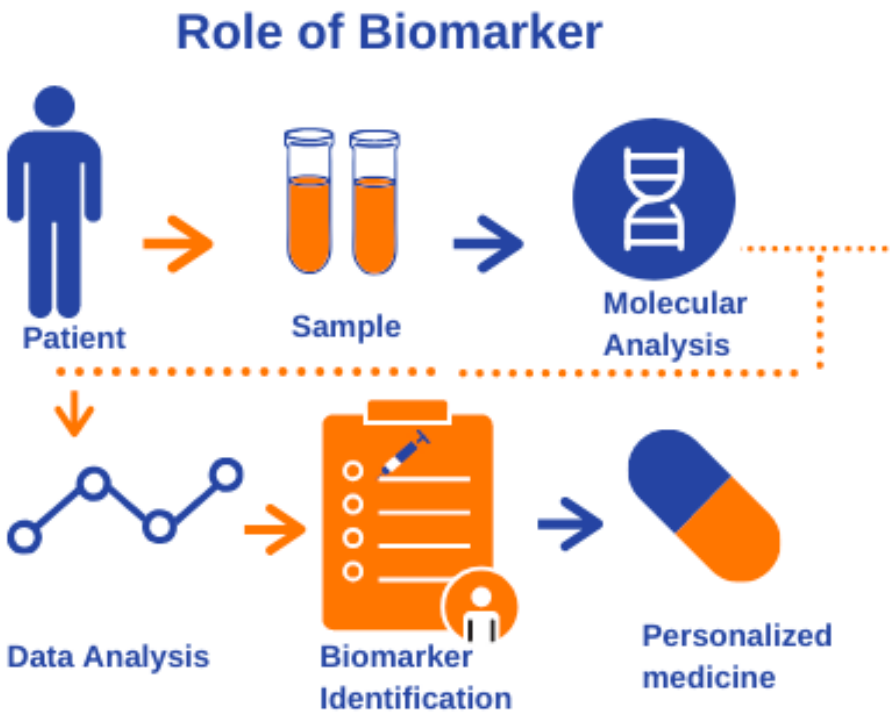
Application of Bayesian networks to multi-omics data for improving the diagnosis of asthma in preschool children

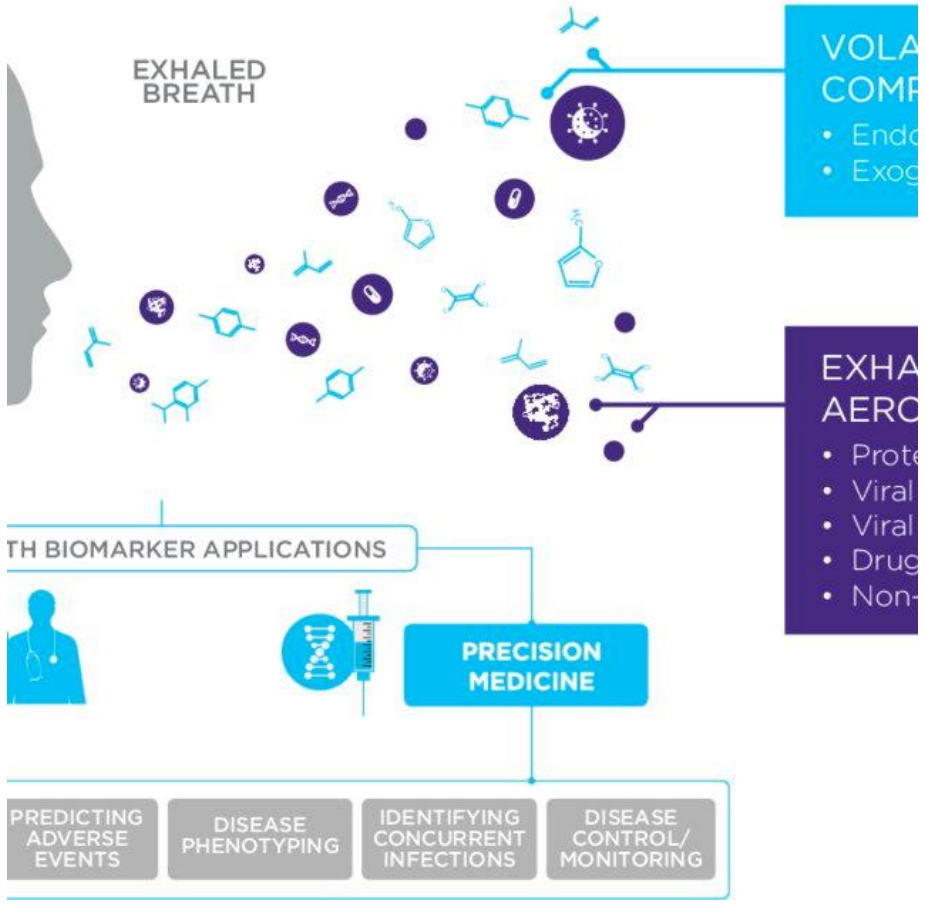
Vincent Dandenault

CHU Sainte-Justine
Le centre hospitalier universitaire mère-enfant

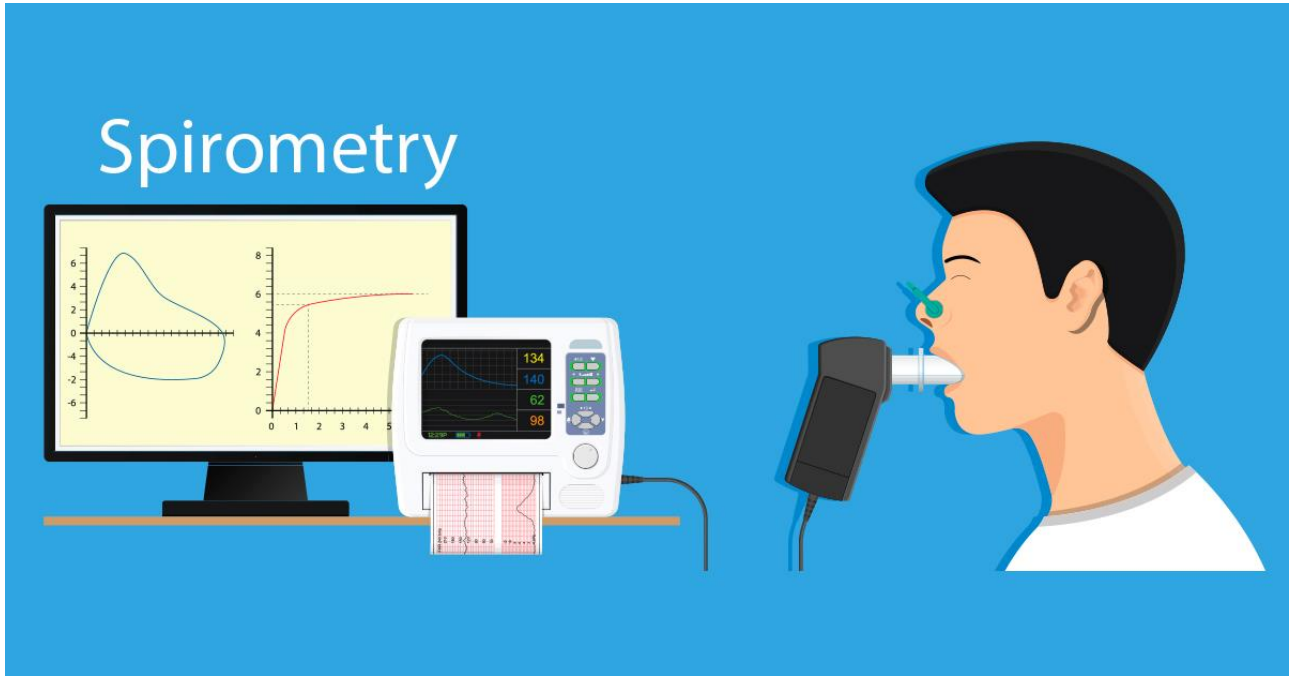
Université de Montréal

Biomarkers and Precision Medicine





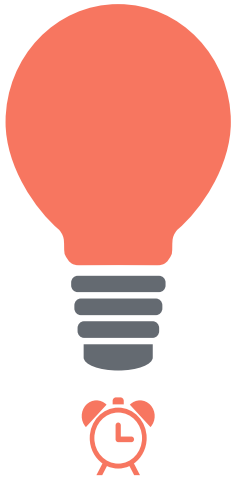
Problem: Asthma in Preschoolers



Objectif

Objectif #1: Feature selection for Asthma and Wheezing Classification

- Using integrated multi-omics and clinical data from preschoolers and asthma with wheeze and asthma
- Building a Naive Bayes Classifier for classification of labels
- Feature extraction for our results to motive the next objective



Objectif #2: Identify molecular endotypes from probabilistic graphical models structures

OMICS Data

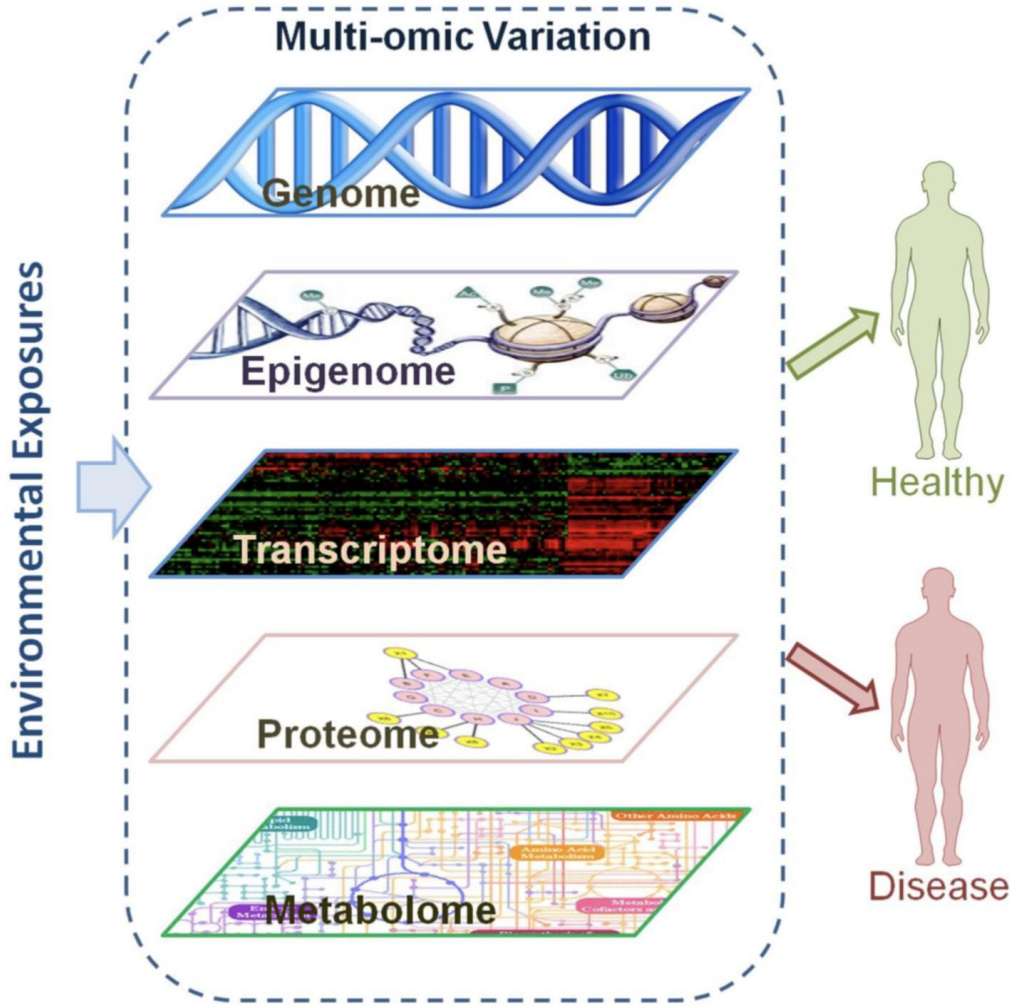
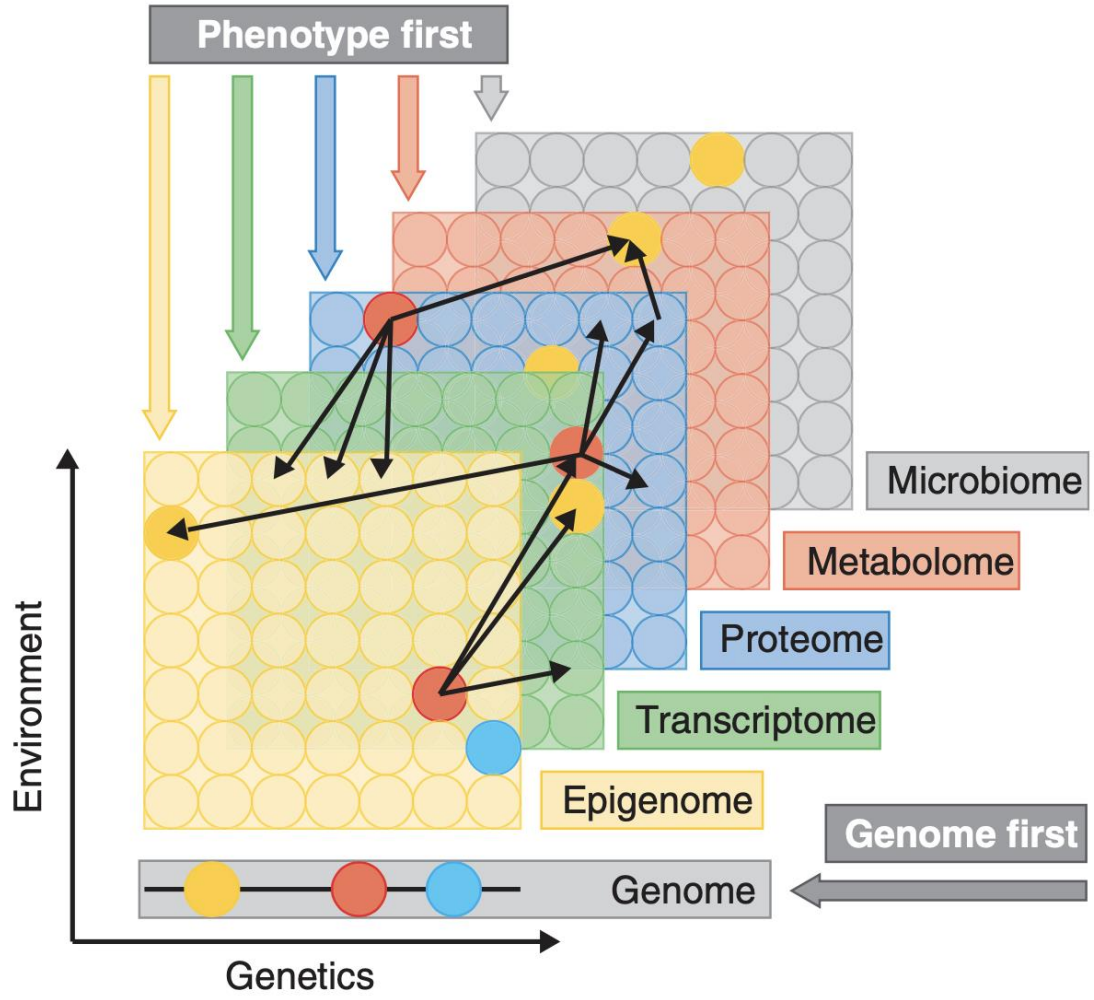
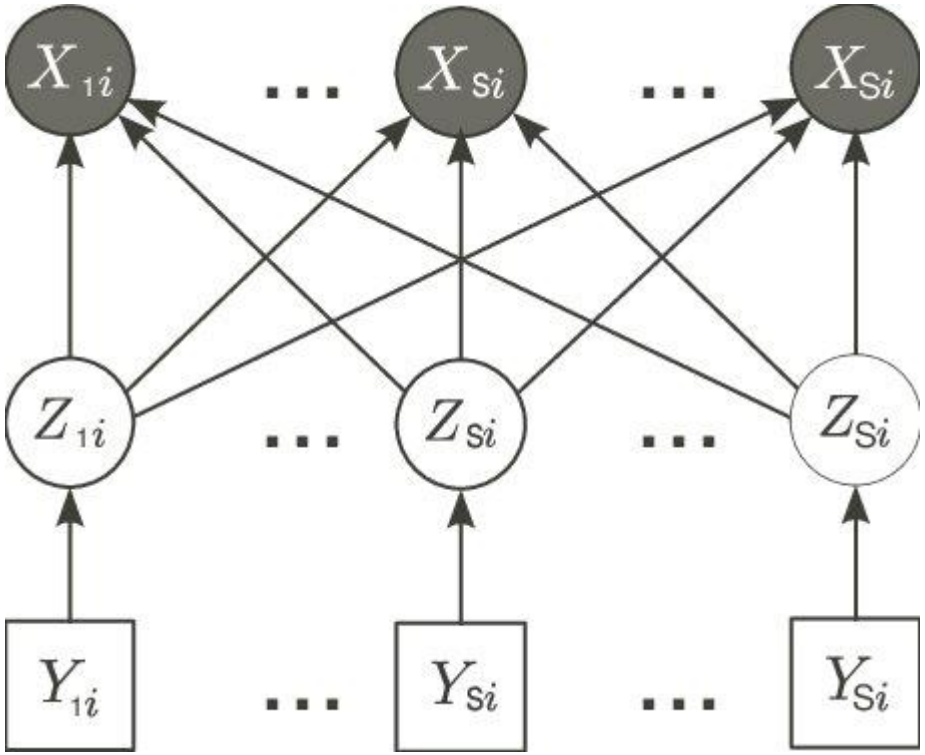
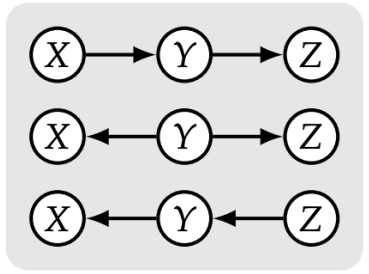


Figure 1. Conceptual model of multi-omics and human disease

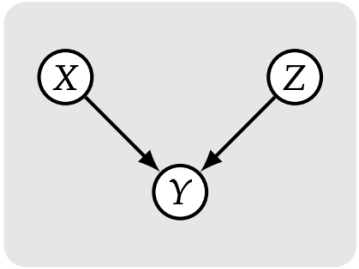


Probabilistic Graphical Models





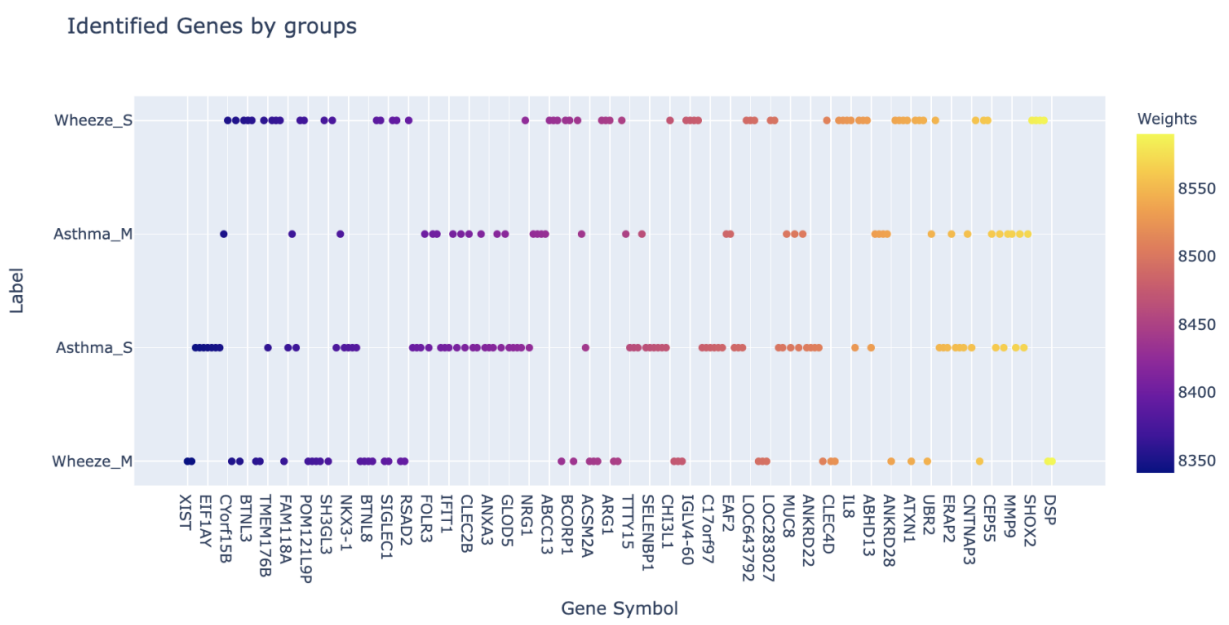
$X \not\perp\!\!\!\perp Z$ and $X \perp\!\!\!\perp Z \mid Y$

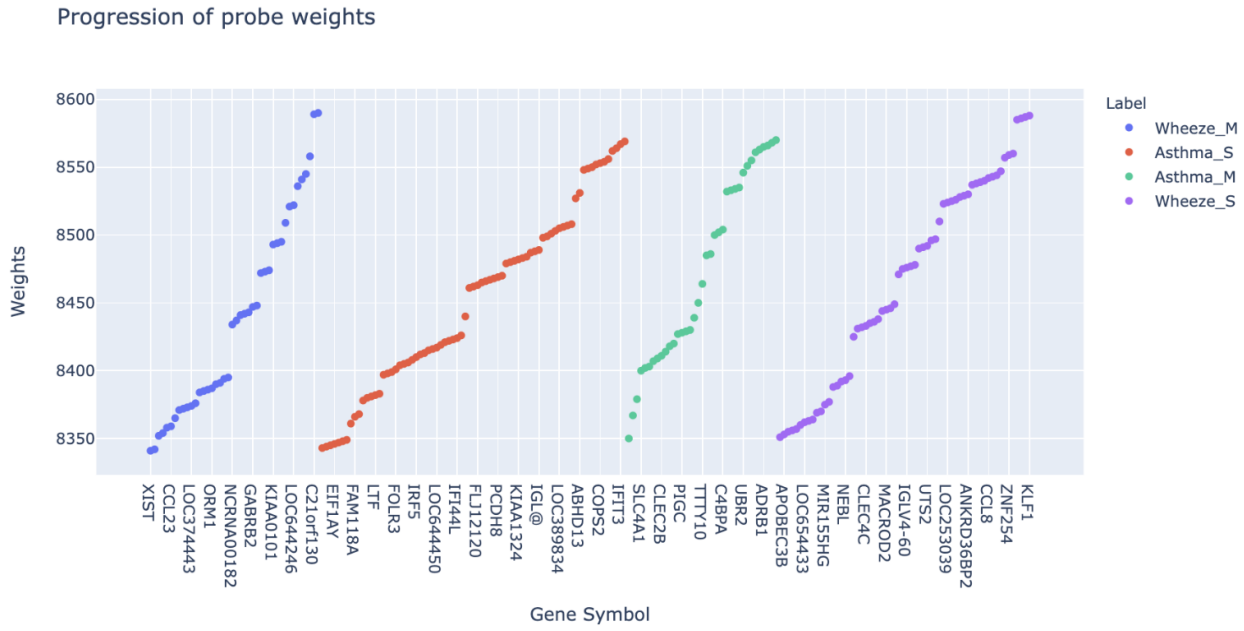


$X \perp\!\!\!\perp Z$ and $X \not\perp\!\!\!\perp Z \mid Y$

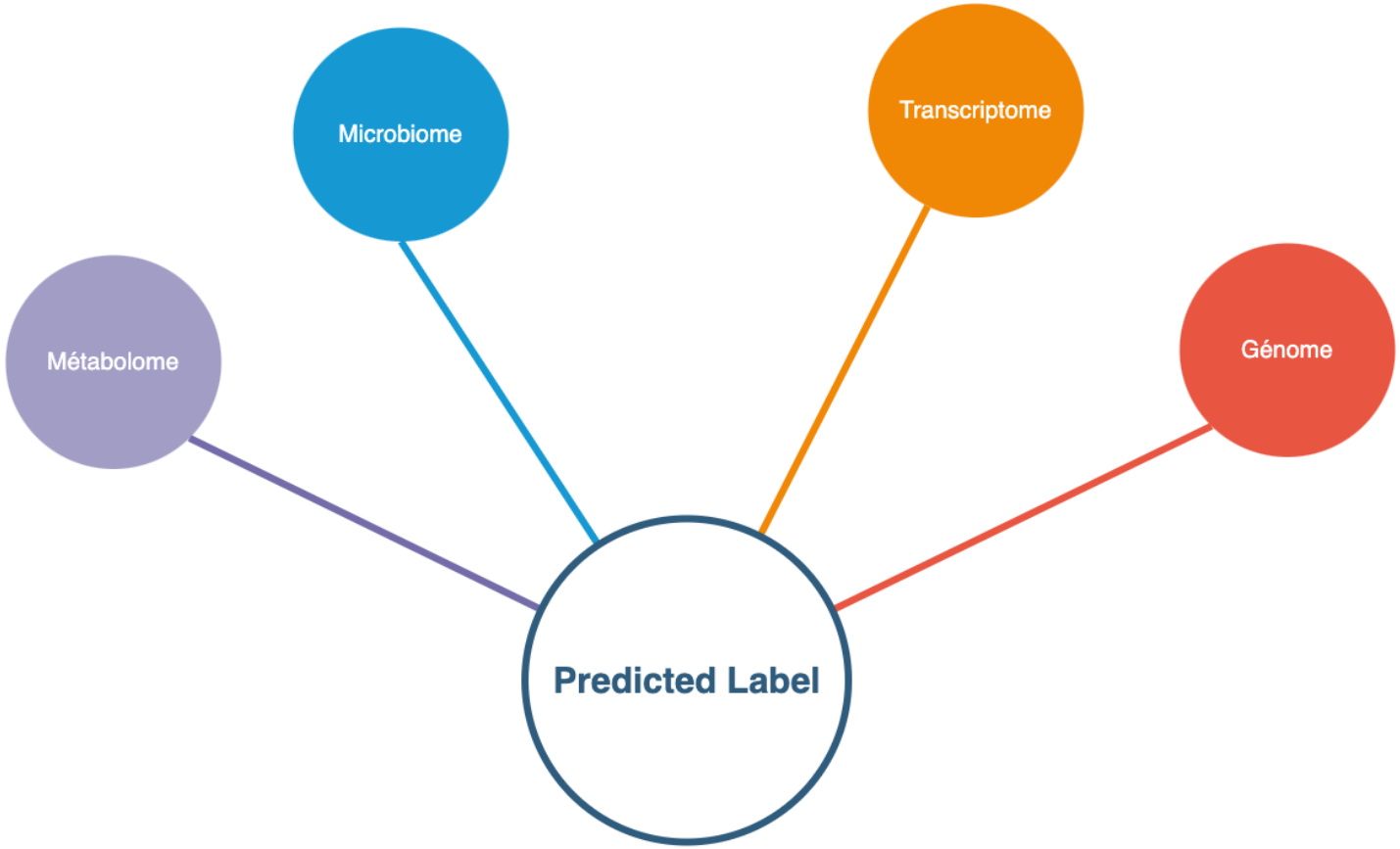
TensorFlow

Example - Transcriptomics





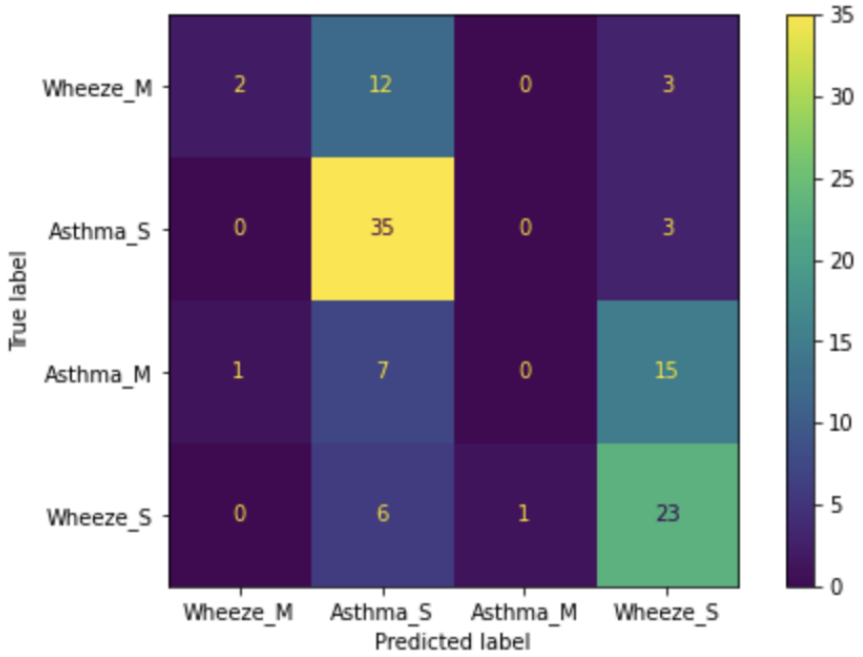
Naive Bayes Classifier



{Mild Asthma, Severe Asthma, Mild Wheeze, Severe Wheeze}

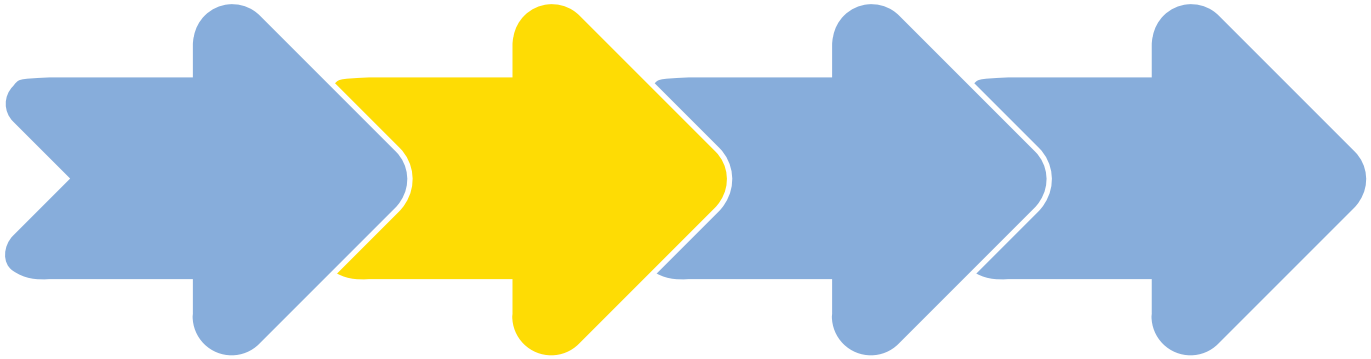
Preliminary Results

```
Number of mislabeled points out of a total 108 points : 48
```



	precision	recall	f1-score	support
Asthma_M	0.50	0.12	0.19	17
Asthma_S	0.56	0.87	0.68	38
Wheeze_M	0.00	0.00	0.00	23
Wheeze_S	0.47	0.70	0.56	30
accuracy			0.52	108
macro avg	0.38	0.42	0.36	108
weighted avg	0.41	0.52	0.42	108

Next Steps



01

Step 1: Data Cleaning

Utilized the maximum variability of probing data to clean the identify a short list of potential features (Genes, Functional Groups, etc.)

02

Step 2: Baseline NBN and VAE model

Implementation of a Naive Bayesian Network as a baseline model to understand feature selection, then using a variational approach to find approximate real underlying distributions

03

Step 3: Structure Learning & Clinical Priors

his is an NP-hard problem, since the number of dag's on N variables is super-exponential in N. Hopefully Clinical insights can allow shortcuts in the finding the ideal hidden nodes.

04

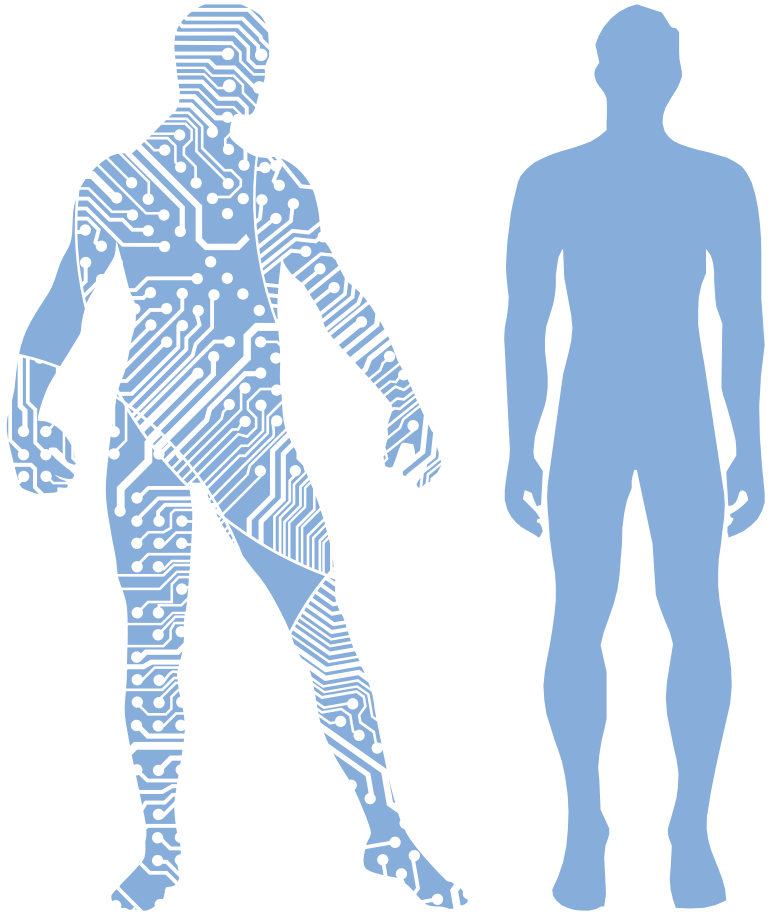
Step 4: Network Analysis and Causal Discovery

Use probabilistic graphical models to uncover causal mechanisms and pathways in complex diseases. Node and edge *centrality* measures in the graph will be used to identify key biomarker

Conclusion

Perspects

- Utilise a hierarchical to identify a better underlying structure of clinical data
- Use variational methods to estimate underlying probability density
- From this, Identify molecular endotypes for to lead further studies in preschool asthma



Conclusion

- Integrated OMICs data for specific prior densities in a Bayesian Network
- Classification tasks as a baseline task
- Acceptable results, need for better probability densities

Thank you!

Any Questions?

A special thank you so both of my supervisors, Dr Cristina Longo and Dr Simon de Montigny from CHU Sainte-Justine, Montréal, Qc.

References

[1] Garrett-Bakelman, F. E., Darshi, M., Green, S. J., Gur, R. C., Lin, L., Macias, B. R., ... & Turek, F. W. (2019). The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight. Science, 364(6436), eaau8650

[2] Tian, D., Gu, Q., & Ma, J. (2016). Identifying gene regulatory network rewiring using latent differential graphical models. Nucleic acids research, 44(17), e140-e140

[3] Conrad, L. A., Cabana, M. D., & Rastogi, D. (2021). Defining pediatric asthma: phenotypes to endotypes and beyond. Pediatric research, 90(1), 45–51. https://doi.org/10.1038/s41390-020-01231-6

[4] Haworth, J. J., Pitcher, C. K., Ferrandino, G., Hobson, A. R., Pappan, K. L., & Lawson, J. L. (2022). Breathing new life into clinical testing and diagnostics: perspectives on volatile biomarkers from breath. Critical Reviews in Clinical Laboratory Sciences, 1-20.

[5] Koller, Daphne, and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.

[6] Ashley, Euan A. "Towards precision medicine." Nature Reviews Genetics 17.9 (2016): 507-522.

[7] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[8] Sun, Y. V., & Hu, Y. J. (2016). Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. Advances in genetics, 93, 147-190.

[9] Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. Genome biology, 18(1), 1-15.

[10] Logotheti, M., Agioutantis, P., Katsaounou, P., & Loutrari, H. (2021). Microbiome Research and Multi-Omics Integration for Personalized Medicine in Asthma. Journal of Personalized Medicine, 11(12), 1299.

[11] Wheelock, C. E., Goss, V. M., Balgoma, D., Nicholas, B., Brandsma, J., Skipp, P. J., ... & U-BIOPRED Study Group. (2013). Application of'omics technologies to biomarker discovery in inflammatory lung diseases. European Respiratory Journal, 42(3), 802-825.

[12] Oukhellou, Latifa, et al. "Semi-supervised feature extraction using independent factor analysis." 2011 10th International Conference on Machine Learning and Applications and Workshops. Vol. 2. IEEE, 2011.