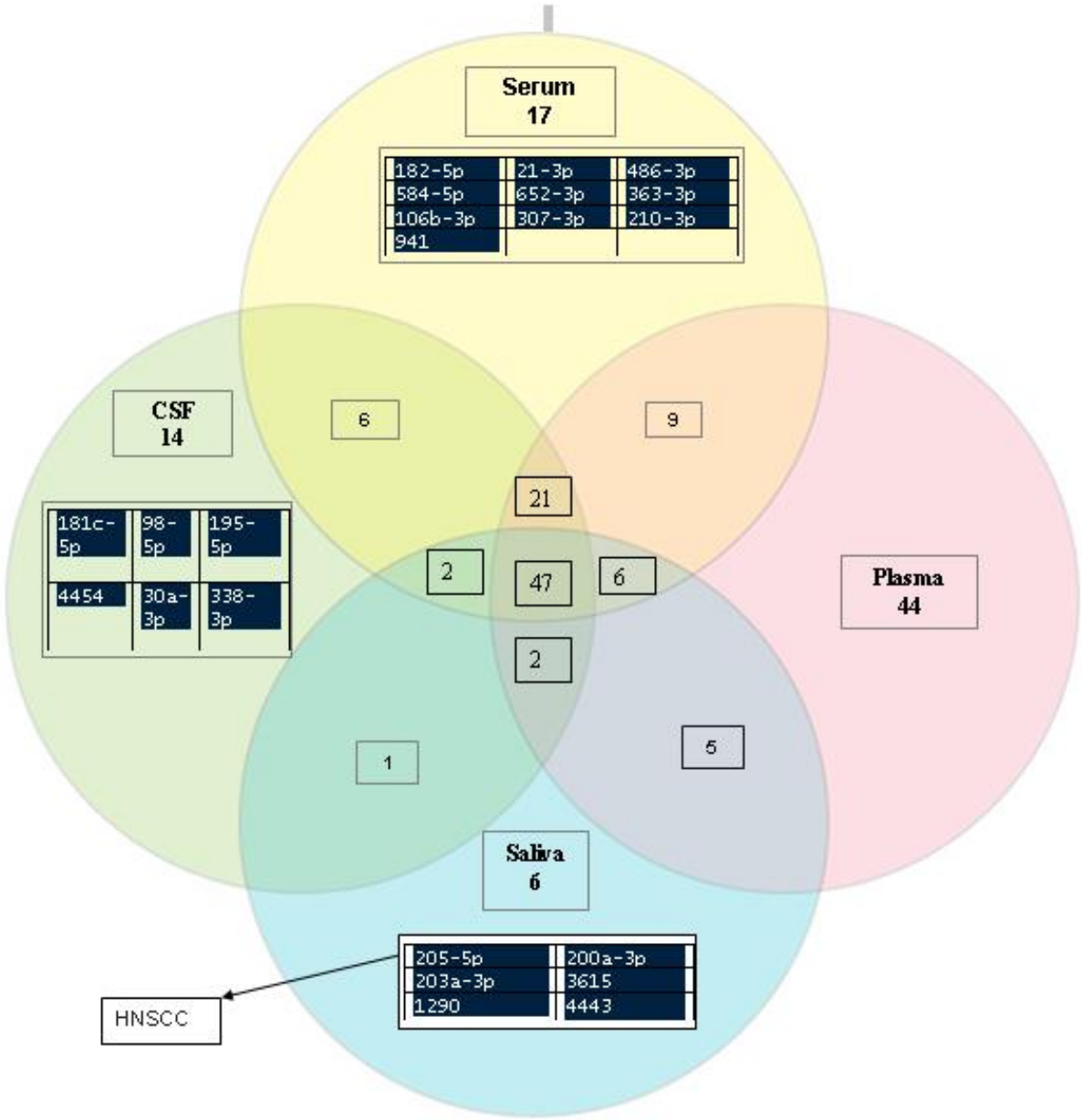
What are those MIRs?

Are they shared between biofluids?

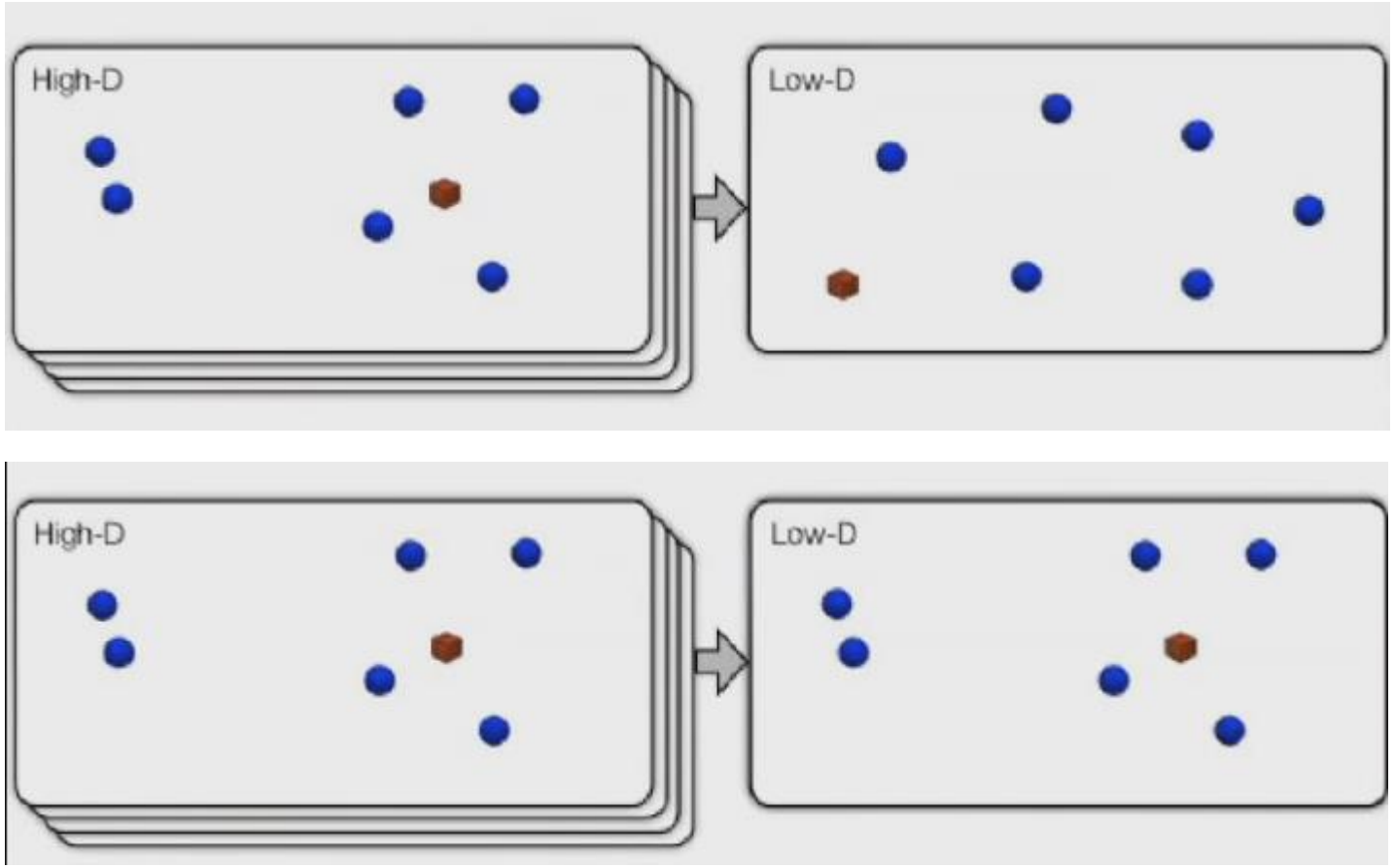
Threshold 25



Threshold 100

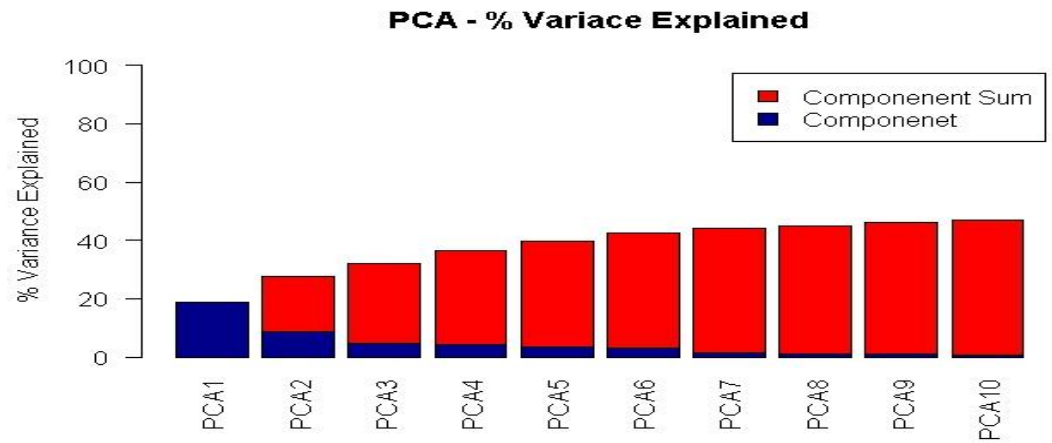


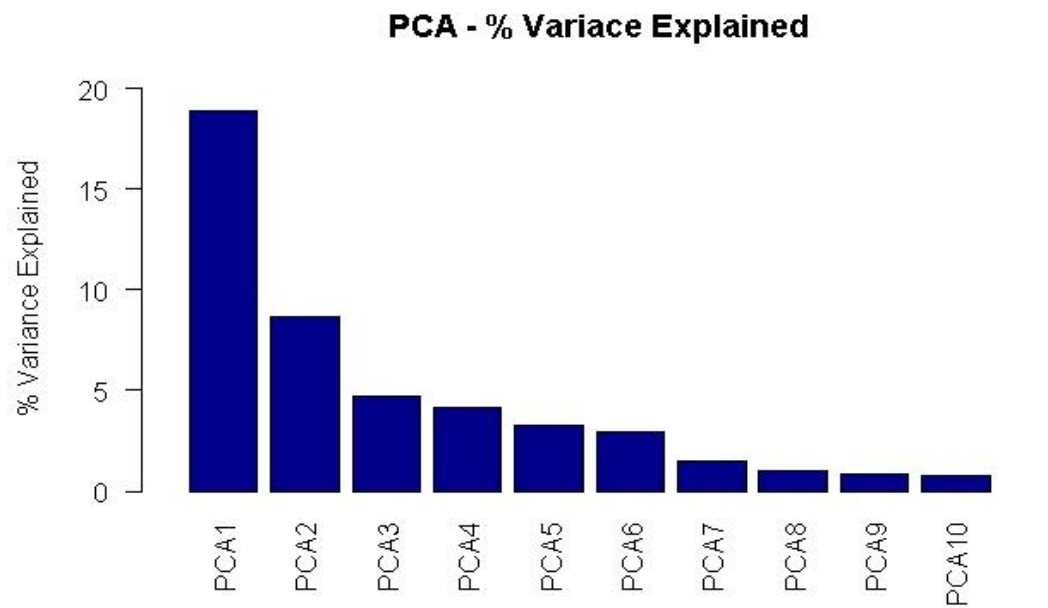
Visualization - Dimensionality reduction



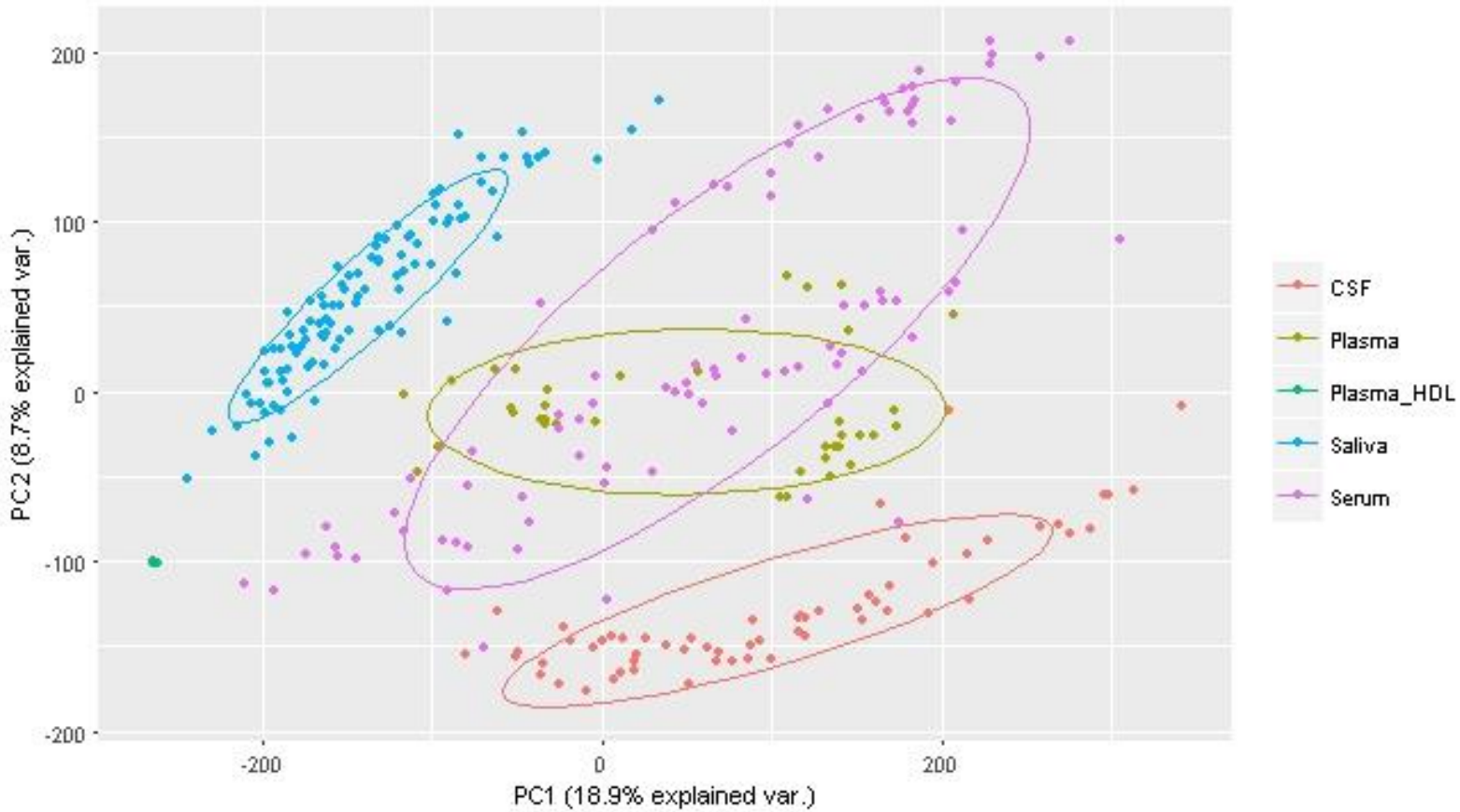
PCA

- 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

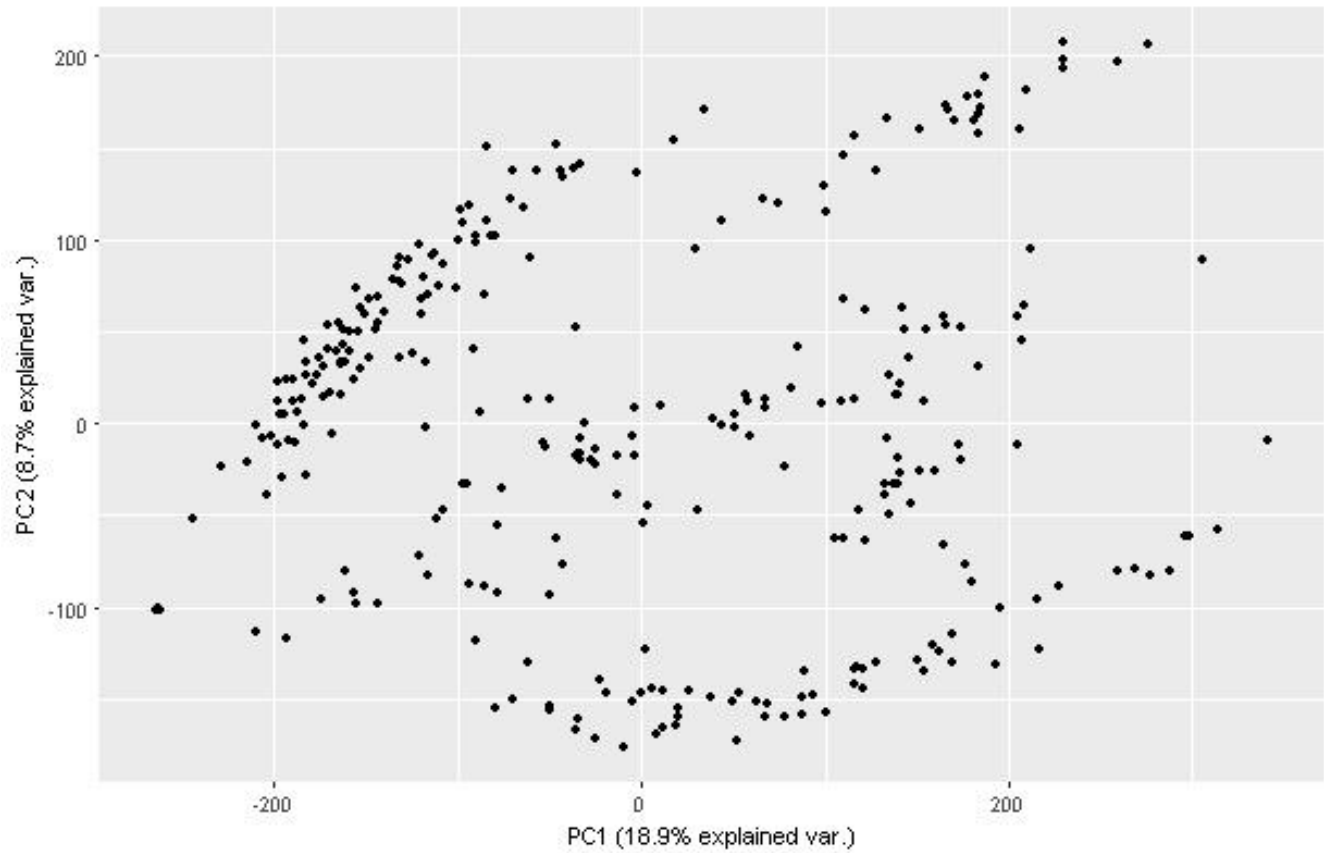


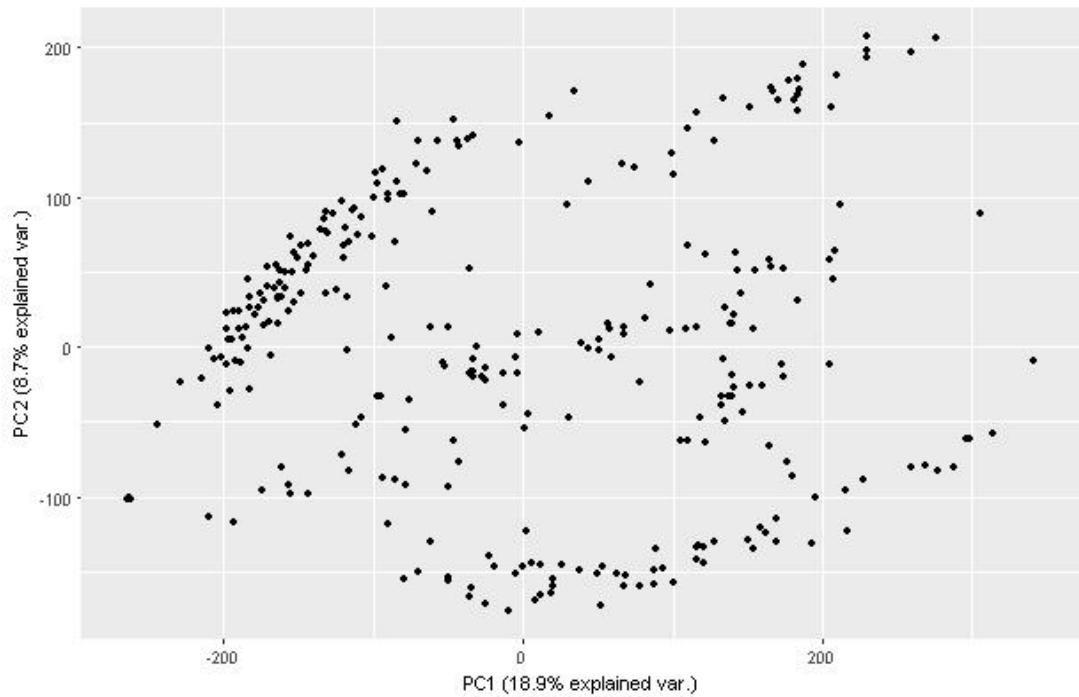


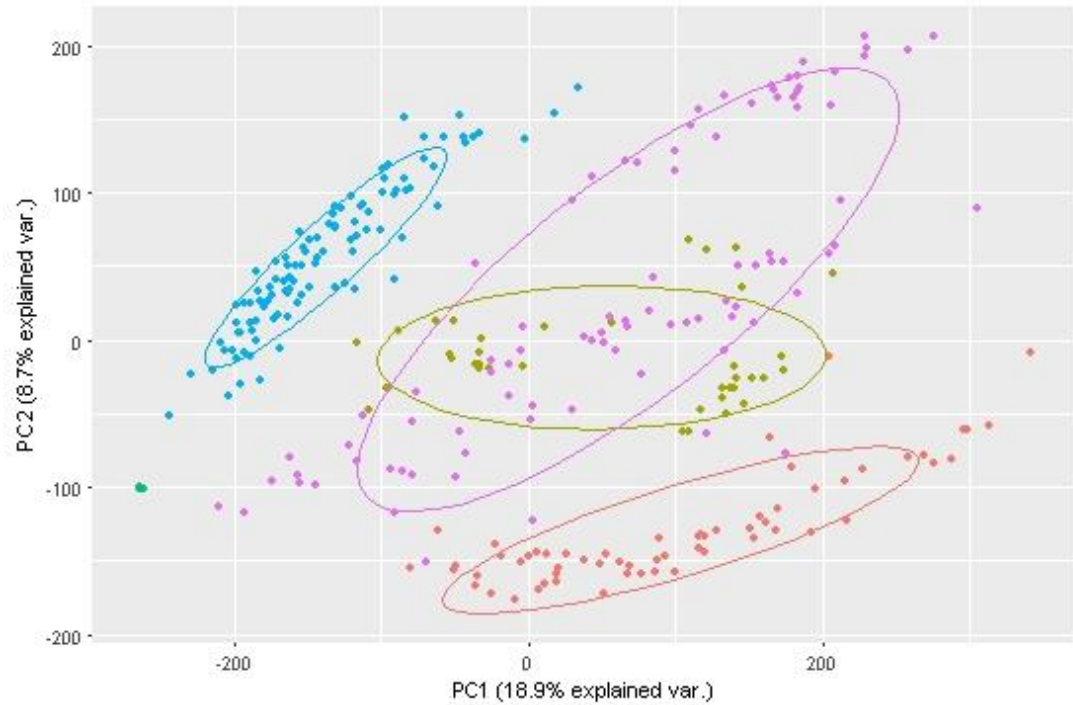
PC2 vs. PC1


200
100
0
-100
-200
PC2 (8.7% explained var.)
-200
0
200
PC1 (18.9% explained var.)
CSF
Plasma
Plasma_HDL
Saliva
Serum

PC2 vs. PC1

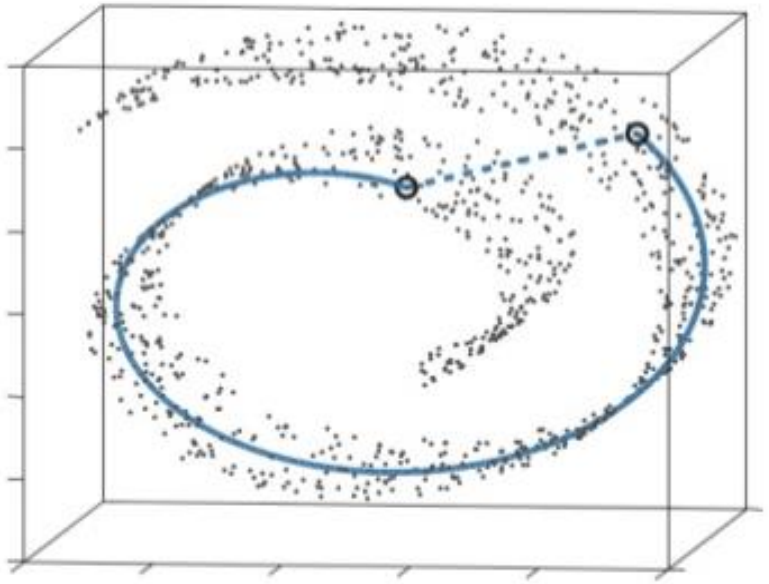

200
100
0
-100
PC2 (8.7% explained var.)
-200
0
200
PC1 (18.9% explained var.)


200
100
0
-100
PC2 (8.7% explained var.)
-200
0
200
PC1 (18.9% explained var.)

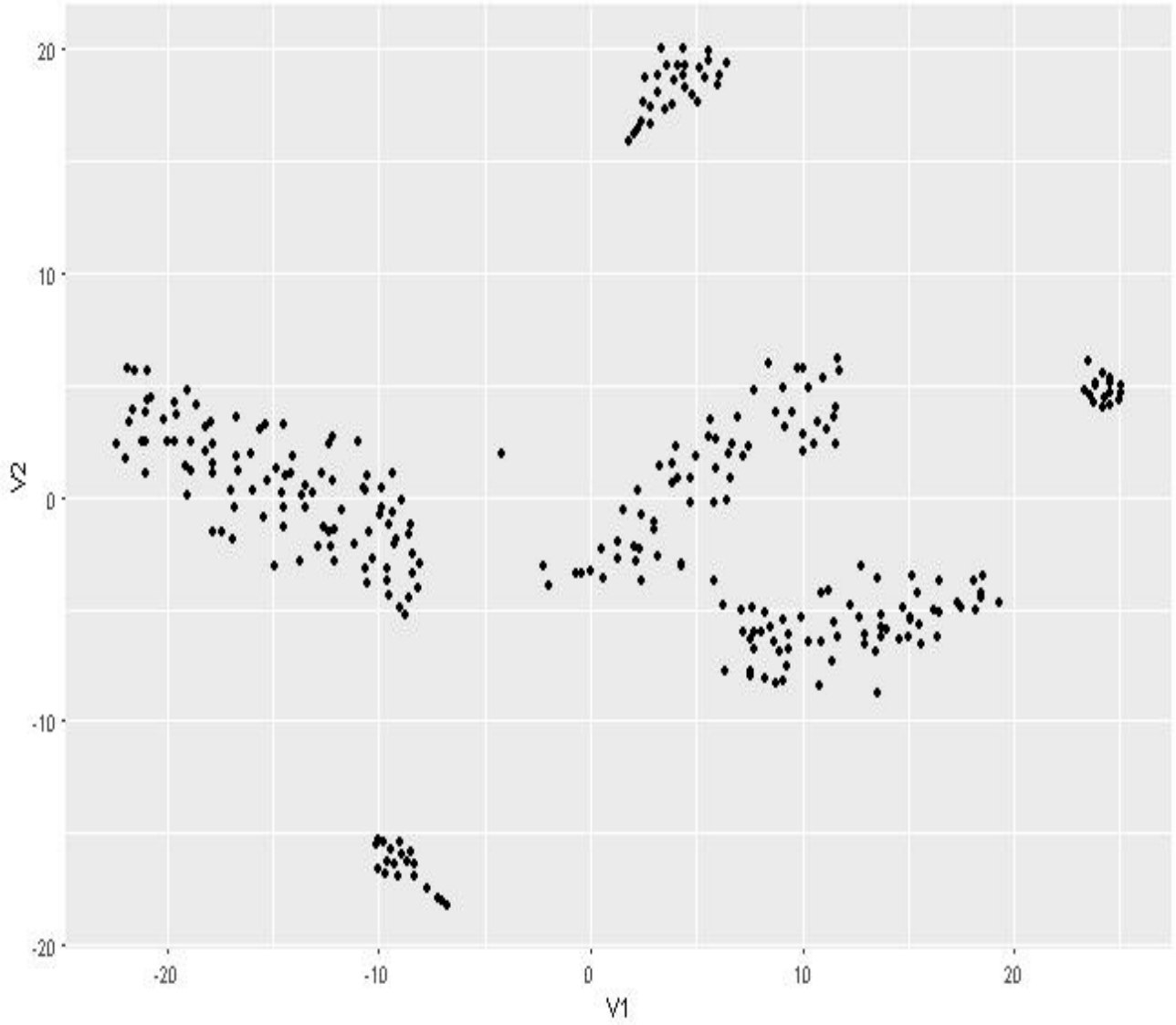

200
100
0
-100
-200
PC2 (8.7% explained var.)
-200
0
200
PC1 (18.9% explained var.)
CSF
Plasma
Plasma_HDL
Saliva
Serum

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

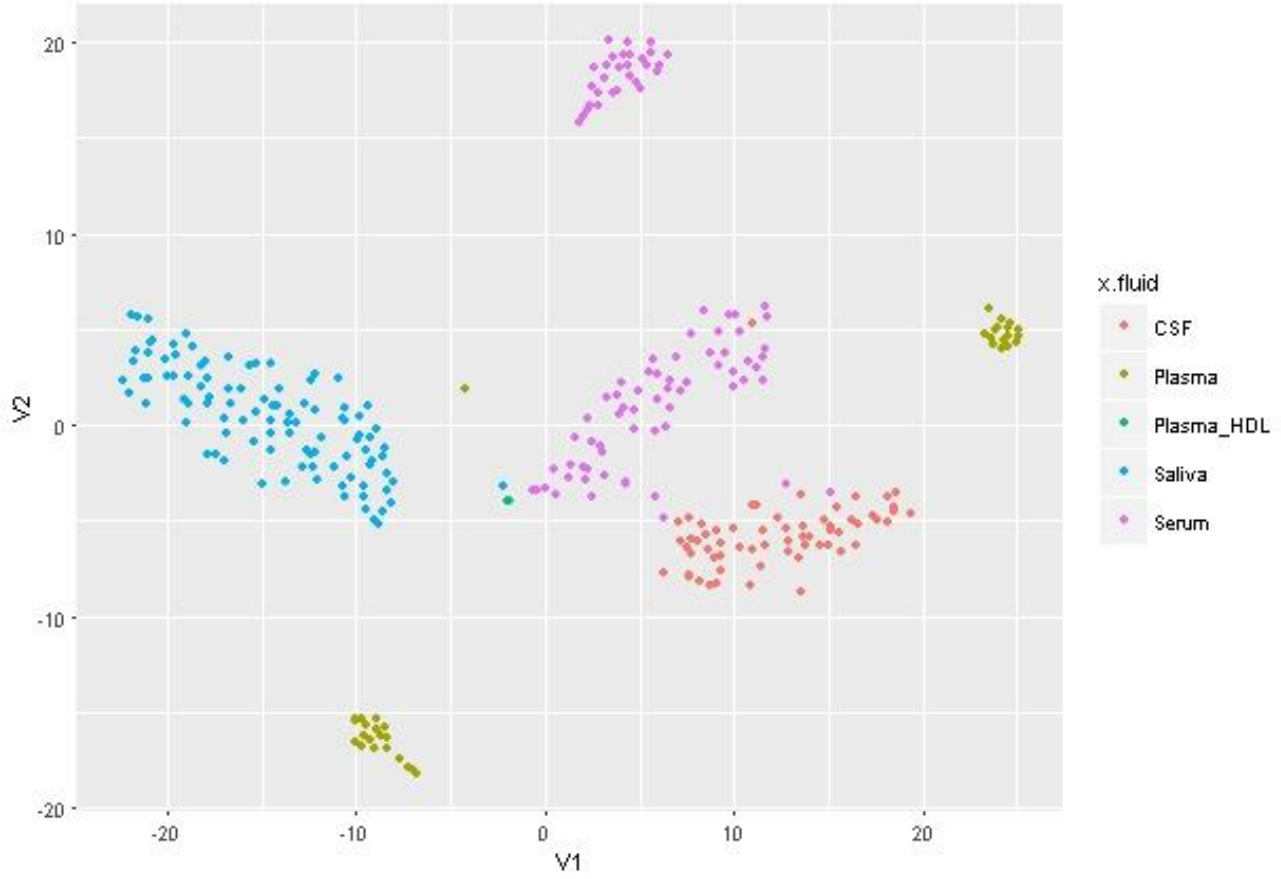


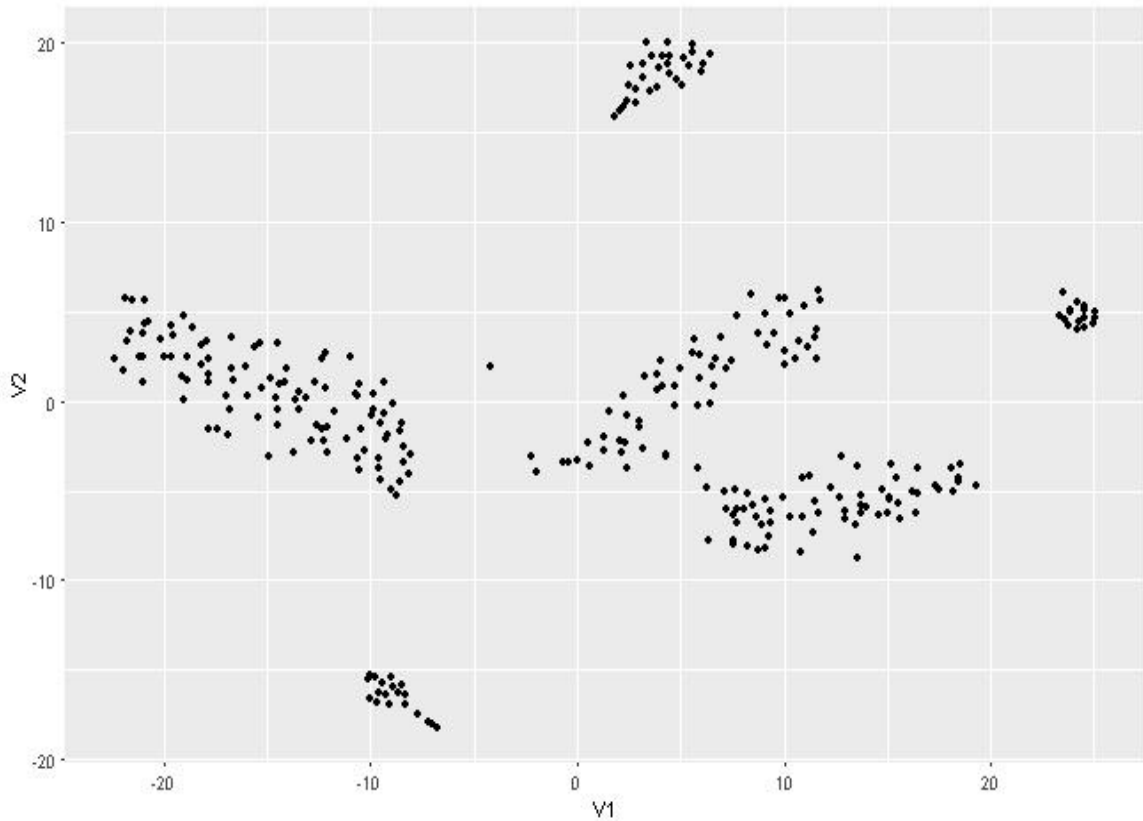
TSNE output

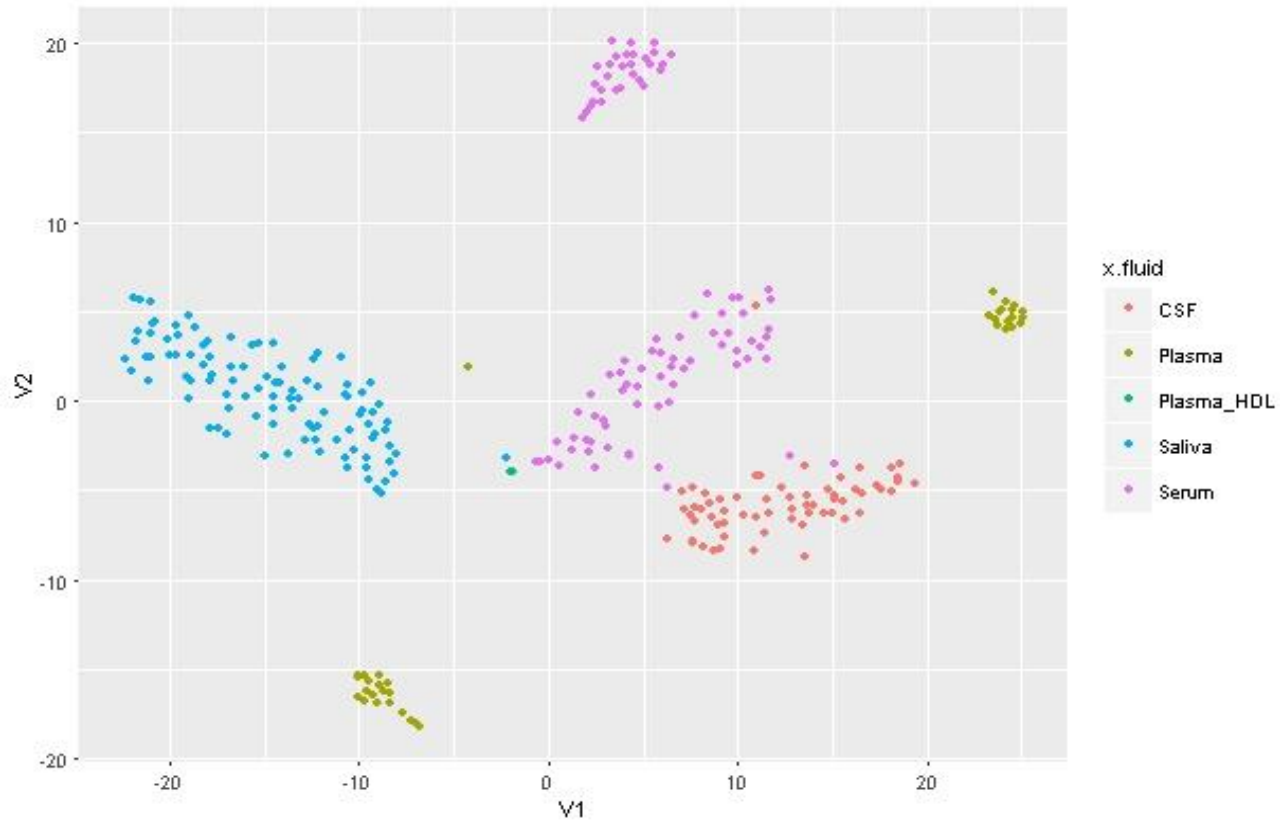


TSNE output

With biofluids labels

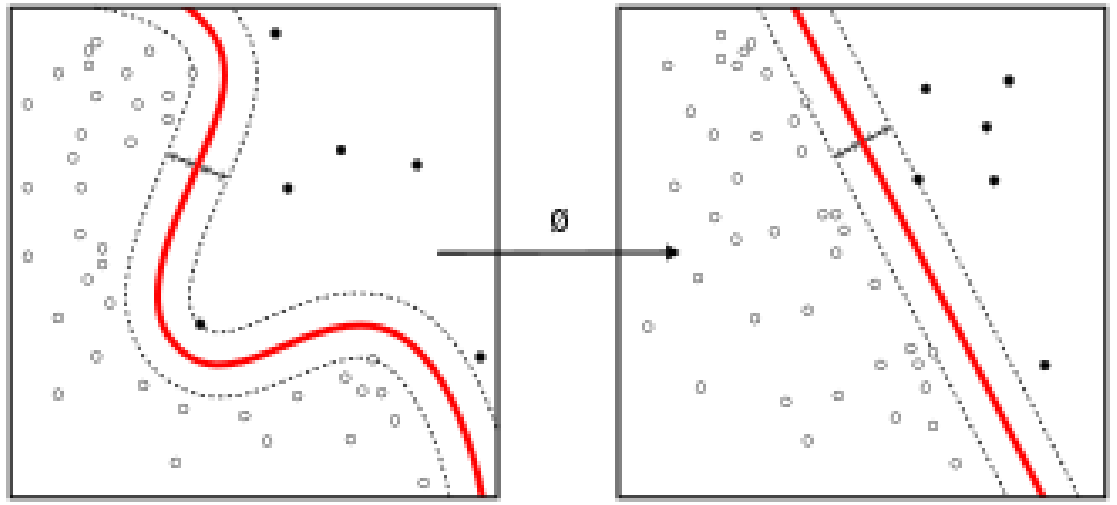



20
10
0
-10
-20
V2
-20
-10
0
10
20
V1


20
10
0
-10
-20
V2
-20
-10
0
10
20
V1
x.fluid
CSF
Plasma
Plasma_HDL
Saliva
Serum

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

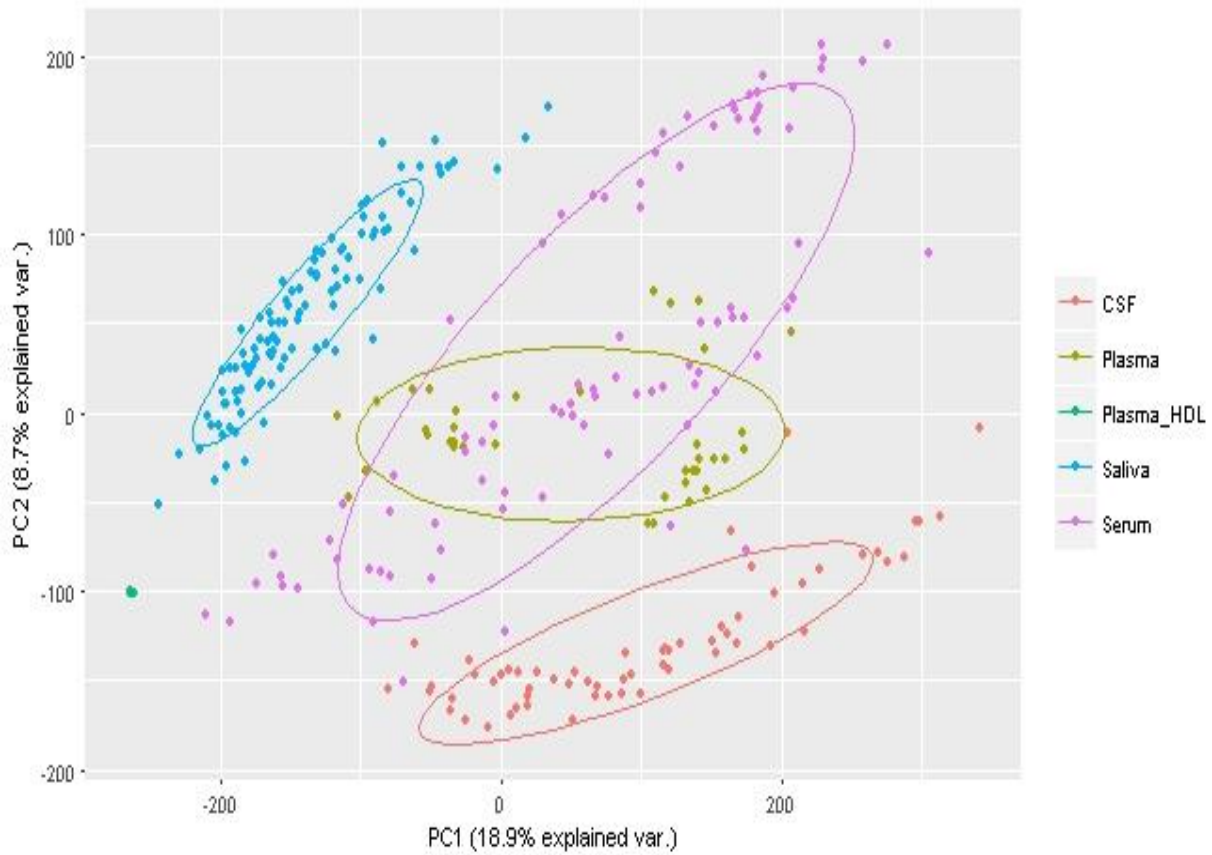
- Look at the low dimensional space
- Lay points in that map
- Represent each high dimensional object by a point on the map
- A probability that measures similarity of 2 points in the low dimensional map
- We want Qij to reflect the similarities in Pij as well as possible

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

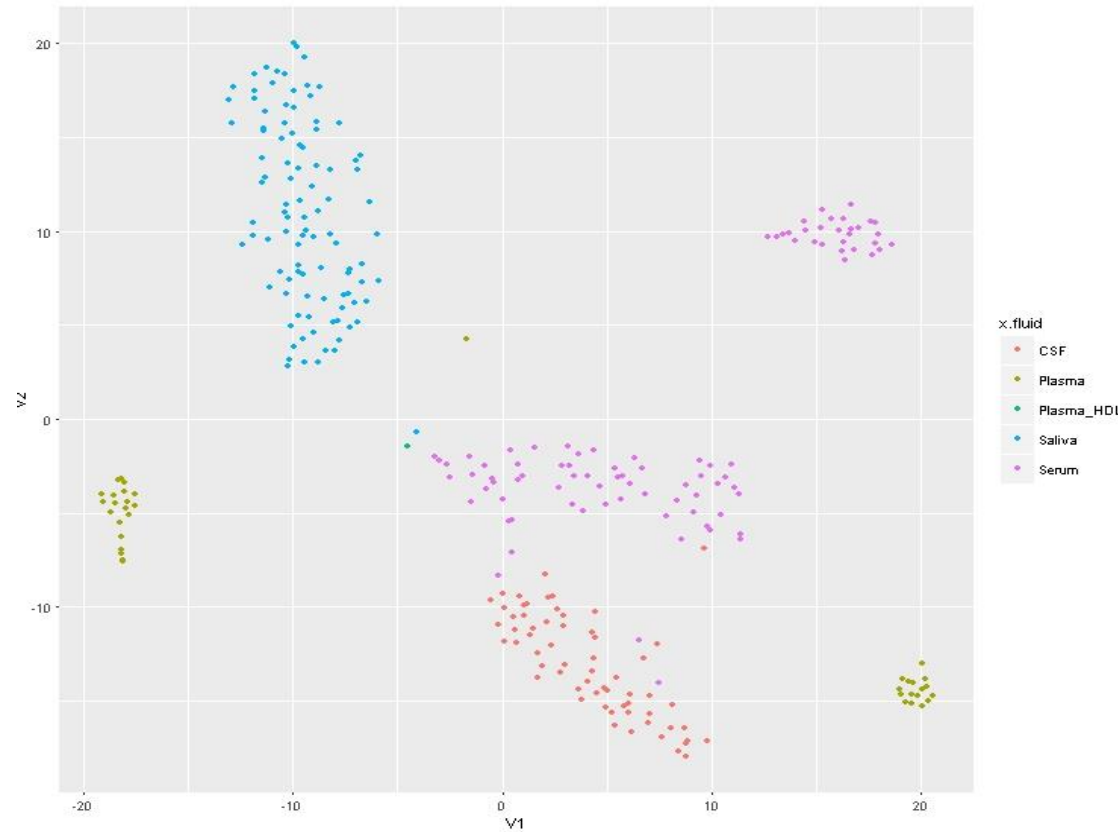
$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Comparison

PCA



TSNE



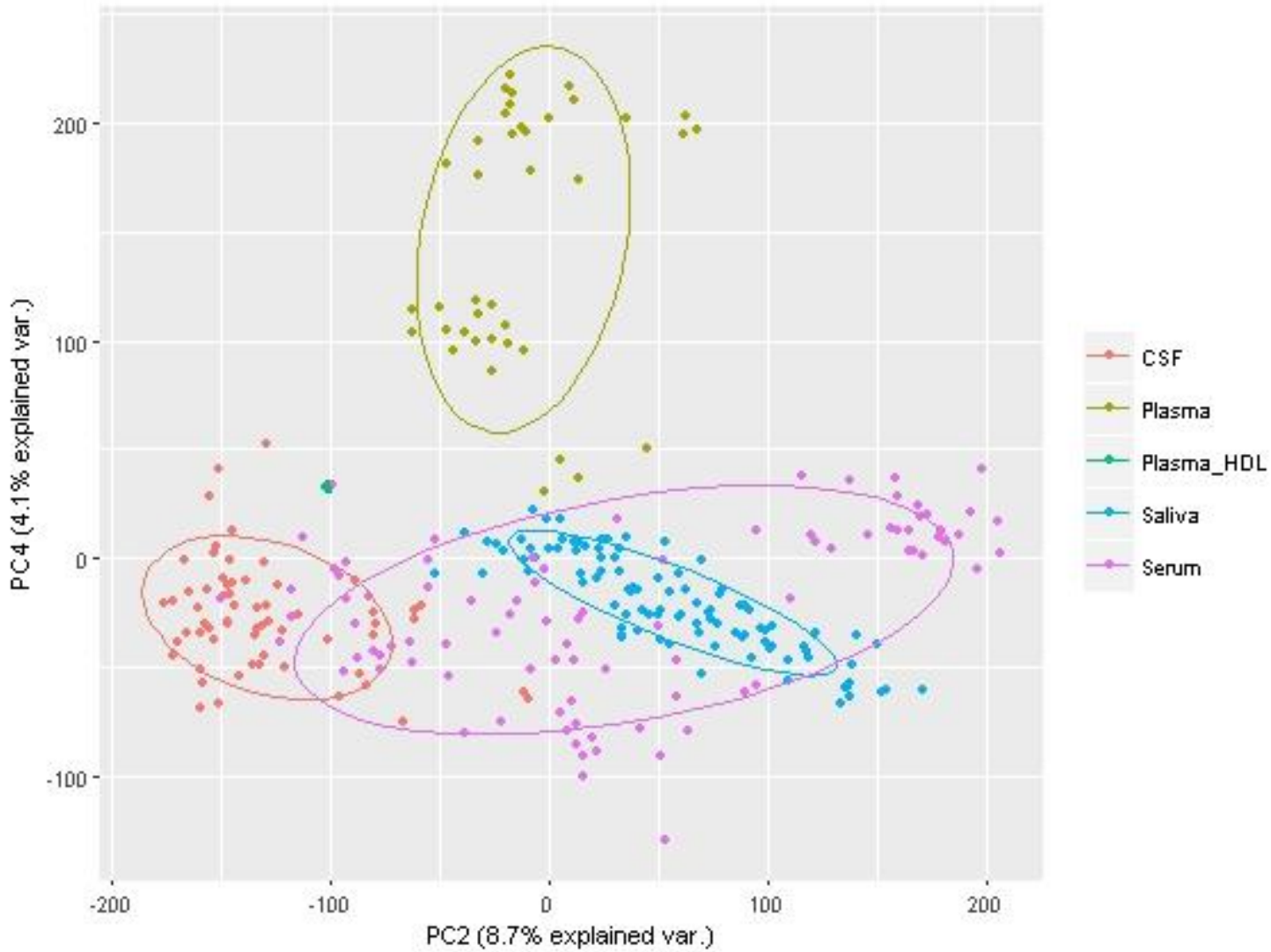
Conclusion & Potential directions

- 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

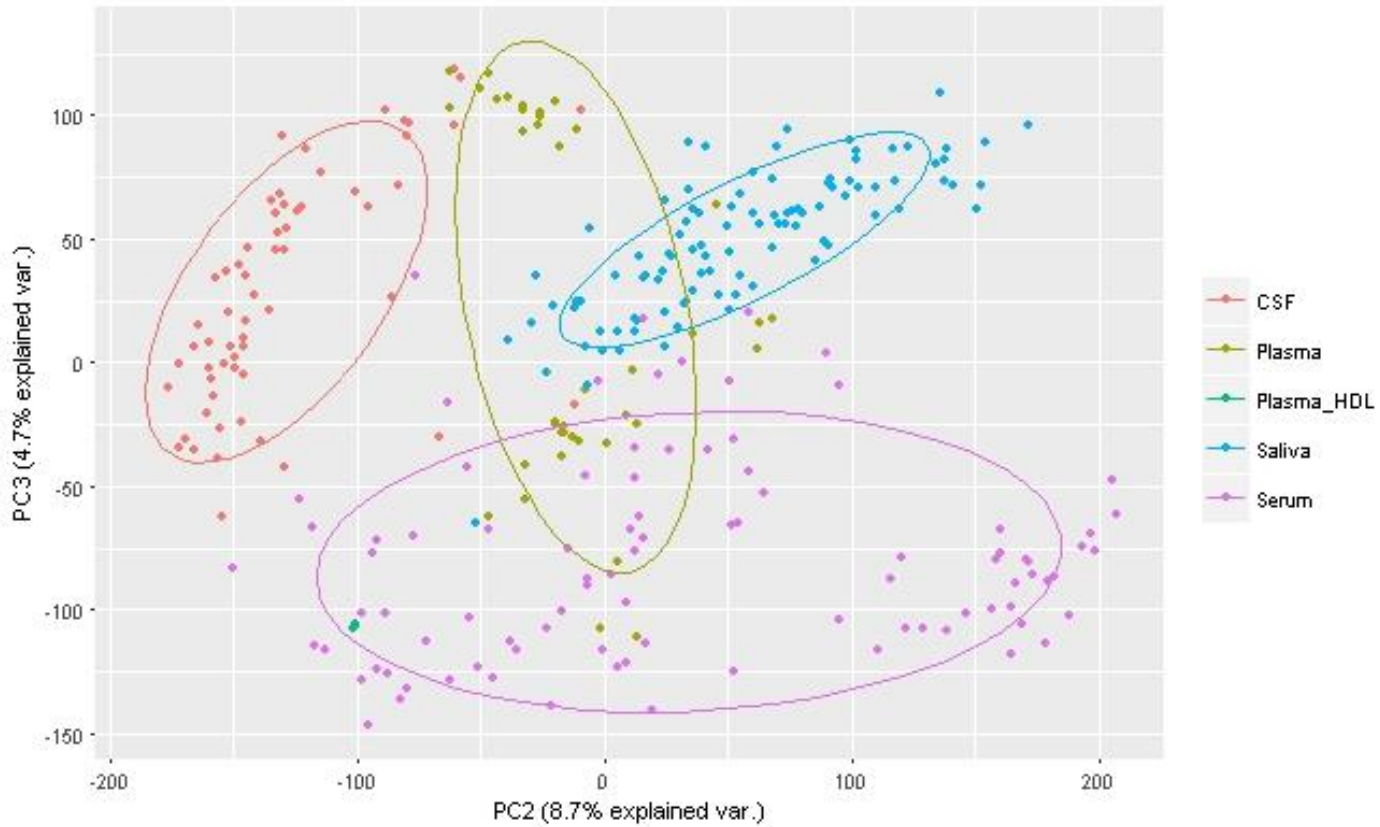
Thank you

- Joel Rozowsky
- Mark Gerstein

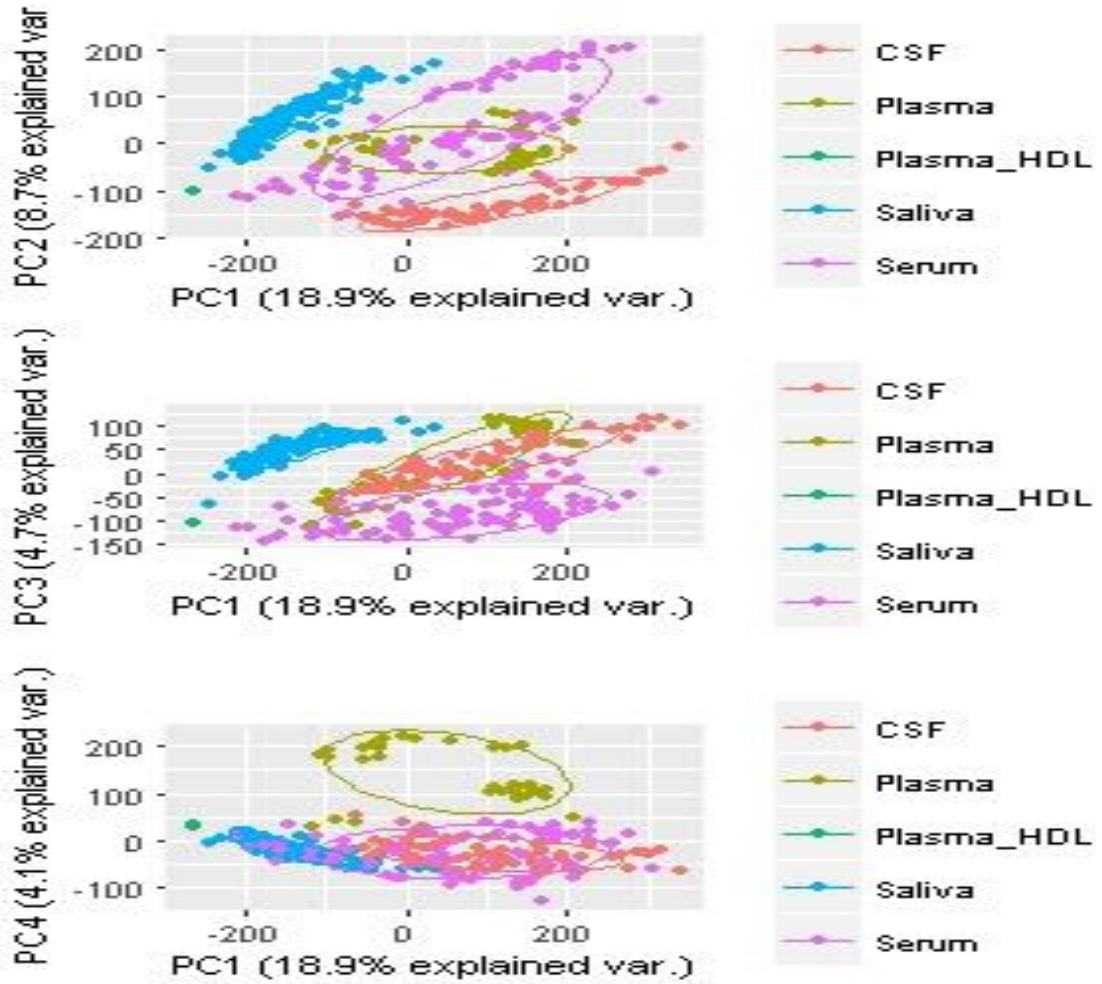
PC4 Vs. PC2

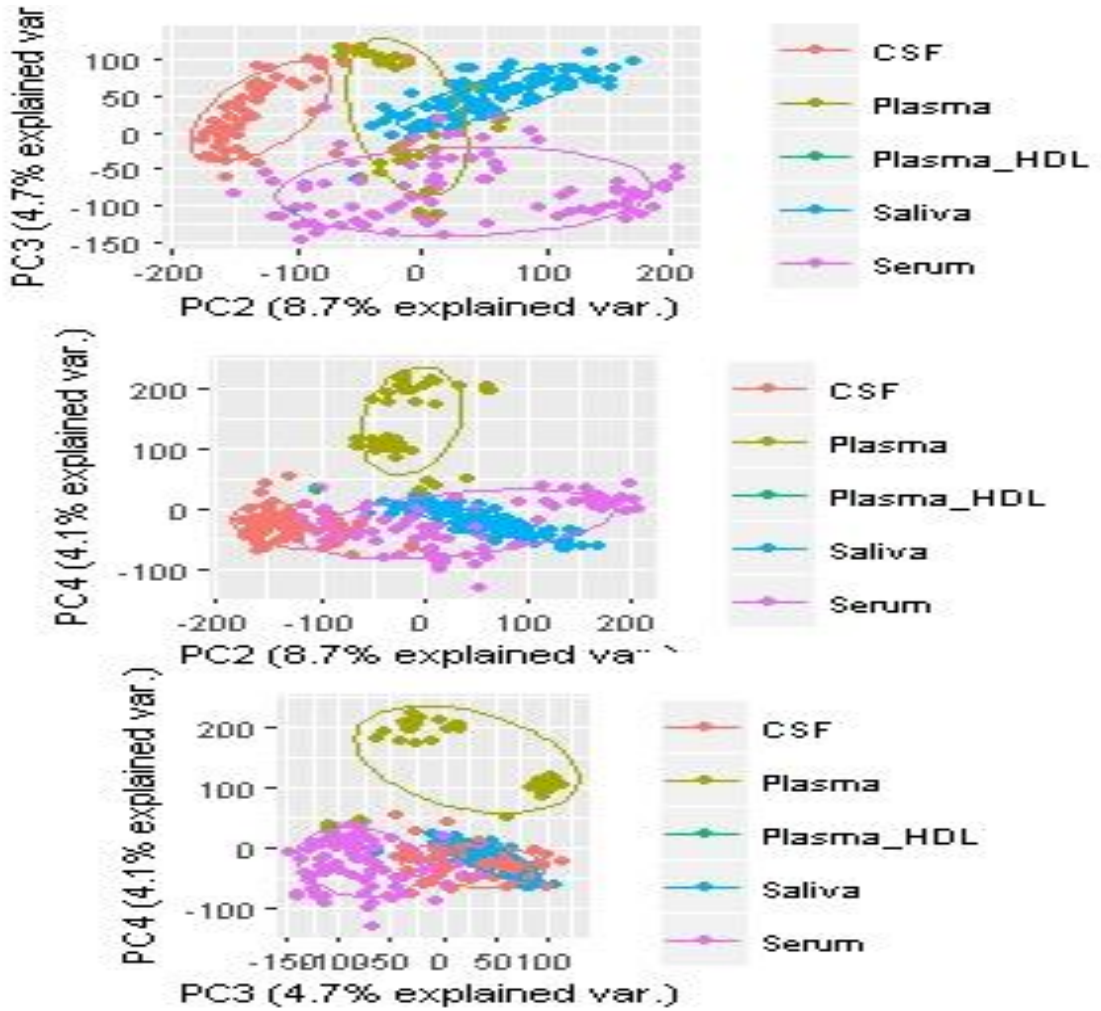

200
100
0
-100
PC4 (4.1% explained var.)
-200
-100
0
100
200
PC2 (8.7% explained var.)
CSF
Plasma
Plasma_HDL
Saliva
Serum

PC3 Vs. PC2

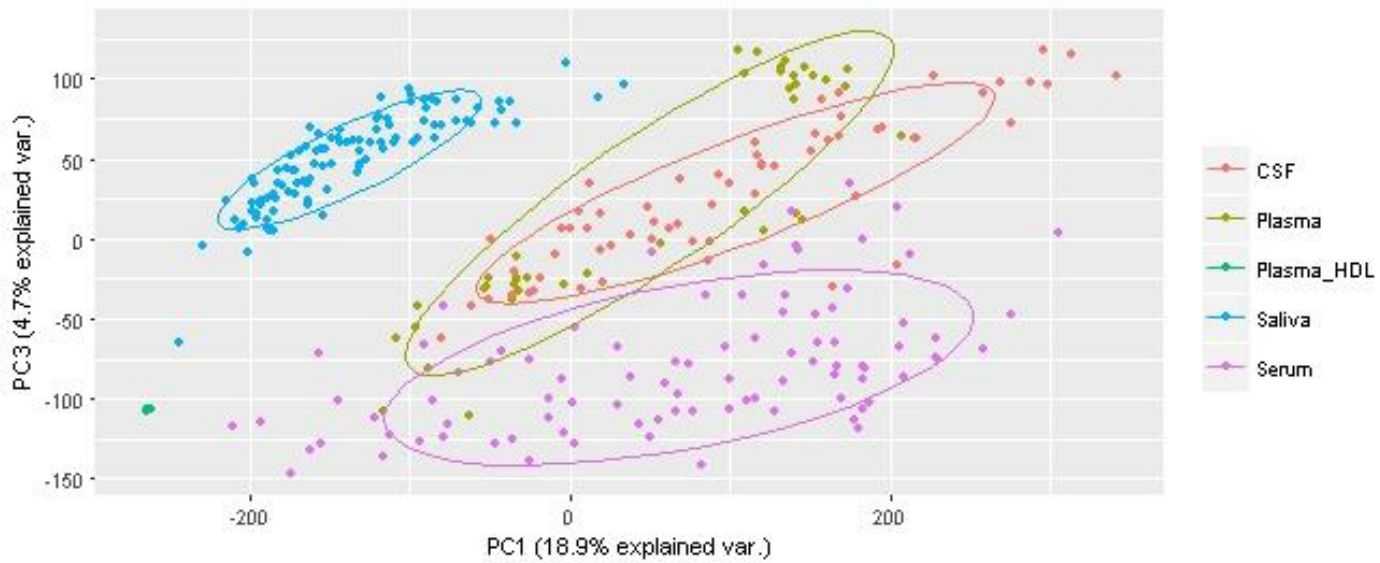

PC3 (4.7% explained var.)
PC2 (8.7% explained var.)
100
50
0
-50
-100
-150
-200
-100
0
100
200
CSF
Plasma
Plasma_HDL
Saliva
Serum

PCA – Center = True; Scale = False

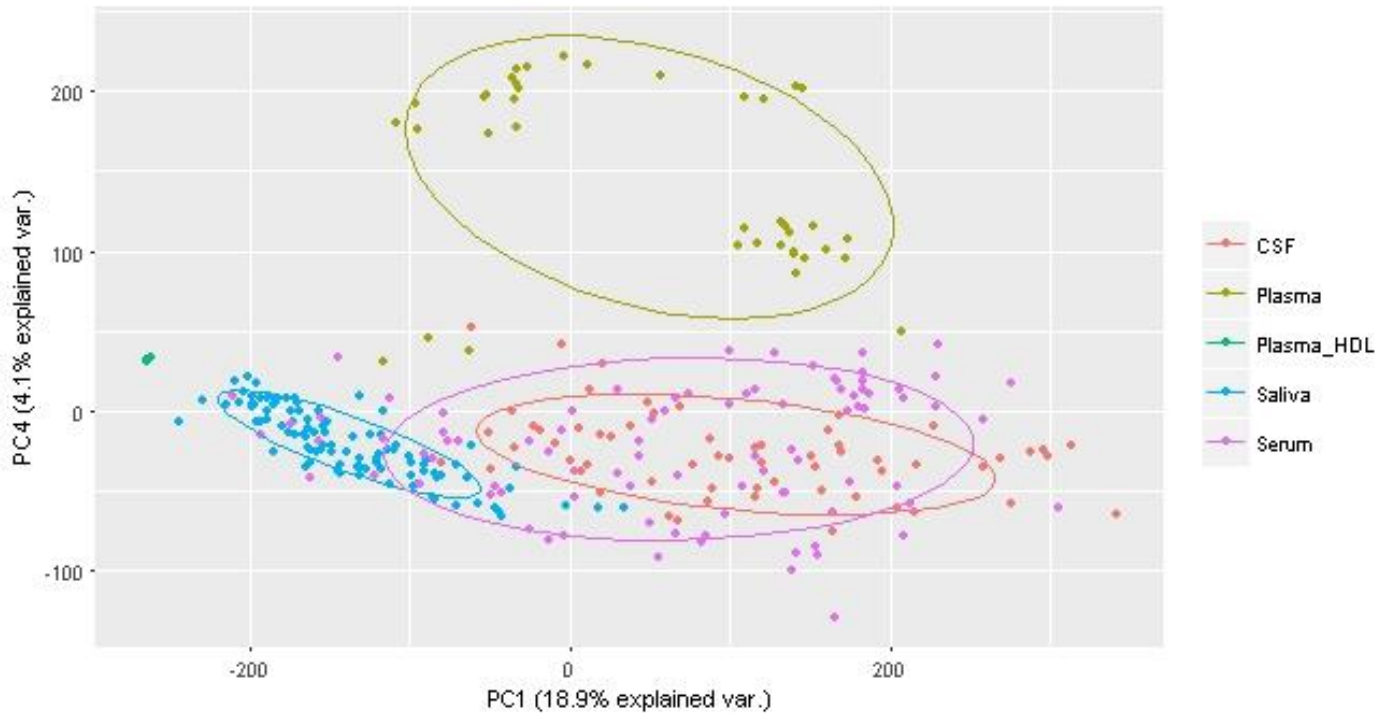




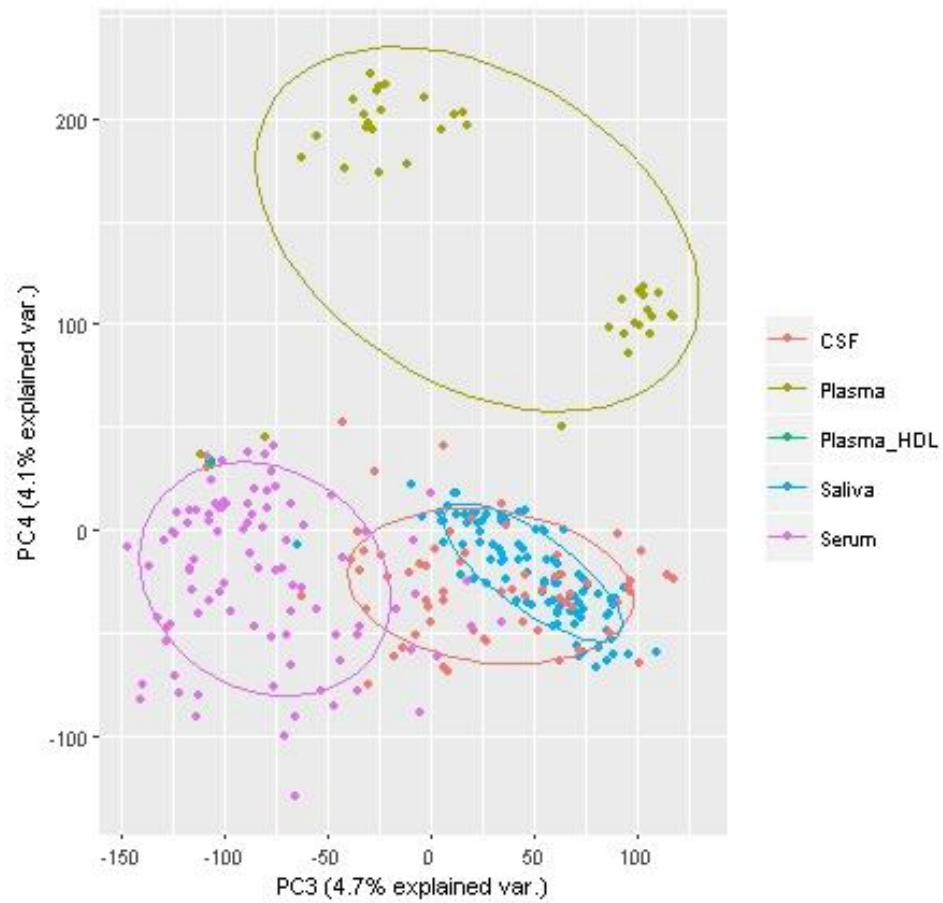
PC3 Vs. PC1


100
50
0
-50
-100
-150
PC3 (4.7% explained var.)
-200
0
200
PC1 (18.9% explained var.)
CSF
Plasma
Plasma_HDL
Saliva
Serum

PC4 Vs. PC1


200
100
0
-100
PC4 (4.1% explained var.)
-200
0
200
PC1 (18.9% explained var.)
CSF
Plasma
Plasma_HDL
Saliva
Serum

PC3 Vs. PC4


200
100
0
-100
PC4 (4.1% explained var.)
-150
-100
-50
0
50
100
PC3 (4.7% explained var.)
CSF
Plasma
Plasma_HDL
Saliva
Serum