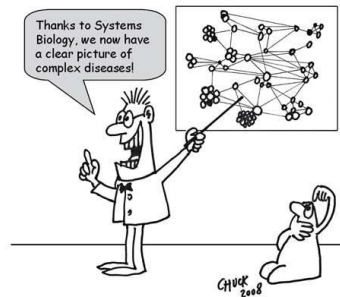


SIMON says: Machine learning for everyone!



Adriana Tomic

Systems Immunology | Oxford Vaccine Group

SIMON training course @Big Data Institute

Part II - May 12th, 2021



@TomicAdriana



adriana.tomic@paediatrics.ox.ac.uk

1

Training course - overview

Part I – SIMON, pattern recognition and knowledge extraction platform (May 5th 2021)

- Machine learning and AI – what is all the fuss about?
- What is SIMON?
- Case study – example 1 (dealing with missing values, overfitting, model performance)

Theoretical part – 15min

- Perform SIMON analysis using provided dataset
- Performance metrics, evaluation and selection of high-quality models

Hands-on – 30min

..... Questions? 15min

Part II – Exploratory analysis (May 12th 2021)

- Feature selection: scoring and elimination
- Correlation and clustering analysis

Hands-on – 30min

- Feature processing methods to avoid 'curse of dimensionality'
- Case studies – examples

Theoretical part – 15min

..... Questions? 15min

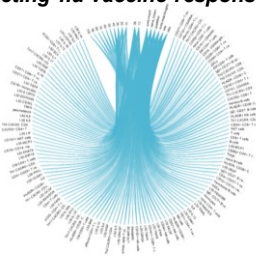
2

Part II. SIMON – usage examples



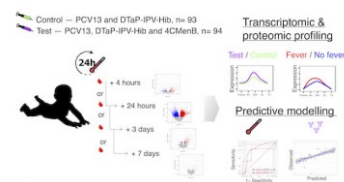
3

Integrative analysis of different data types – predicting flu vaccine responses



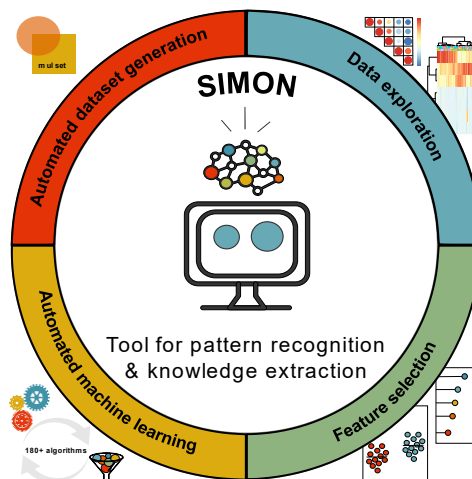
Tomic et al, JI, 2019

Transcriptome data

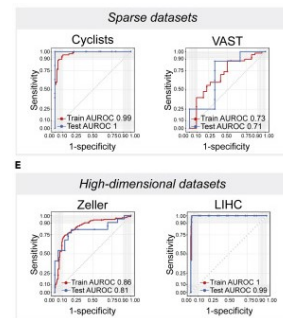


O'Connor et al, Mol Syst Biol, 2020

Datasets with high sparsity or high-dimensionality (transcriptome, microbiome)



Tomic et al, Patterns, 2021



Multi-omics integrative analysis – COVID-19 COMBAT project



COMBAT project, submitted

4

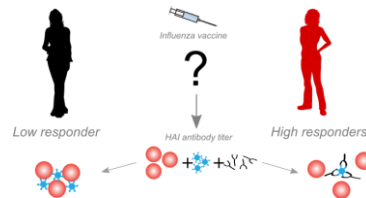
SIMON applications - examples

1. Identification of immunological parameters that can predict responses to flu vaccination

Tomic et al, JI, 2019



<https://fluprint.com/>

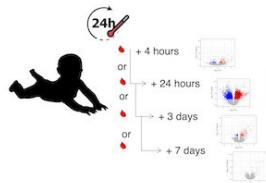


Great example when you are dealing with data collected across studies with many missing values

2. Identification of molecular signature that can predict reactivity in infants to meningococcal vaccine

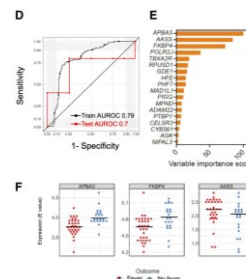
O'Connor et al, Mol Syst Biol, 2020

Control — PCV13 and DTaP-IPV-Hib, n= 93
Test — PCV13, DTaP-IPV-Hib and 4CMenB, n= 94



'curse of dimensionality'

Need to reduce number of features:
-DEG
-PCA
-clustering
-etc.



Great example when you are dealing with transcriptome/proteome data collected on a smaller sample size

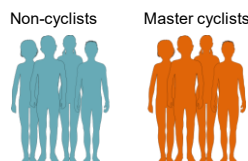
Blood RNAseq (54 infants, 14,000+ transcripts)

5

SIMON applications - examples

3. Identification of immunological parameters that can distinguish between two groups

Tomic et al, Patterns, 2021



Duggal et al, Agging Cell, 2018

Great example how to use SIMON for data mining instead of multiple testing

4. Building models to identify cancer cells (LIHC dataset) or to identify cancer patients based on the microbiome (Zeller dataset)

Tomic et al, Patterns, 2021

Great example how to use SIMON for building diagnostic, predictive models

5. Using SIMON for multi-omics integrative analysis ('COMBAT: COVID-19 Multi-Omics Blood Atlas')

Manuscript under review



6



SIMON says: Thank you!

Don't forget to provide you feedback on SIMON training course:

<https://b4vdw1gm99z.typeform.com/to/ADgBmhU1>