

Integrative analysis of exRNA atlas datasets using dimensionality reduction techniques

Jonathan Warrell
Yale University
2017 - 04 - 05



tSNE: t-Distributed Stochastic Neighbor Embedding

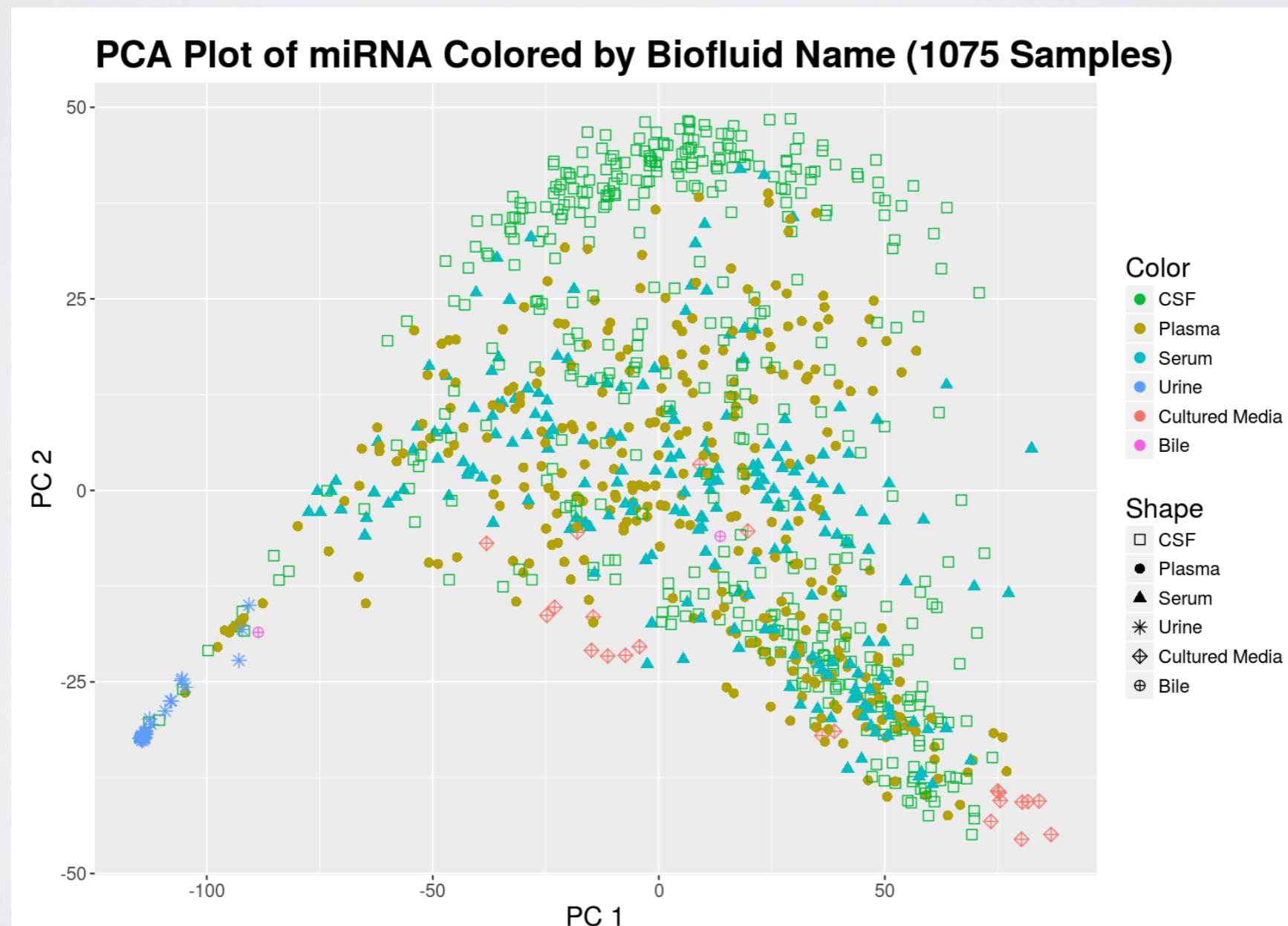
- tSNE performs non-linear dimensionality reduction on high-dimensional data (van der Maaten and Hinton, JMLR 2008) providing more flexibility than linear techniques such as PCA
- Matches the local density around each data point (i) by minimizing the KL -divergence between high (P) and low (Q) dimensional local distributions

$$C = \sum_i KL(P_i || Q_i)$$

- Widely used in machine learning to visualize and discover structure in high dimensional data

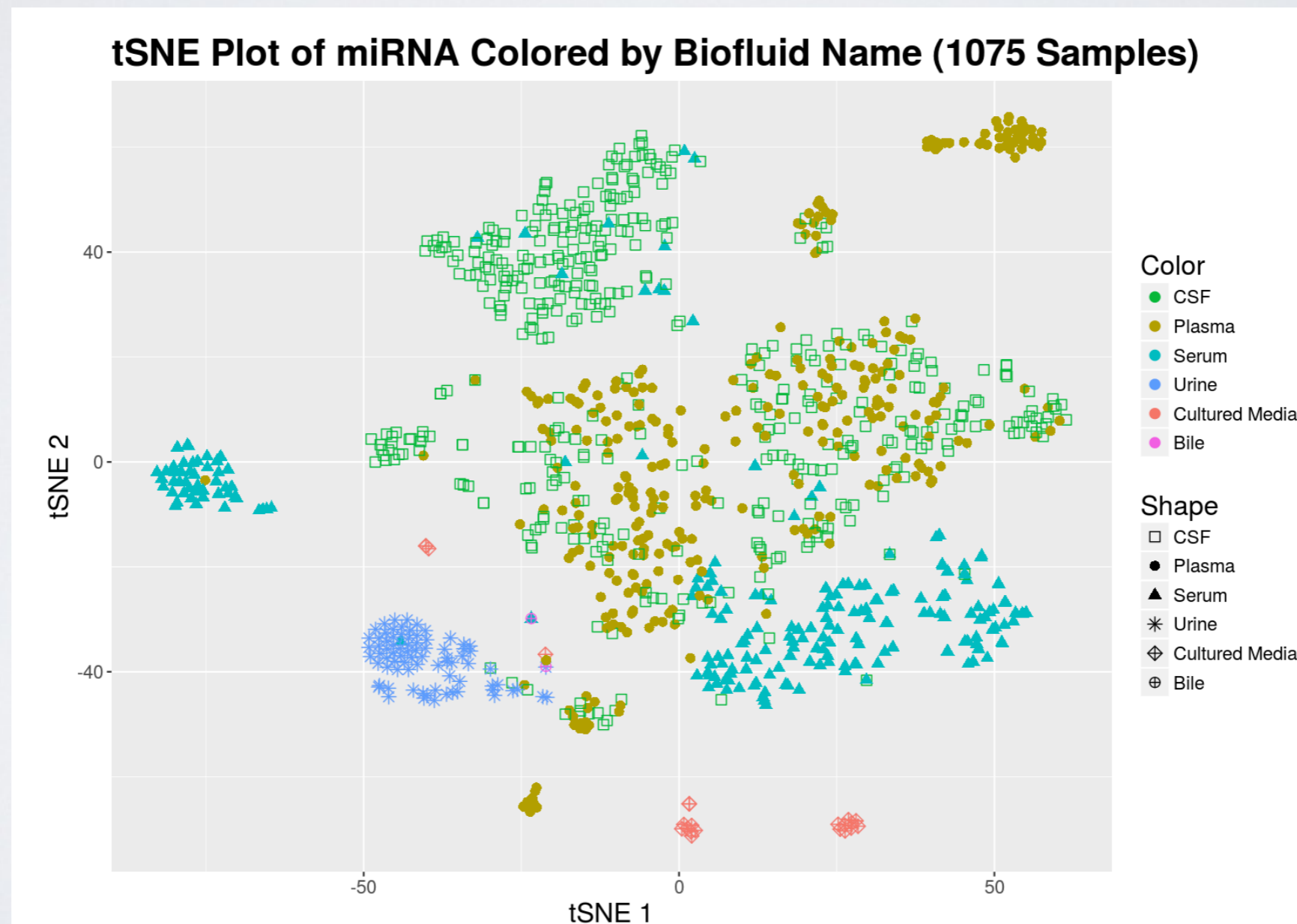
Dimensionality reduction tool for the exRNA atlas

- Tool in Genboree for performing tSNE and PCA on exRNA atlas data



Dimensionality reduction tool for the exRNA atlas

- Tool in Genboree for performing tSNE and PCA on exRNA atlas data



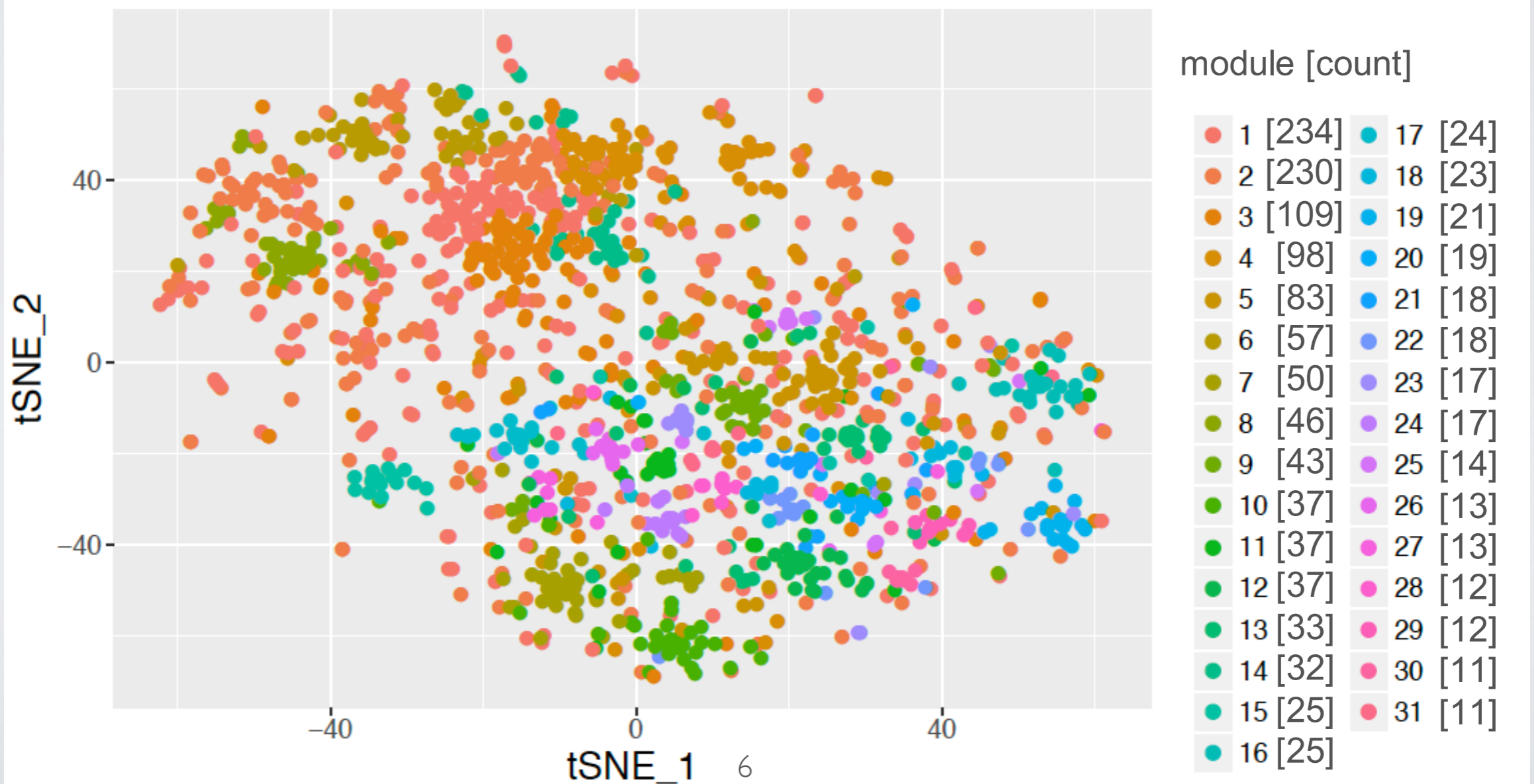
WGCNA: Weighted gene coexpression network analysis

- WGCNA finds modules of coexpressed genes by clustering expression patterns across samples/tissues/biofluids (Zhang and Horvath, 2005)
- First builds network using correlations between expression patterns, then hierarchically clusters genes by connectivity and applies a dynamic tree cut algorithm
- Number of clusters is controlled by the minimum module size, and the network connectivity
- Module expression patterns summarized by applying linear dimensionality reduction (SVD) to identify ‘eigengenes’

Visualizing miRNA modules from the exRNA atlas using tSNE

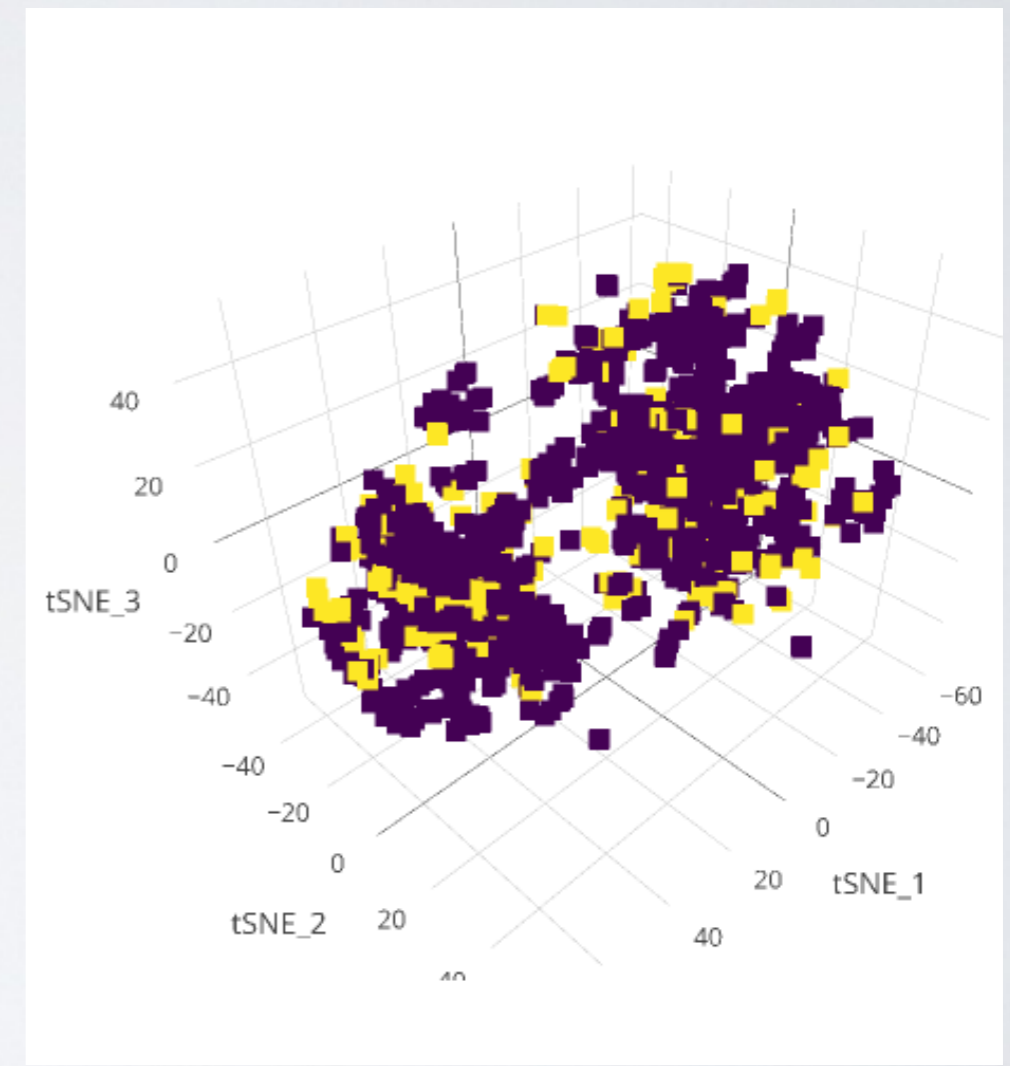
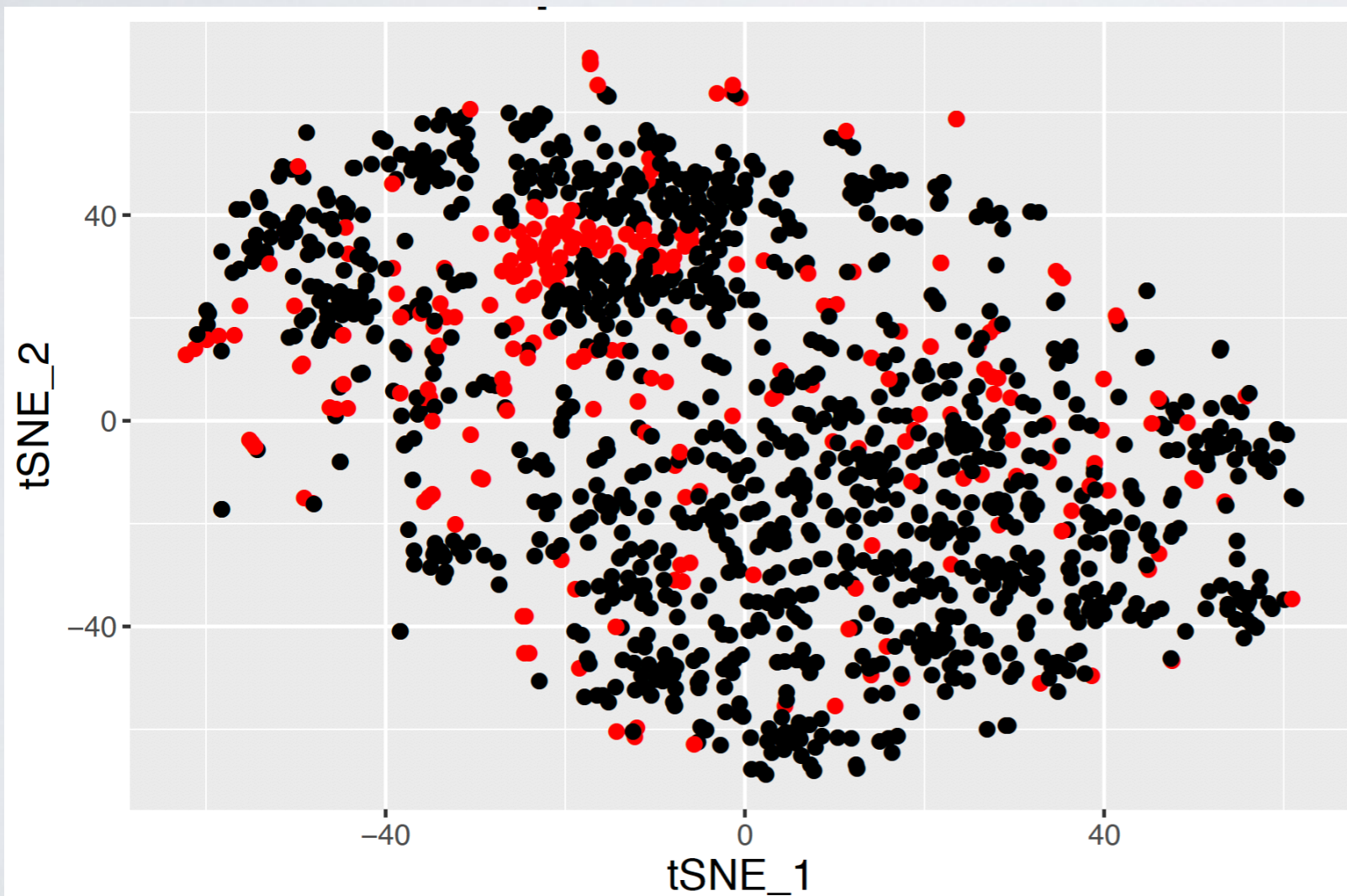
- N=48 samples including controls and 5 placental dysfunction conditions (Louise Laurent); 2411 miRNAs total, 1419 miRNAs clustered

miRNA modules, tSNE visualization



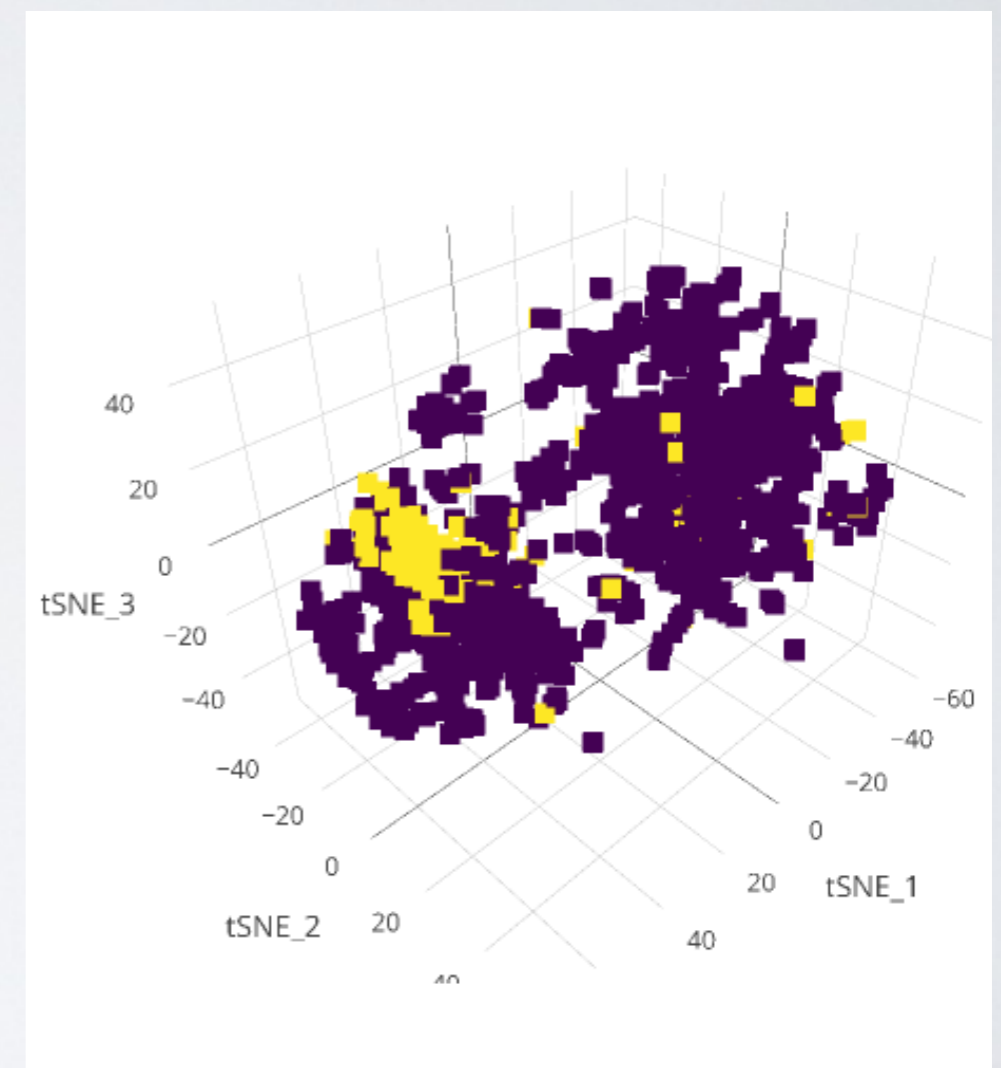
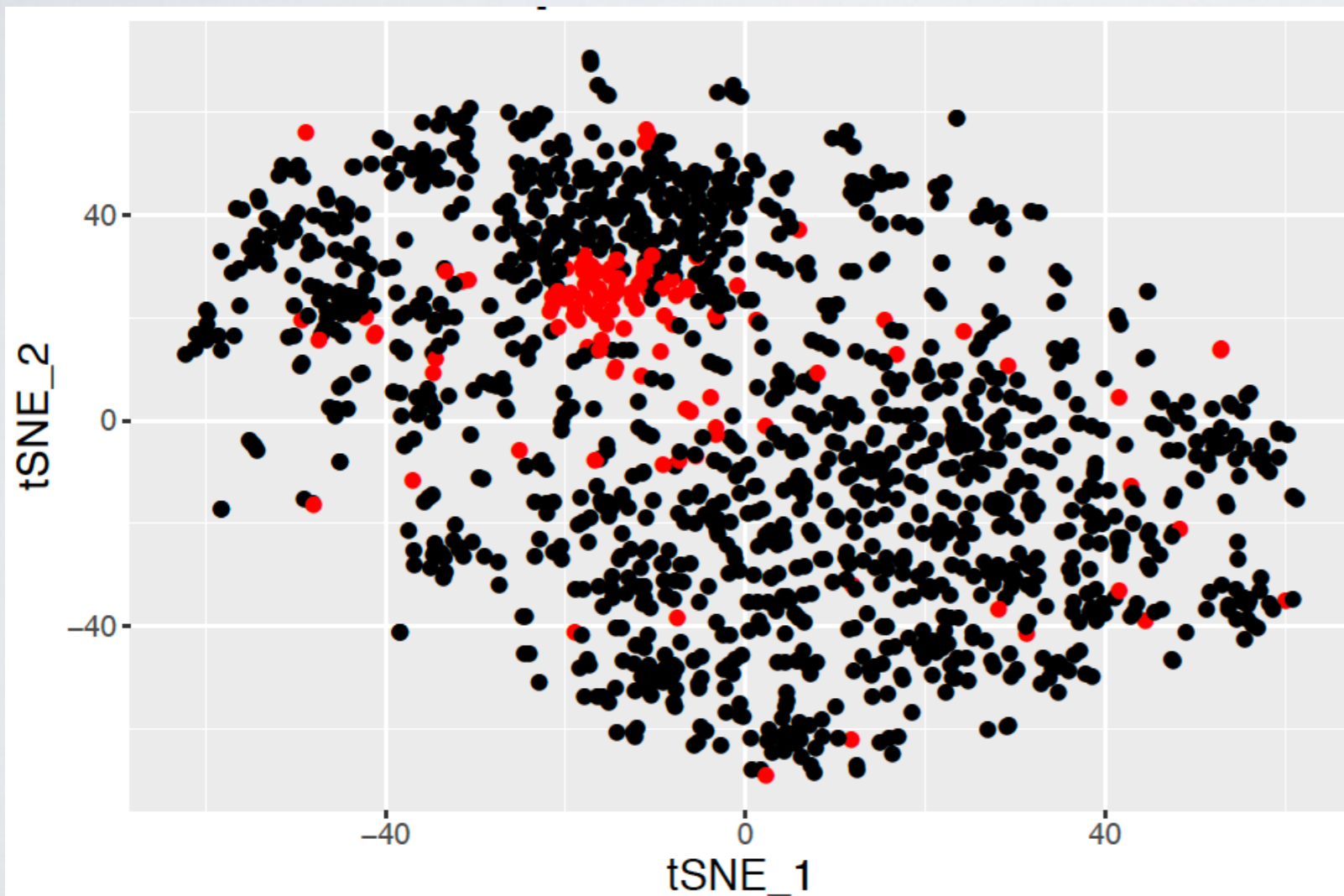
2d/3d tSNE for individual modules

- Module 1, 234 miRNAs: hsa-miR-92a-3p, hsa-miR-25-3p, hsa-let-7a-5p, hsa-let-7g-5p, hsa-miR-191-5p...



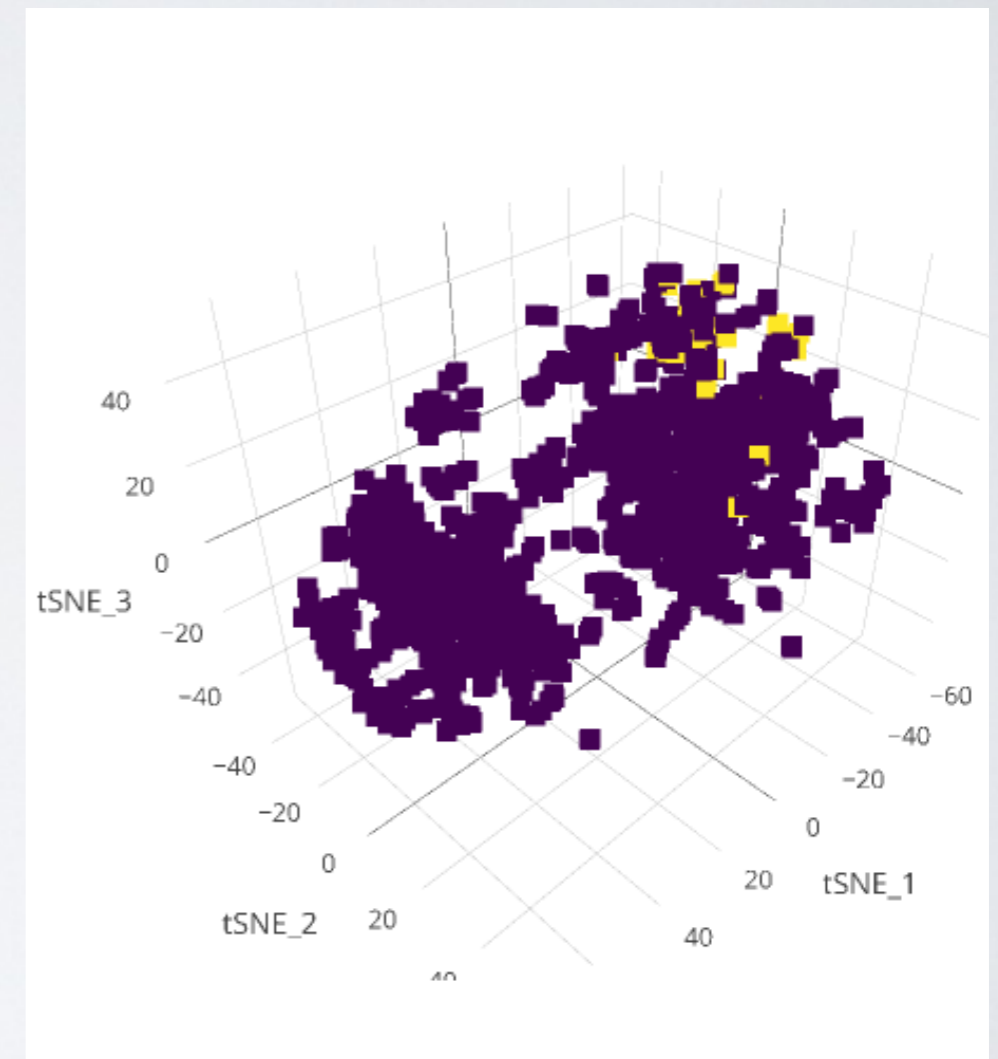
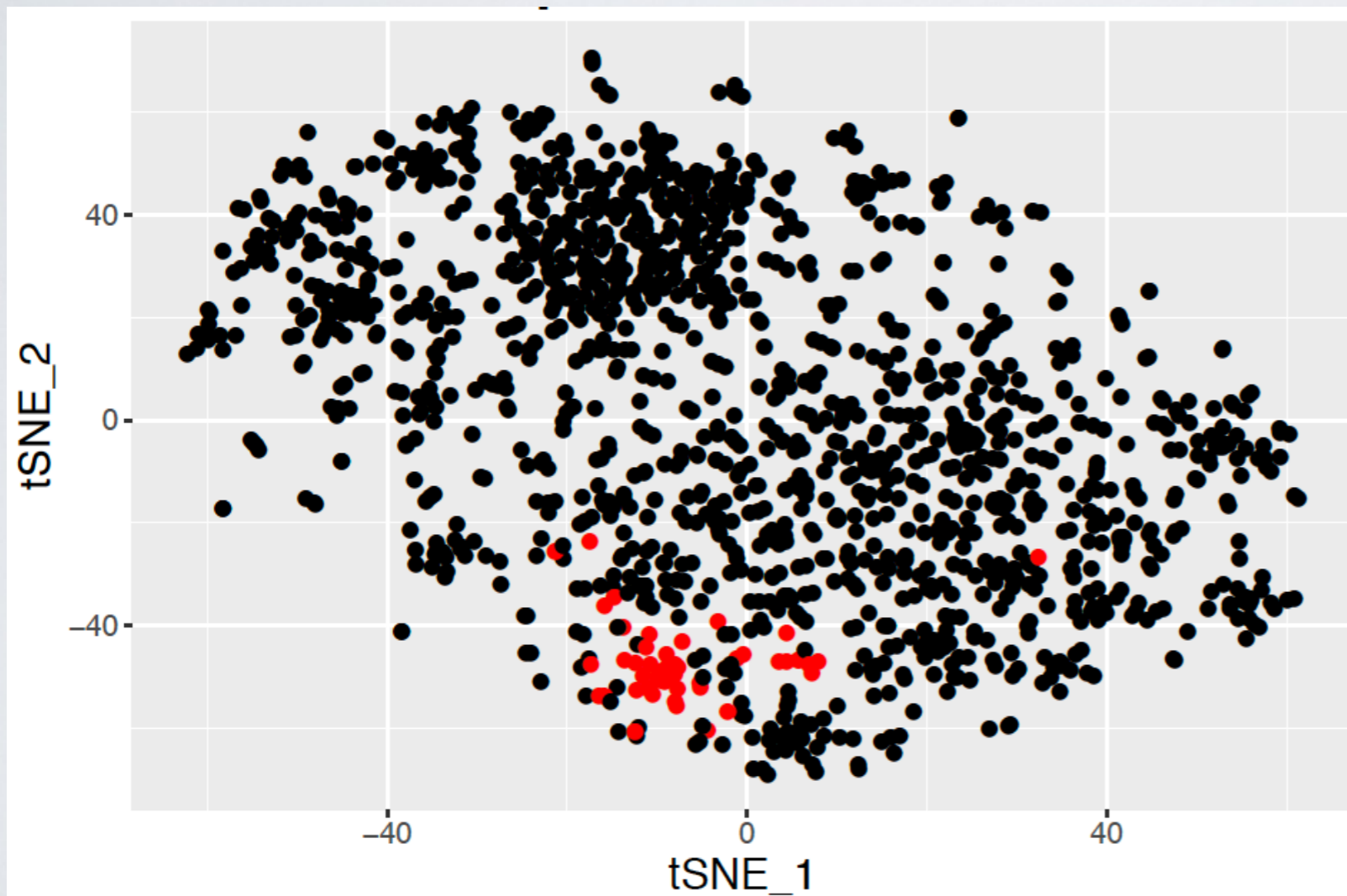
2d/3d tSNE for individual modules

- Module 3, 109 miRNAs: hsa-miR-486-5p, hsa-miR-451a, hsa-let-7b-5p, hsa-let-7i-5p, hsa-miR-22-3p ...



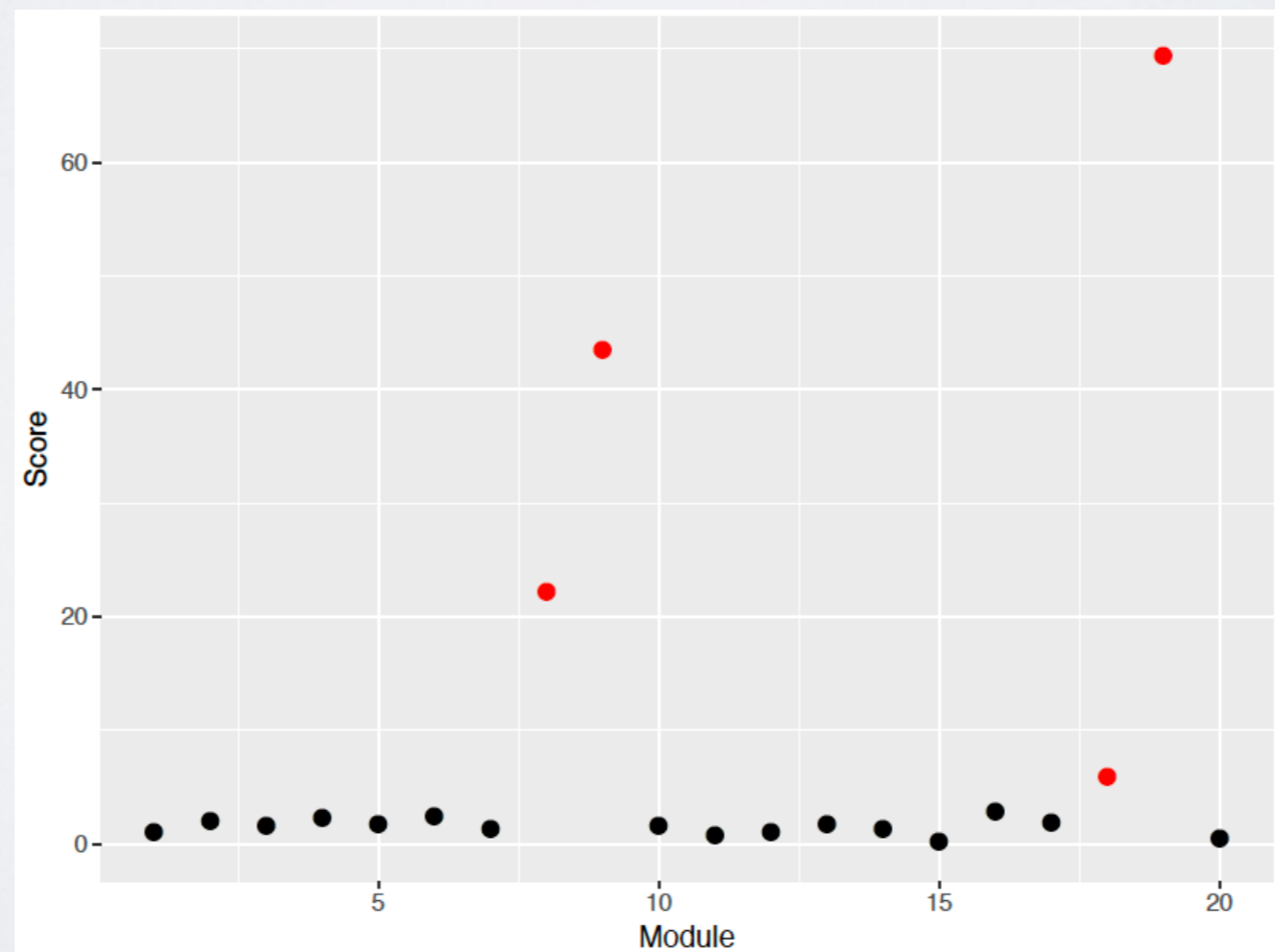
2d/3d tSNE for individual modules

- Module 7, 50 miRNAs: hsa-miR-1258, hsa-miR-1247-5p, hsa-miR-520g-5p, hsa-miR-4677-5p ...



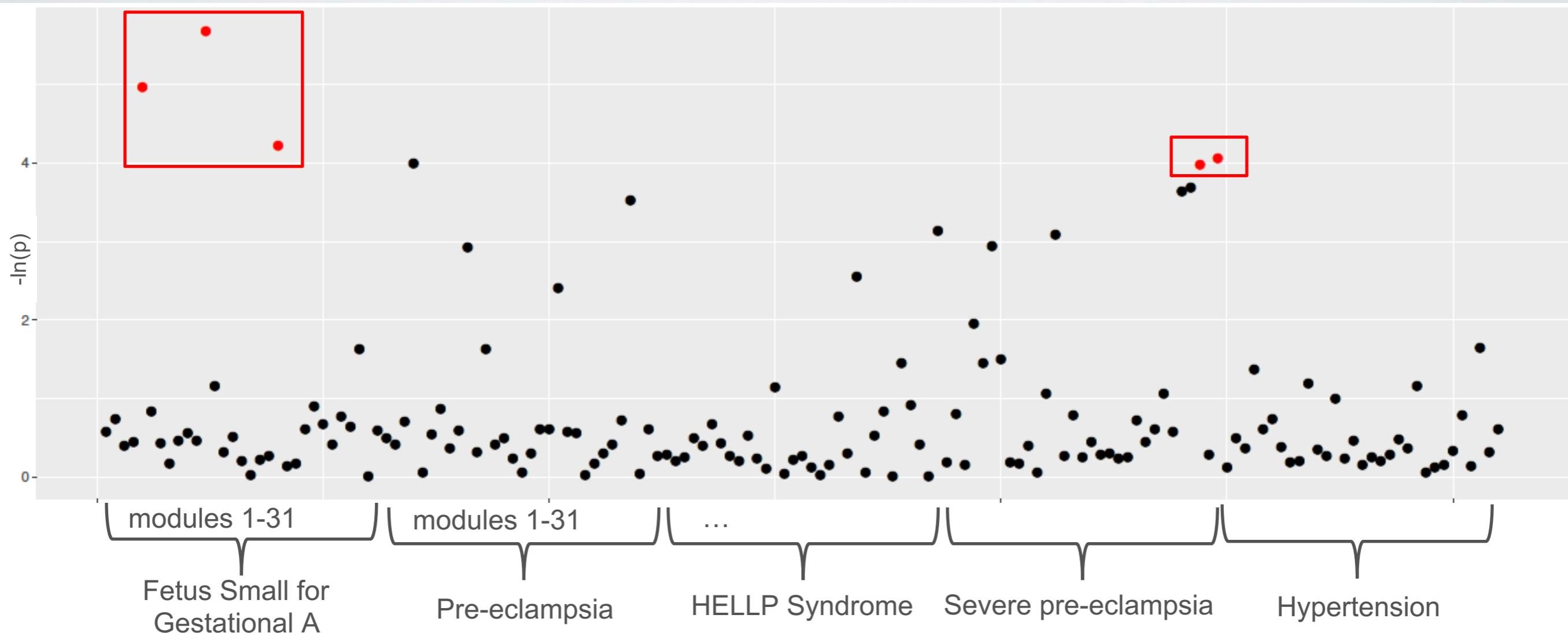
ANOVA test for module relevance to chosen factor (biofluid)

- Apply ANOVA to module eigengene values to test which modules are relevant to a chosen factor
- Example: Biofluid discrimination across atlas (N=1075 samples; 7 fluids: CSF, Plasma, Serum, Saliva, Urine, Cultured media, Bile; red = significant at FDR = 0.05)



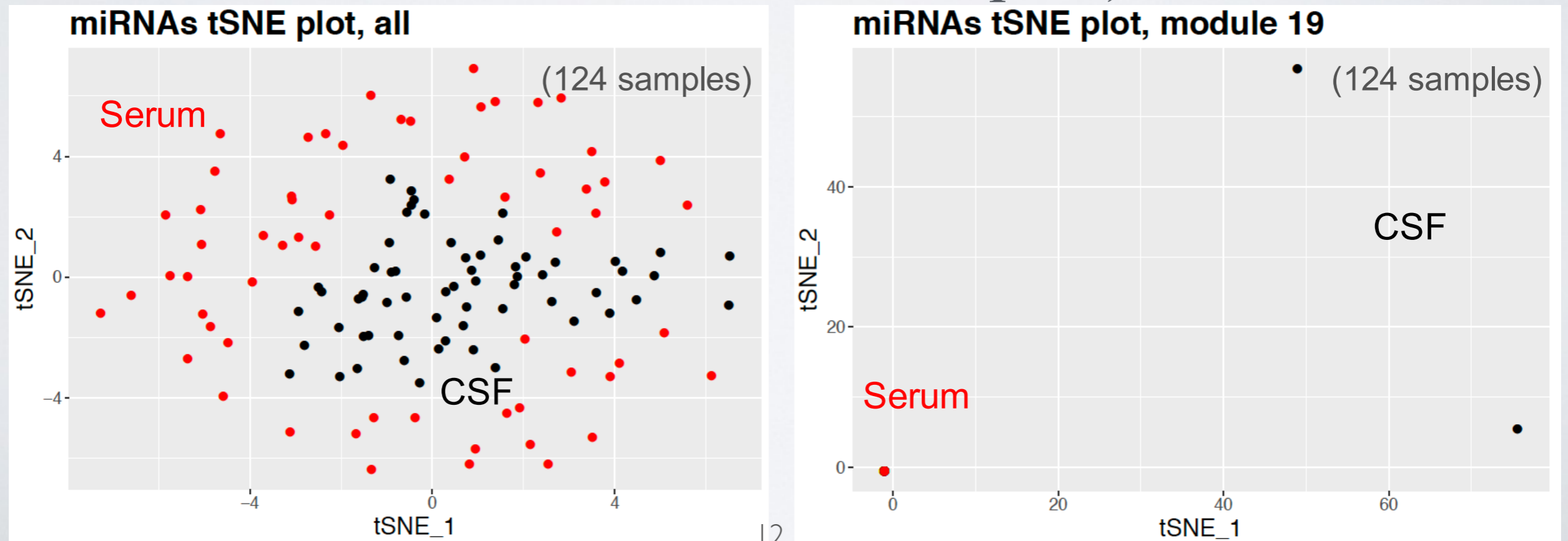
ANOVA test for module relevance to chosen factor (Placental dysfunction)

- Example: Placental dysfunction condition discrimination (N=48 samples (Louise Laurent); each condition tested against control; red = significant at FDR = 0.25)



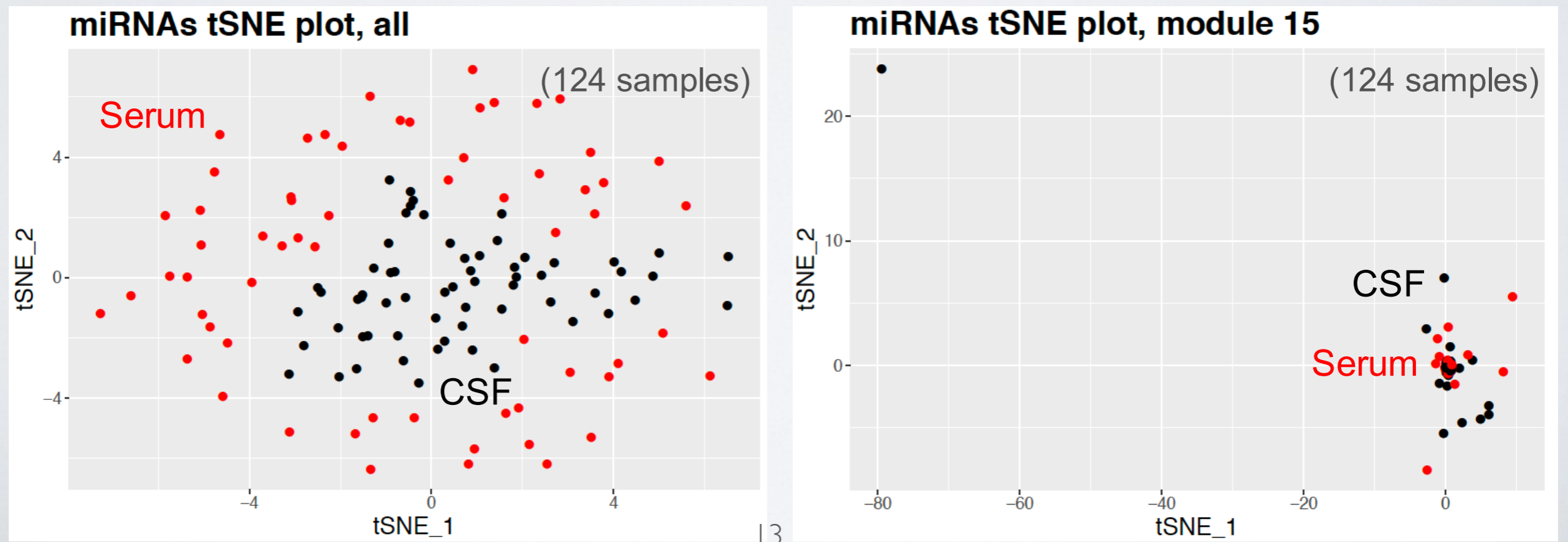
tSNE validation of discriminative modules for biofluids

- 124 samples (Kendall Van Keuren-Jensen) from CSF/Serum control samples
- Comparing sample-based tSNE using all miRNAs (left) and using only most discriminative module (right, 10 miRNAs: hsa-miR-4640-5p, hsa-mir-30b, hsa-mir-92a-1, hsa-miR-4644, hsa-miR-3942-3p ...)



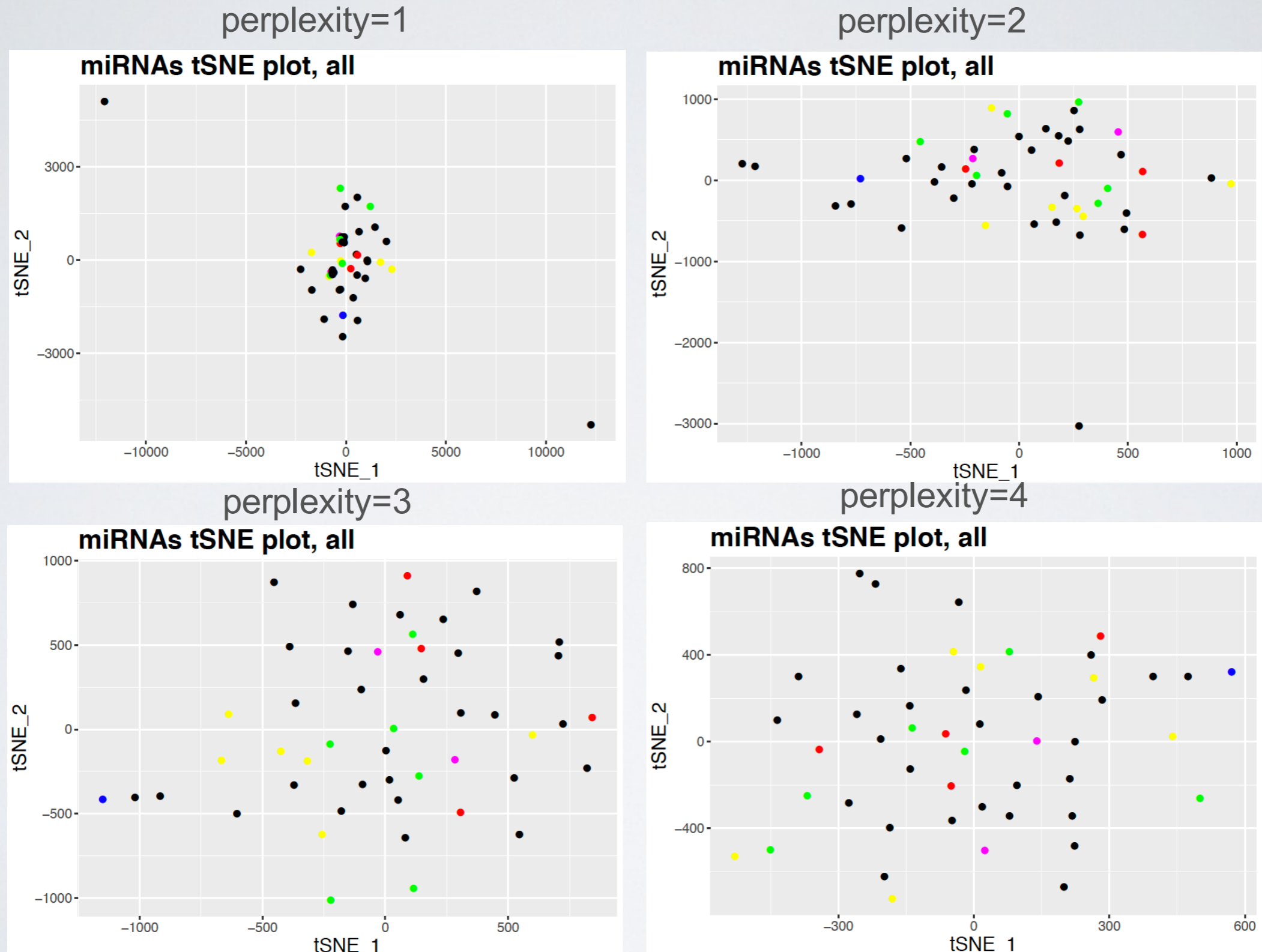
tSNE validation of discriminative modules for biofluids

- Comparing sample-based tSNE using all miRNAs (left) and using only **least** discriminative module (right, 20 miRNAs: hsa-miR-4289, hsa-miR-4515, hsa-miR-5004-3p, hsa-miR-6814-3p, hsa-miR-6759-5p ...)



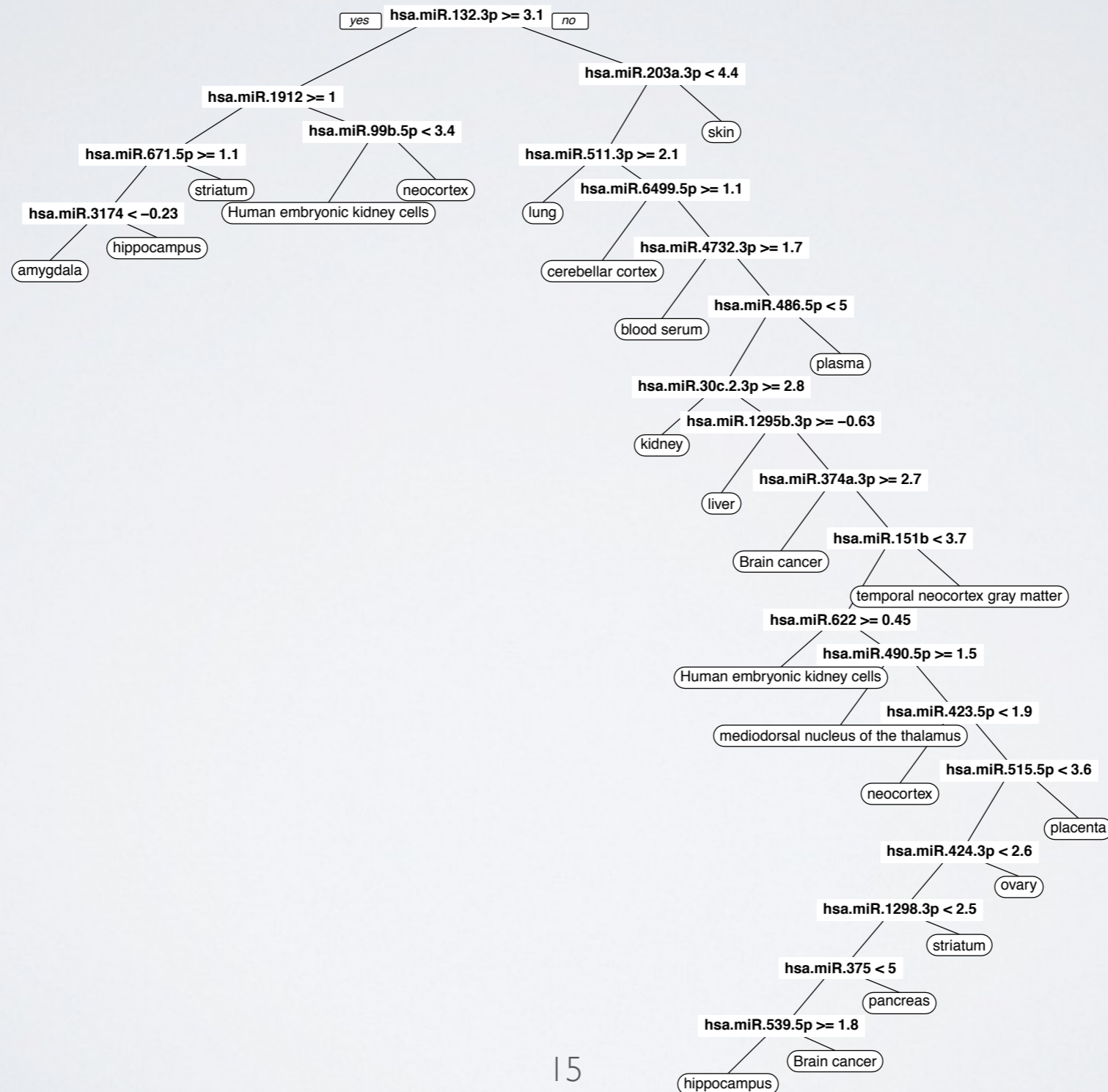
Varying the tSNE perplexity

- Placental dysfunction samples (black = control, 5 colors = conditions)



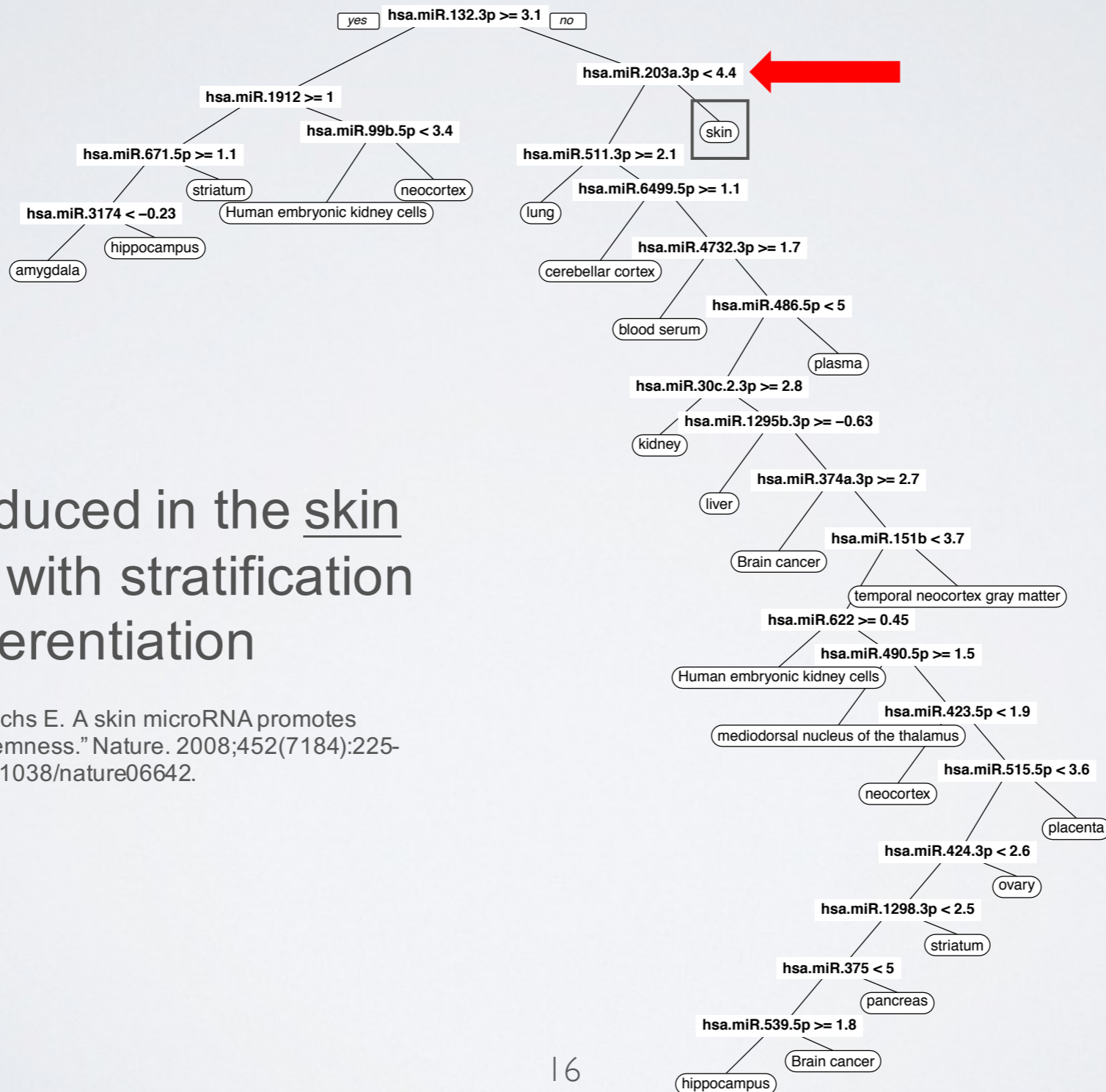
Decision trees to identify tissue/cell-types using cellular miRNAs

- Decision tree for cellular miRNA contents



Decision trees to identify tissue/cell-types using cellular miRNAs

- Decision tree for cellular miRNA contents



miR-203 is induced in the skin concomitantly with stratification and differentiation

Yi R, Poy MN, Stoffel M, Fuchs E. A skin microRNA promotes differentiation by repressing "stemness." Nature. 2008;452(7184):225-229. doi:10.1038/nature06642.

Decision trees to identify tissue/cell-types using cellular miRNAs

- Decision tree for cellular miRNA contents

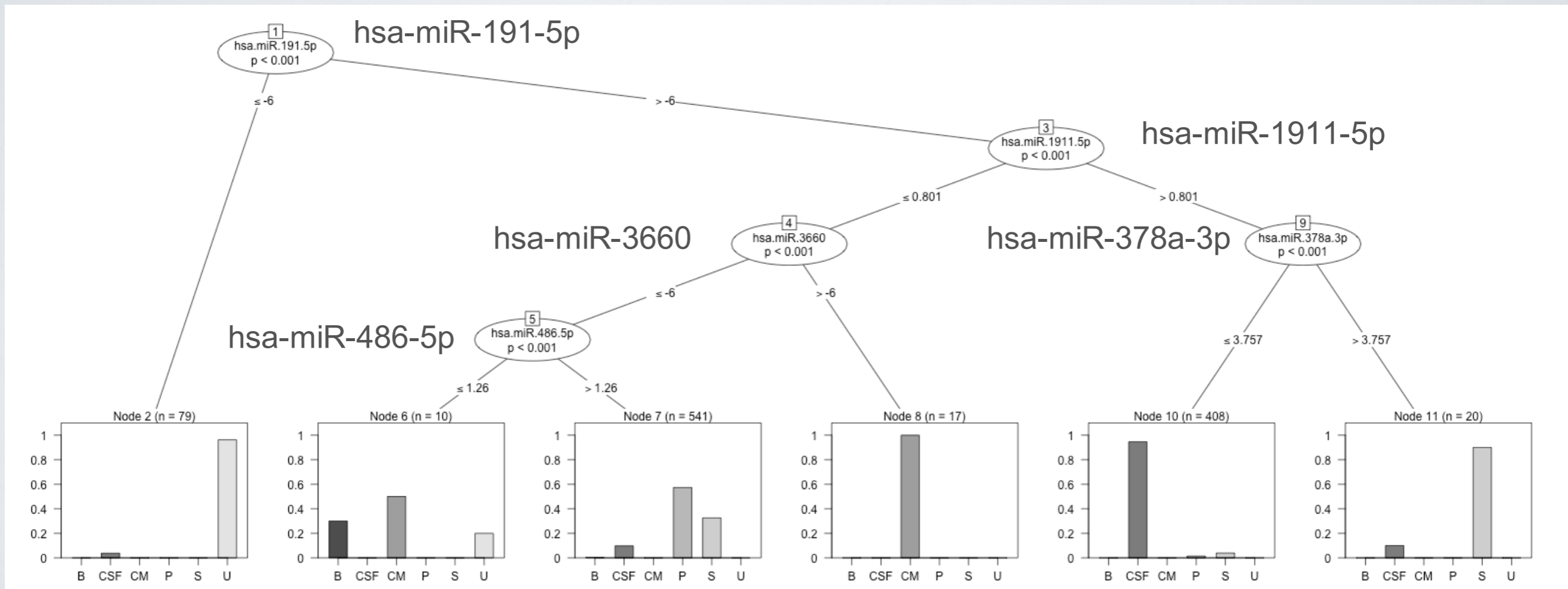


miR-515-5p is a highly specific placental microRNA and showed significant increase in maternal plasma concentrations throughout gestation

Katerina Kotlabova, Jindrich Doucha, Ilona Hromadnikova, Placental-specific microRNA in maternal circulation – identification of appropriate pregnancy-associated microRNAs with diagnostic potential, Journal of Reproductive Immunology, Volume 89, Issue 2, May 2011, Pages 185-191, ISSN 0165-0378, <http://dx.doi.org/10.1016/j.jri.2011.02.006>.

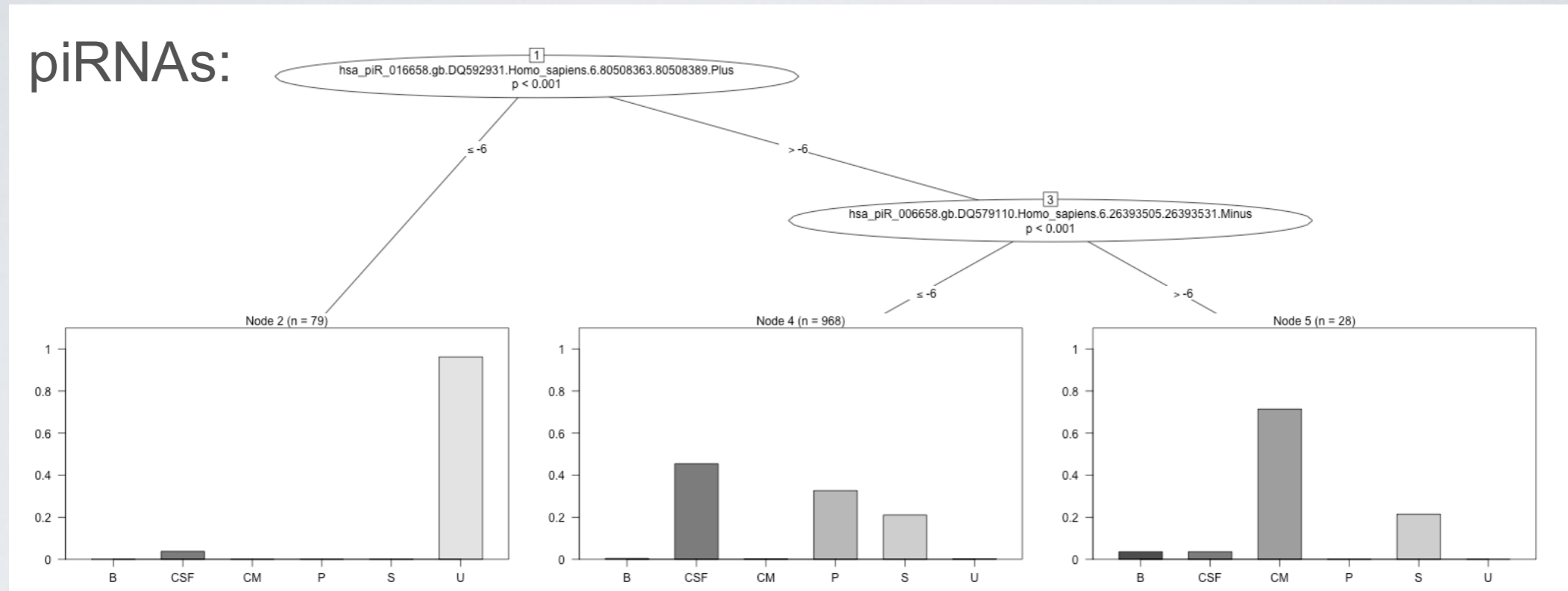
Biofluid decision trees: miRNAs

- Learn decision tree for discrimination of 6 biofluids using individual miRNAs

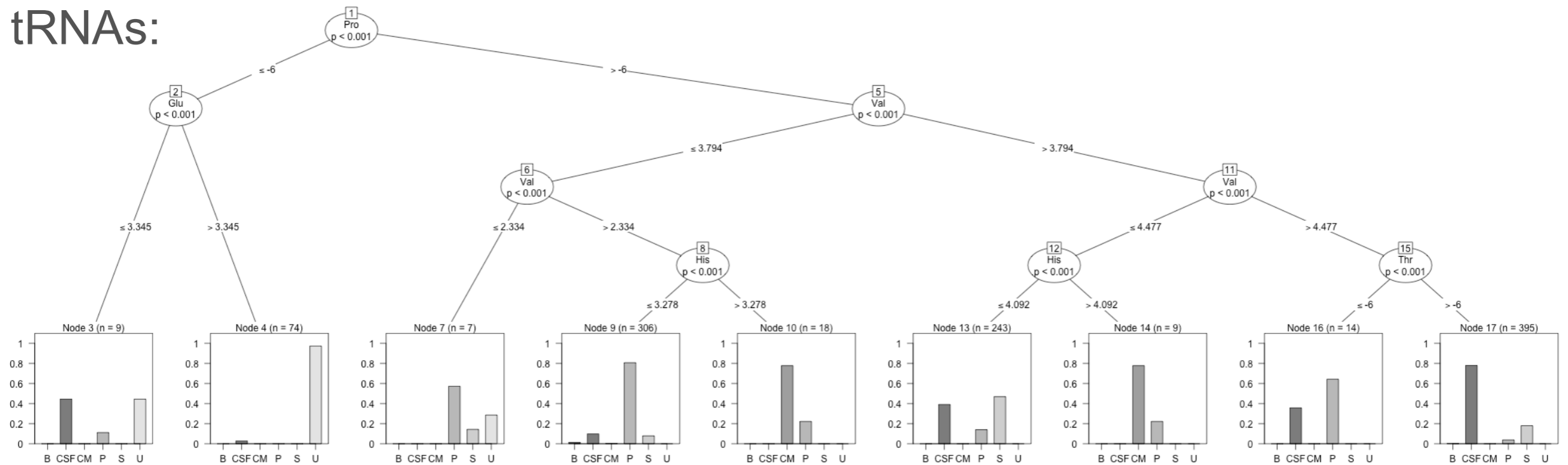


Biofluid decision trees: piRNAs/tRNAs

piRNAs:

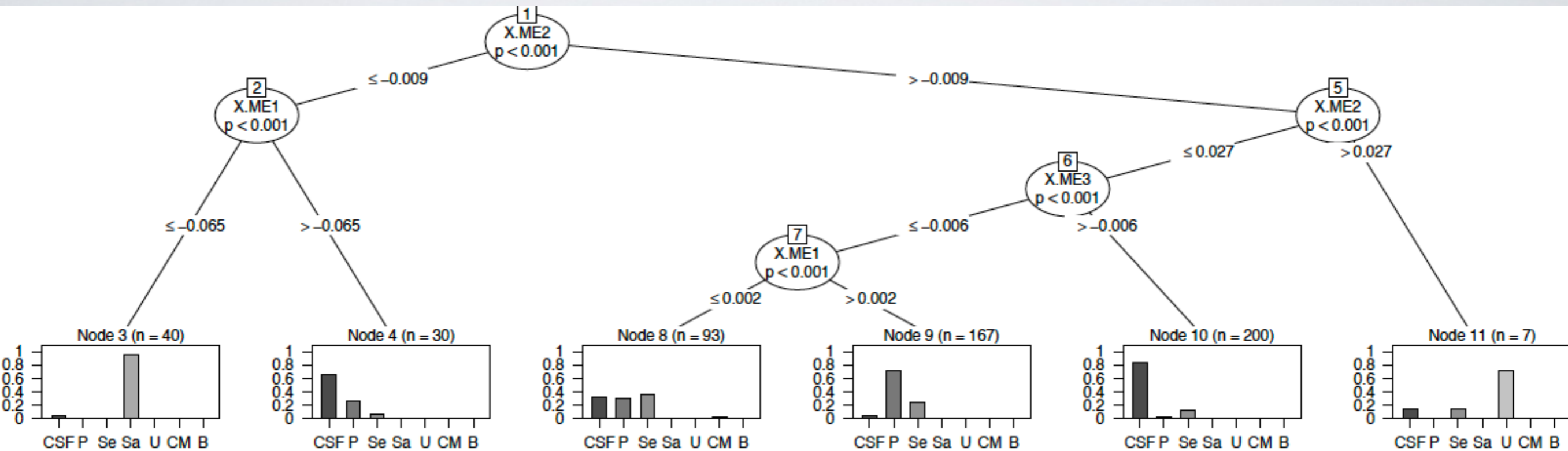


tRNAs:



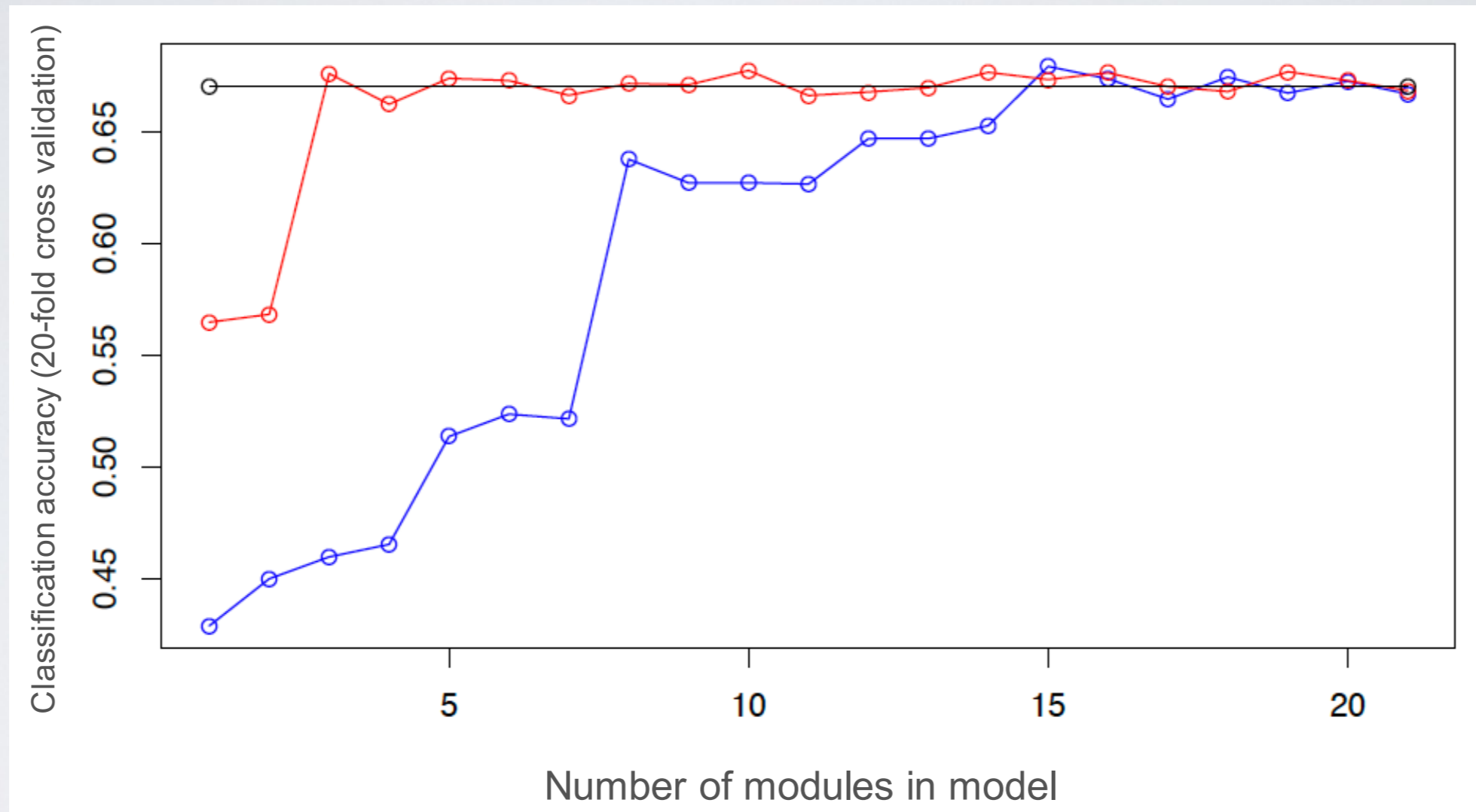
Biofluid decision trees: miRNA-modules

- Decision tree for biofluids learnt using module eigengene values



- 3 modules dominate (8, 5 and 7 miRNAs in each: hsa-miR-5584-3p, hsa-miR-6507-3p, hsa-miR-6717-5p ...)

Comparing miRNA and miRNA-module based decision trees



- Module-based, order by ANOVA score
- Module-based, random ordering
- individual miRNAs, all miRNAs used

Summary

- tSNE may be used to probe for structure in data using sample-based and miRNA-based views
- Modules may provide discriminative information as biomarkers of pathological conditions, and provide insight into exRNA regulation/function across biofluids
- Validation of discriminative power of modules/miRNAs via decision tree analysis: cell-type/tissue and biofluid classification

Acknowledgements

- James Diao
- Jonathan Park
- Joel Rozowsky
- Timur Galeev
- Rob Kitchen
- William Thistlethwaite
- Aleks Milosavljevic
- Mark Gerstein

Thanks for your attention!