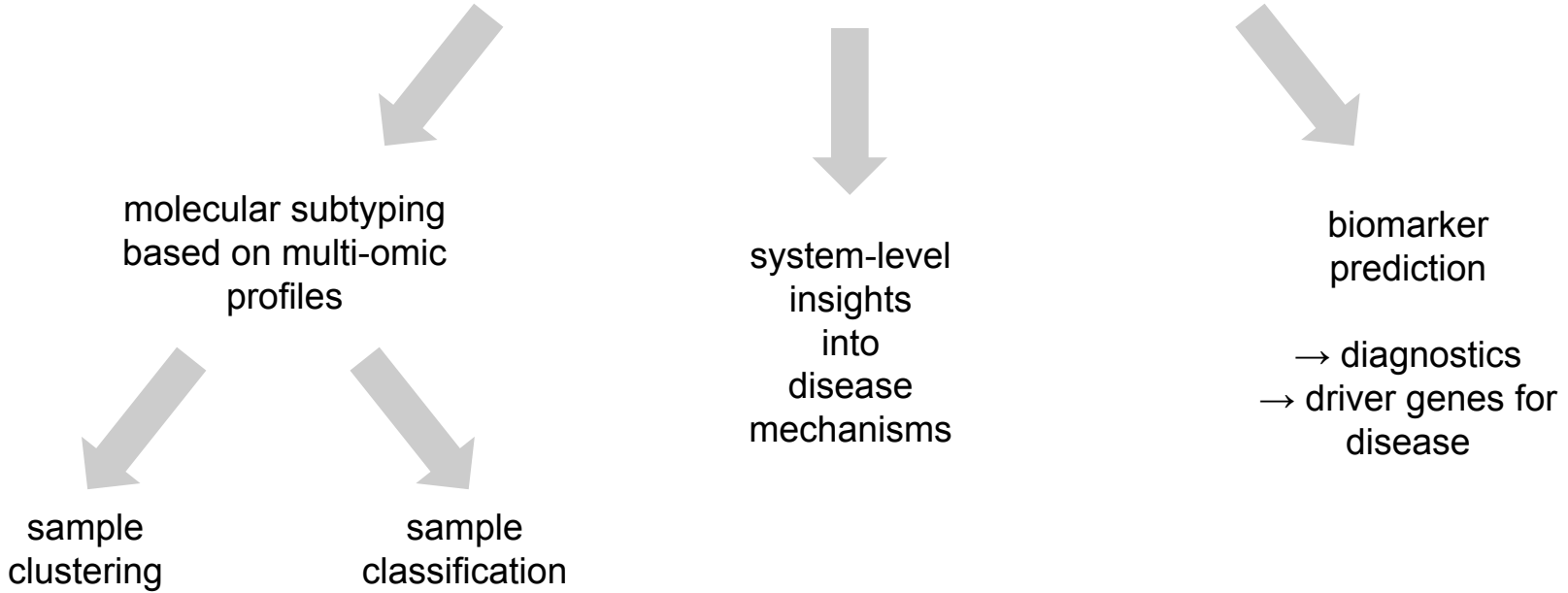
Multi-omics data integration hackathon

September 8 - 15, 2021

Karolina Sienkiewicz
Weill Cornell Graduate School of Medical Sciences,
Cornell University

Why are we integrating multi-omic datasets?



Integration challenges

Different number of samples

Data heterogeneity

Missing values in feature matrices

Normalization across data types following different distributions

Sparse datasets

Feature selection

Methods for multi-omic data integration

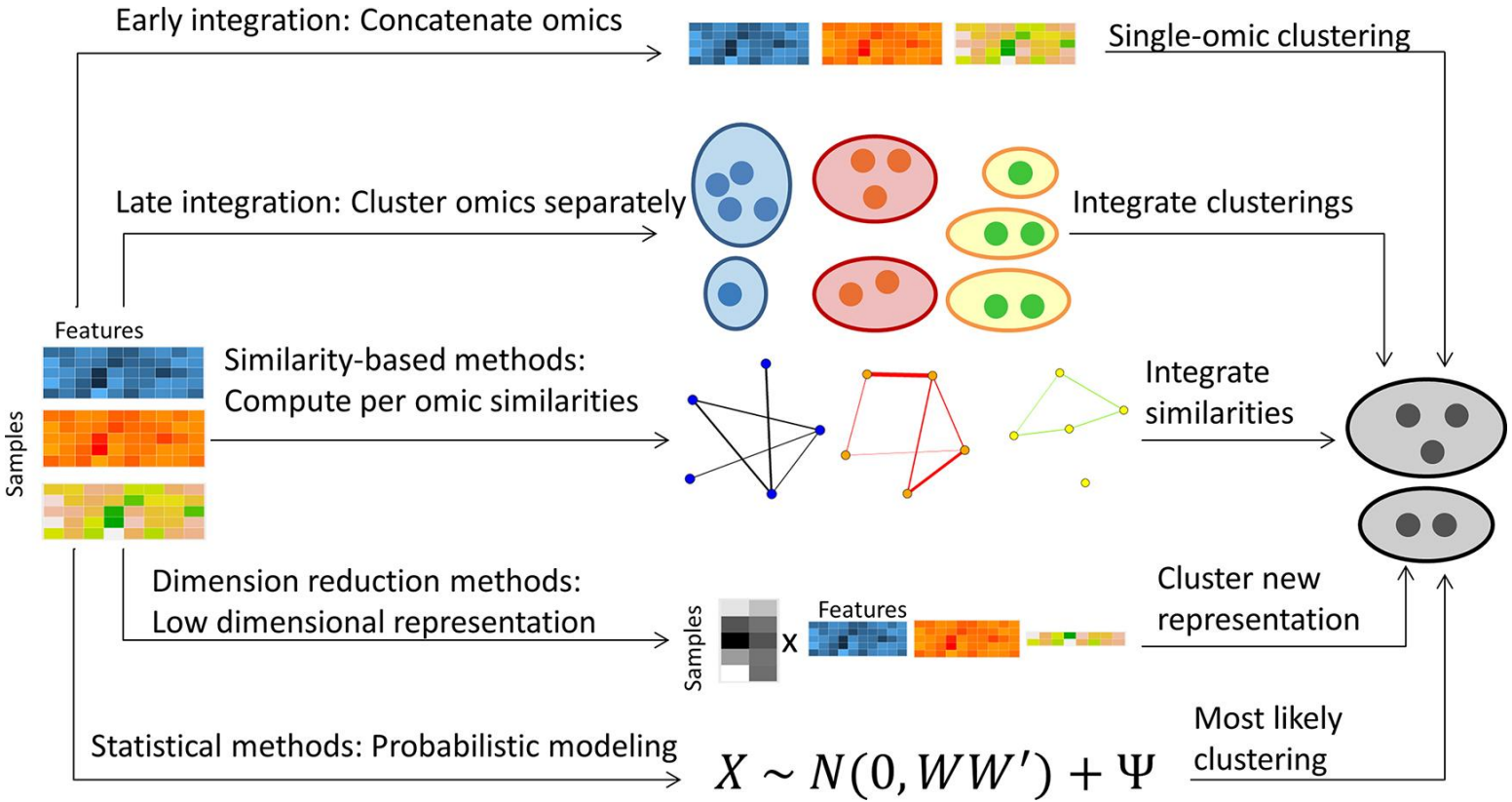


Figure source: Rappoport, N., & Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Research, 46(20), 10546–10562. https://doi.org/10.1093/nar/gky889

Methods for multi-omic data integration

Multiple Factor Analysis (MFA)

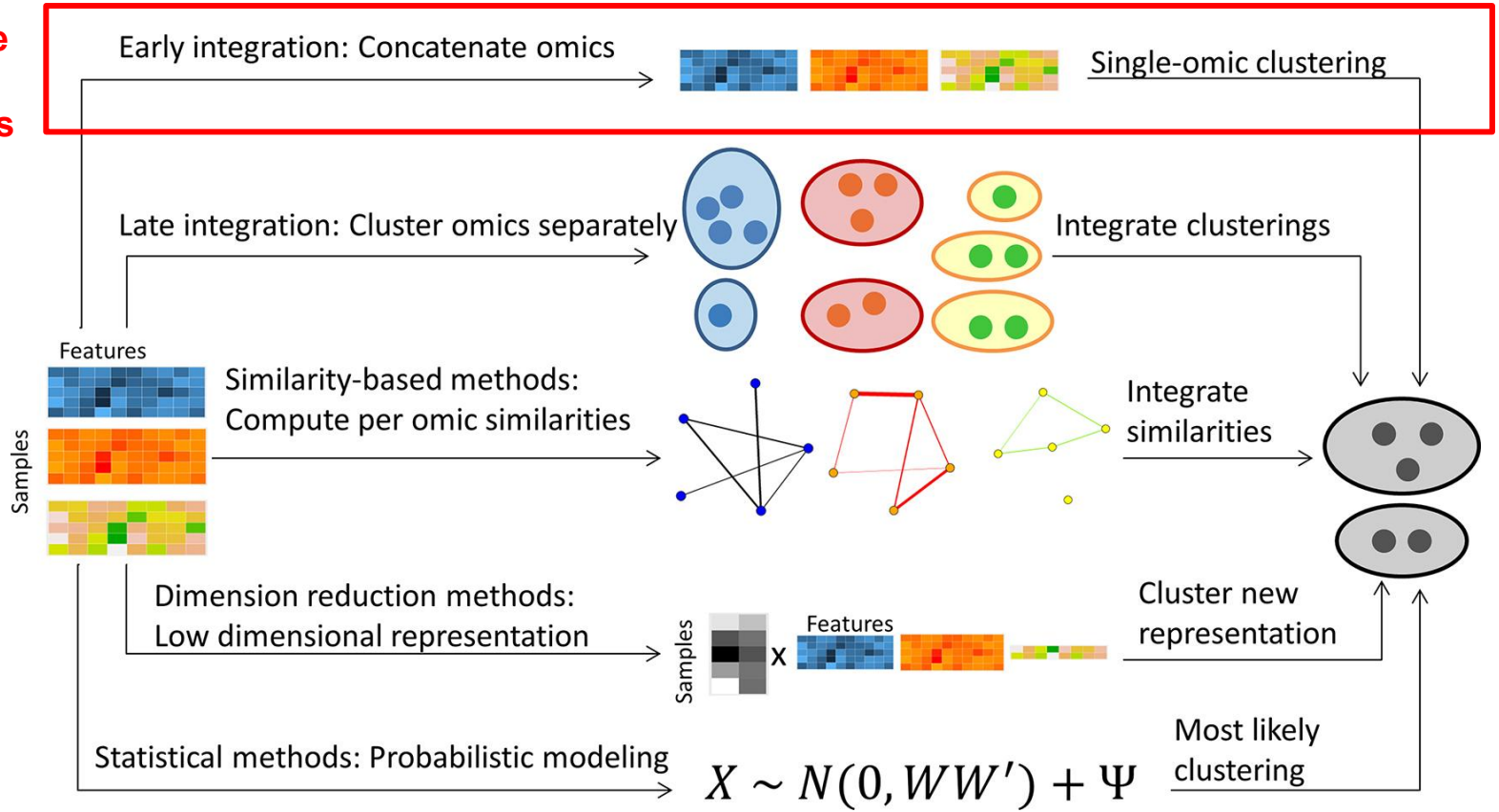


Figure source: Rappoport, N., & Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Research, 46(20), 10546–10562. https://doi.org/10.1093/nar/gky889

Methods for multi-omic data integration

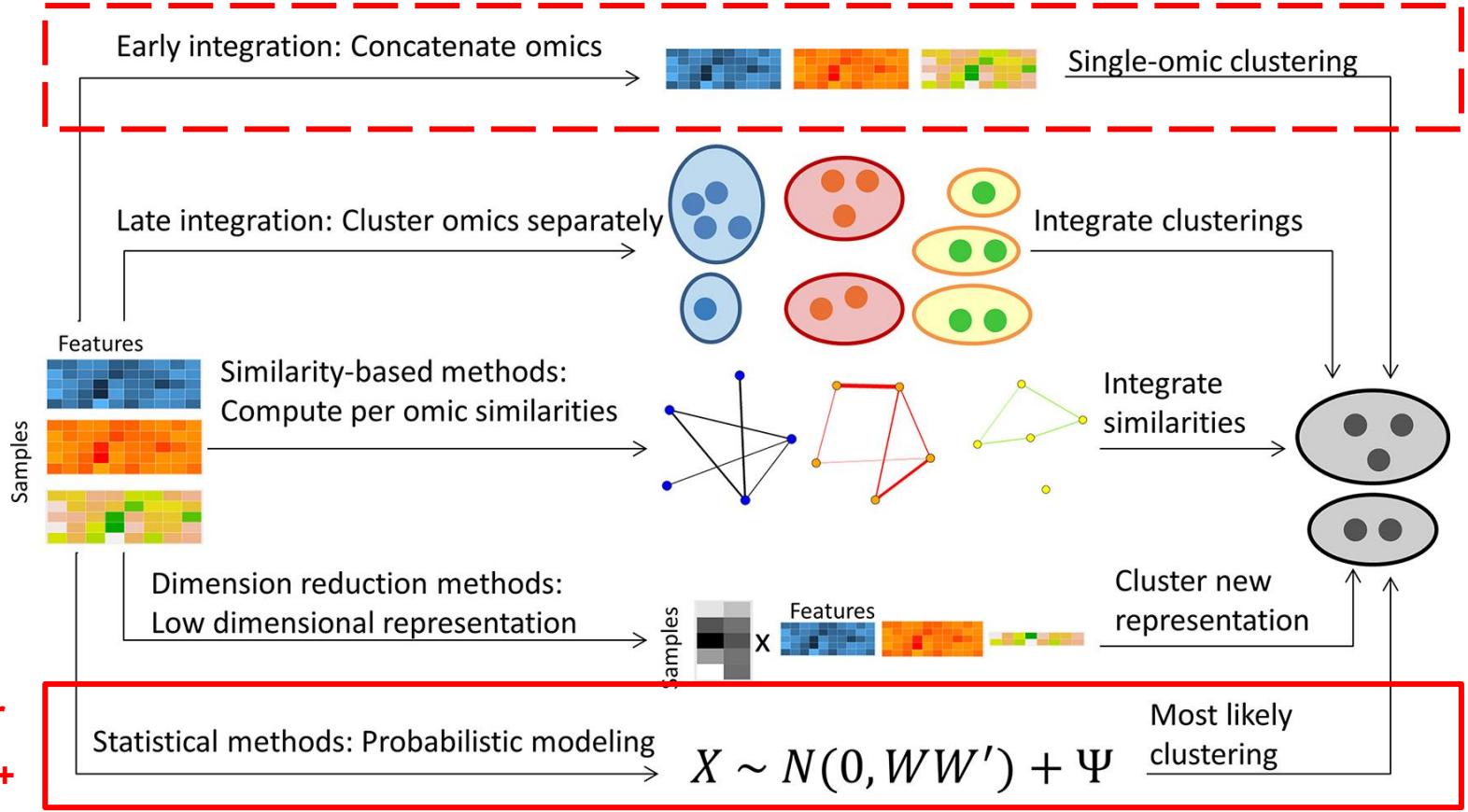


Figure source: Rappoport, N., & Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Research, 46(20), 10546–10562. https://doi.org/10.1093/nar/gky889

Methods for multi-omic data integration

PINS

PINS+

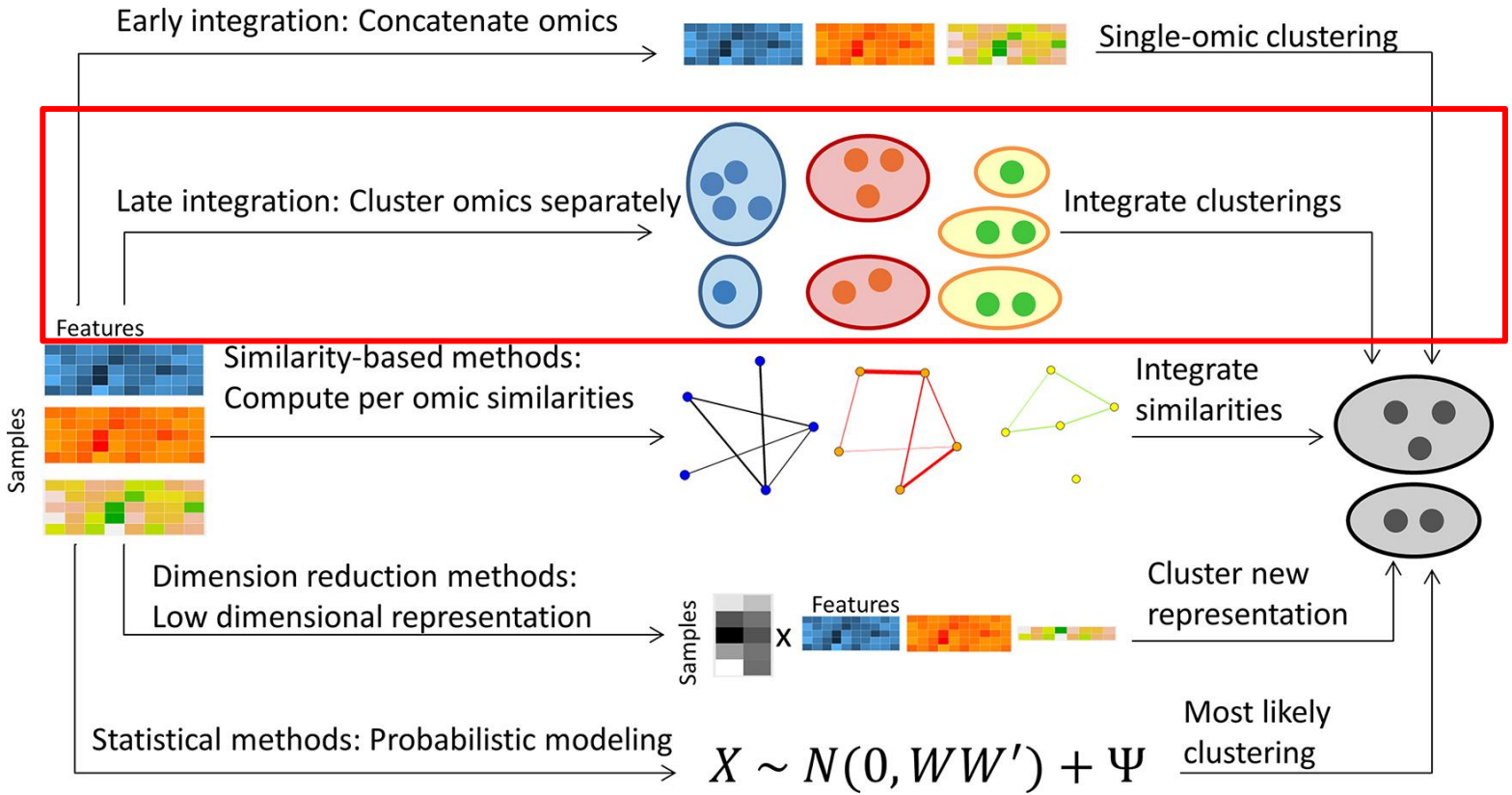


Figure source: Rappoport, N., & Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Research, 46(20), 10546–10562. https://doi.org/10.1093/nar/gky889

Methods for multi-omic data integration

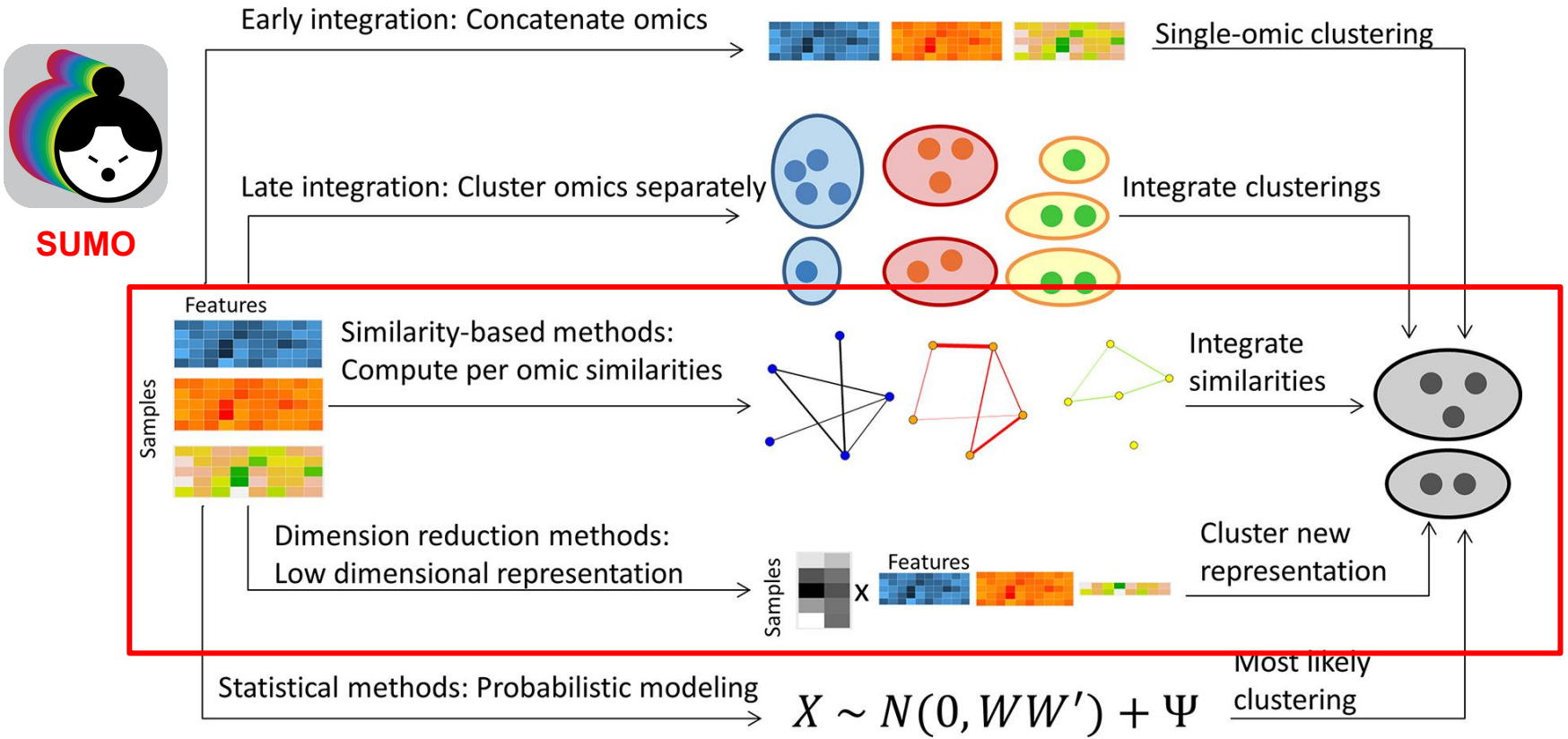


Figure source: Rappoport, N., & Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Research, 46(20), 10546–10562. https://doi.org/10.1093/nar/gky889

Hackathon schedule

NGSprint21

Time (CET)	Day 1 Wed Sep 8	Day 2 Th Sep 9	Day 3 Fr Sep 10	Day 4-5 Sat-Sun Sep 11-12	Day 6 Mon Sep 13	Day 7 Tue Sep 14	Day 8 Wed Sep 15
2 - 3	Opening remarks	Lecture #1	Networking/Social		Lecture #3		
3 - 4	Hackathon meeting #1 3pm-5:30pm	Lecture #2	Hackathon meeting #3 3:30pm-6pm		Lecture #4	Hackathon meeting #5 2:30pm-5:30pm	Hackathon presentation
4 - 5		Hackathon meeting #2 4pm-6:30pm			Hackathon meeting #4 4pm-6:30pm		Closing remarks
5 - 6							
6 - 7	Networking/Social						

Hackathon materials: https://github.com/sienkie/multi-omics-hackathon

Multi-omics data integration

This repository contains materials for **Multi-omics data integration** hackathon for #NGSprint2021.

Running the tutorial materials

To run the tutorial materials you will need:

1. The Jupyter notebook with the R language support:

- Jupyter Notebook (installation information)
- R from CRAN-R (download from here)
- IRKernel which enables the usage of R from the notebooks (see tutorial here)

2. R packages (see the instructions in the *set_up.R* file)

3. python3.6+ (download from here)

4. SUMO: the subtyping tool for multi-omic data (installation information)

R (tidyverse) resource:
https://github.com/sienkie/R_for_data_science