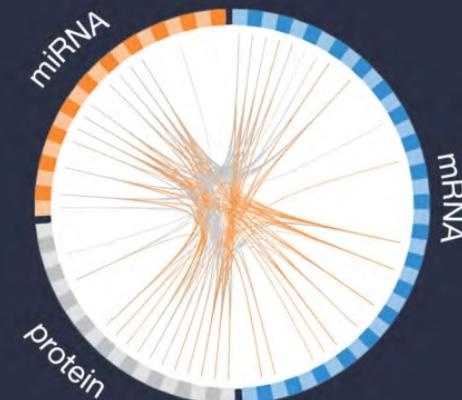


# mixOmics R-WORKSHOP

**Patrick Buerger**  
Marine Omics and Technology Lab  
Applied BioSciences



Contact: [patrick.buerger@mq.edu.au](mailto:patrick.buerger@mq.edu.au)



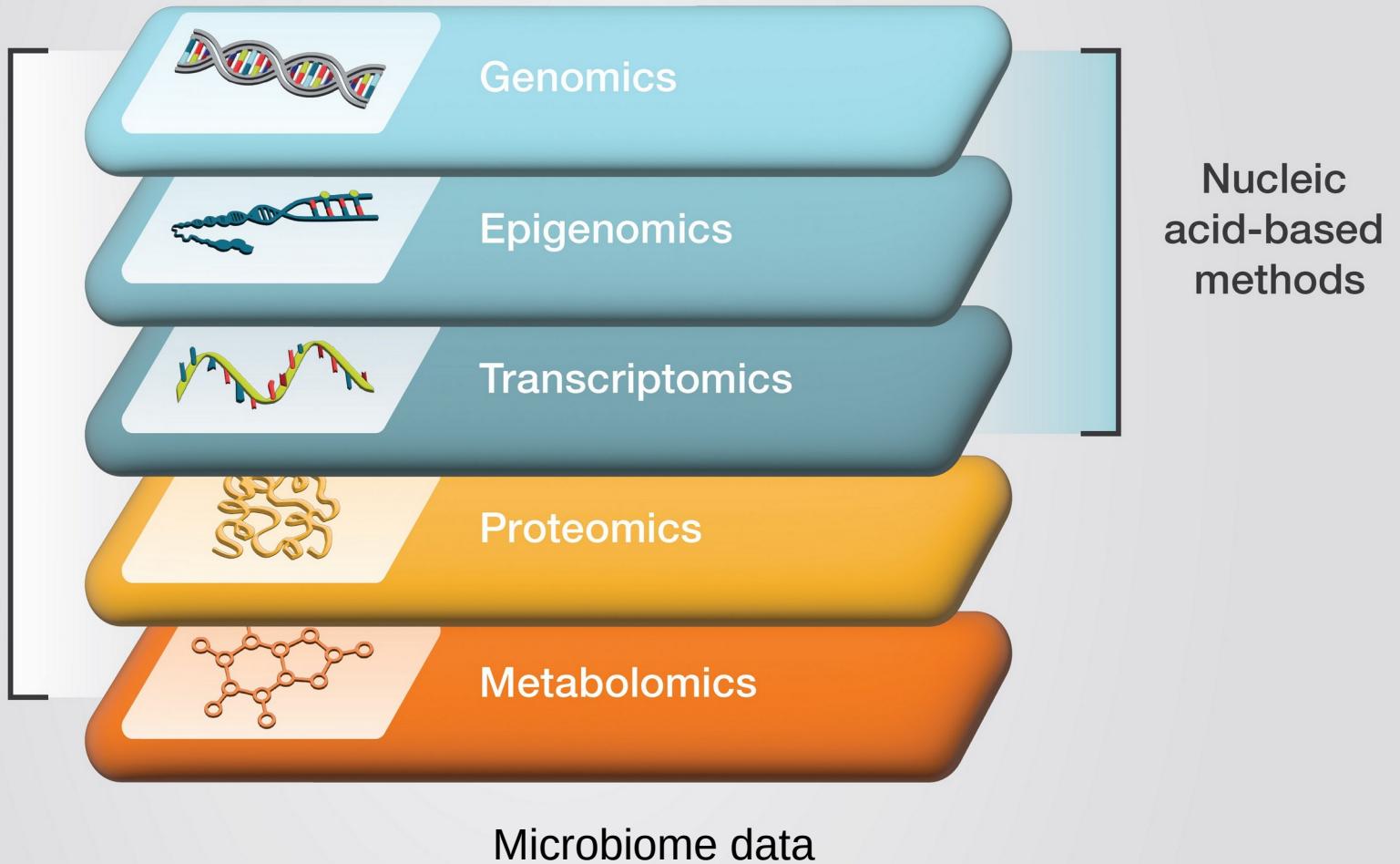
## MULTIVARIATE DATA INTEGRATION USING R

Methods and Applications  
with the mixOmics Package

Kim-Anh Lê Cao  
Zoe Welham

 CRC Press  
Taylor & Francis Group  
A CHAPMAN & HALL BOOK

Integration  
of multi-omics  
data

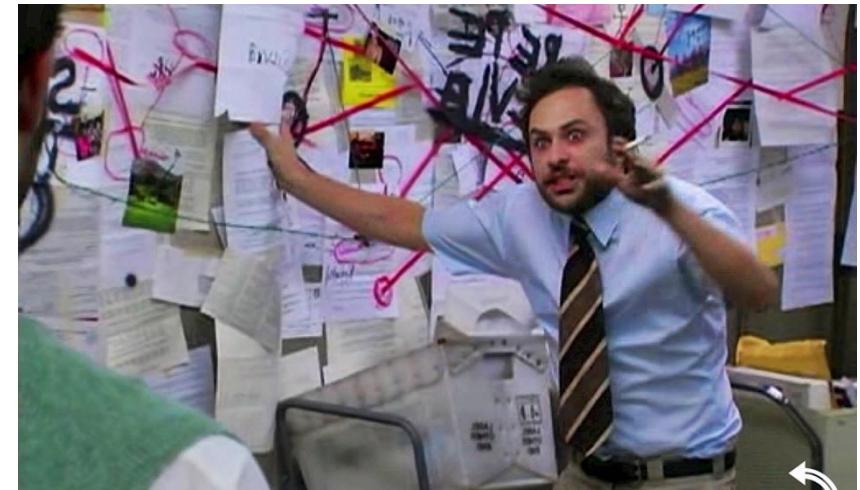


# What to do with your data????

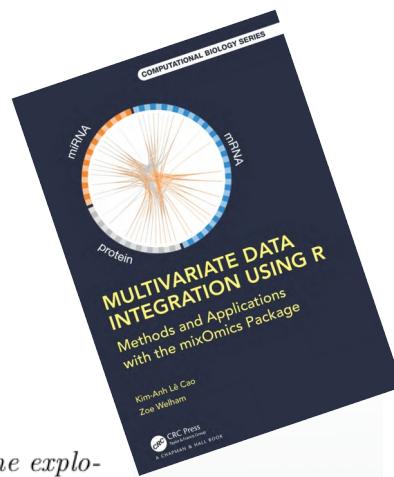
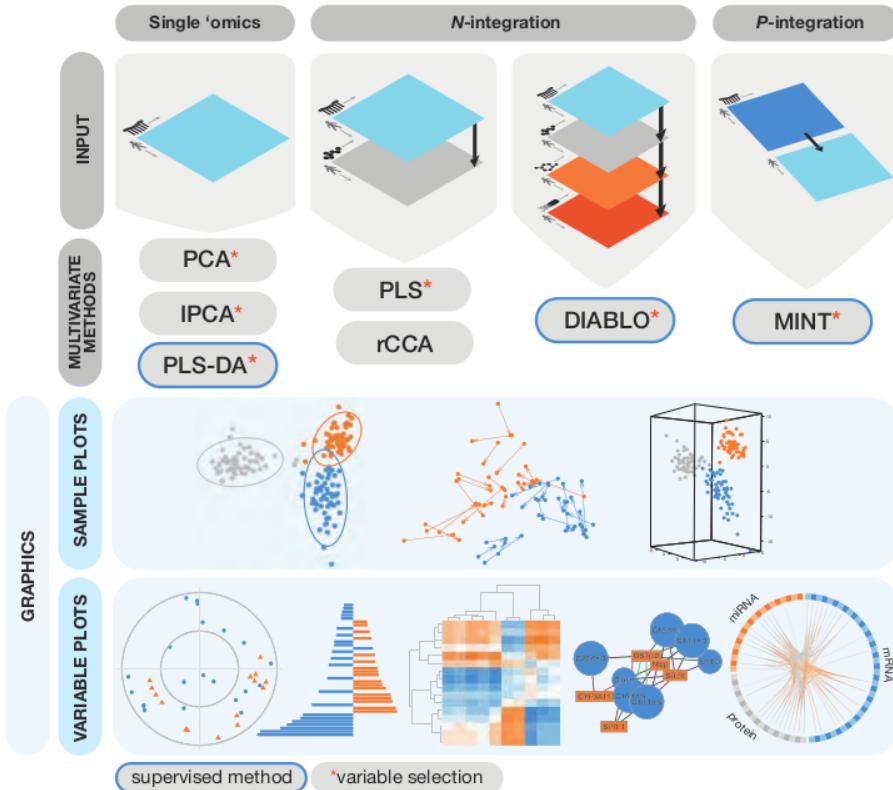


# What do you want to do with your omics types?

- get an **overview** of the role of each data type...
- better understand the **relationship** between the different types...
- identify a molecular signature or **biomarkers**...
- create a **predictive** model...



# What is the mixOmics R-package?



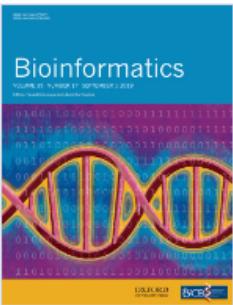
Lab head: [Prof Kim-Anh Lê Cao](#)  
Director of research (Maths & Stats)  
Melbourne Integrative Genomics (MIG) & School  
of Mathematics and Statistics  
Building 184 ground floor | University of  
Melbourne | Parkville VIC 3010

@: [kimanh.lecao@unimelb.edu.au](mailto:kimanh.lecao@unimelb.edu.au) | [twitter](#):  
@mixOmics\_team | Ph: +61 3 8344 3971

**FIGURE 1:** Overview of the methods implemented in the *mixOmics* package for the exploration and integration of multiple data sets. This book aims to guide the data analyst in constructing the research question, applying the appropriate multivariate techniques, and interpreting the resulting graphics.

Framework		Function name	Sparse	Prediction
Single 'omics	unsupervised	pca	-	-
		ipca	-	-
		sipca	✓	-
		spca	✓	-
	supervised	plsda	-	✓
		splsda	✓	✓
<i>N</i> -integration	unsupervised (2 'omics)	rcca	-	-
		pls	-	✓
		spls	✓	✓
	unsupervised	wrapper.rgcca	-	-
		wrapper.sgccca	✓	✓
		block.pls	-	✓
		block.spls	✓	✓
	supervised (DIABLO)	block.plsda	-	✓
		block.splsda	✓	✓
<i>P</i> -integration (MINT)	unsupervised	mint.pls	-	✓
		mint.spls	✓	✓
	supervised	mint.plsda	-	✓
		mint.splsda	✓	✓

## EXAMPLE METHOD



Volume 35, Issue 17  
September 2019

JOURNAL ARTICLE

# DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays FREE

Amrit Singh, Casey P Shannon, Benoît Gautier, Florian Rohart, Michaël Vacher,  
Scott J Tebbutt, Kim-Anh Lê Cao

*Bioinformatics*, Volume 35, Issue 17, September 2019, Pages 3055–3062,

<https://doi.org/10.1093/bioinformatics/bty1054>

Published: 18 January 2019 Article history ▾

PDF Split View Cite Permissions Share ▾

## Abstract

## Motivation

In the continuously expanding omics era, novel computational and statistical strategies are needed for data integration and identification of biomarkers and molecular signatures. We present *Data Integration Analysis for Biomarker discovery using Latent cOmponents* (DIABLO), a multi-omics integrative

Framework		Function name	Sparse	Prediction
Single 'omics	unsupervised	pca	-	-
		<a href="#">ipca</a>	-	-
		<a href="#">sipca</a>	✓	-
		<a href="#">spca</a>	✓	-
N-integration	supervised	plsda	-	✓
		<a href="#">splsda</a>	✓	✓
		<a href="#">rcca</a>	-	-
		pls	-	✓
<i>P</i> -integration ( <a href="#">MINT</a> )	unsupervised	<a href="#">spls</a>	✓	✓
		wrapper.rgcca	-	-
		wrapper.sgccca	✓	✓
		<a href="#">block.pls</a>	-	✓
	supervised	<a href="#">block.spls</a>	✓	✓
		<a href="#">block.plsda</a>	-	✓
		<a href="#">block.spl lda</a>	✓	✓
		<a href="#">mint.pls</a>	-	✓
	supervised	<a href="#">mint.spls</a>	✓	✓
		<a href="#">mint.plsda</a>	-	✓
		<a href="#">mint.spl lda</a>	✓	✓

## ARTICLE

[doi:10.1038/nature11412](https://doi.org/10.1038/nature11412)

# Comprehensive molecular portraits of human breast tumours

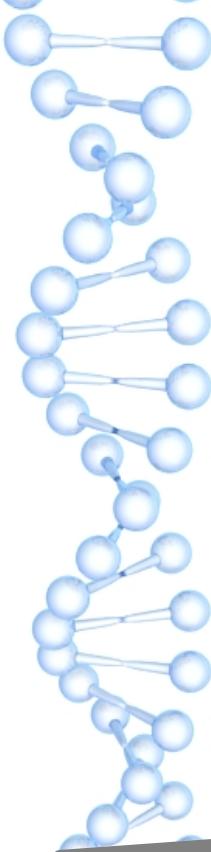
The Cancer Genome Atlas Network\*

We analysed primary breast cancers by genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays. Our ability to integrate information across platforms provided key insights into previously defined gene expression subtypes and demonstrated the existence of four main breast cancer classes when combining data from five platforms, each of which shows significant molecular heterogeneity. Somatic mutations in only three genes (*TP53*, *PIK3CA* and *GATA3*) occurred at >10% incidence across all breast cancers; however, there were numerous subtype-associated and novel gene mutations including the enrichment of specific mutations in *GATA3*, *PIK3CA* and *MAP3K1* with the luminal A subtype. We identified two novel protein-expression-defined subgroups, possibly produced by stromal/microenvironmental elements, and integrated analyses identified specific signalling pathways dominant in each molecular subtype including a HER2/phosphorylated HER2/EGFR/phosphorylated EGFR signature within the HER2-enriched expression subtype. Comparison of basal-like breast tumours with high-grade serous ovarian tumours showed many molecular commonalities, indicating a related aetiology and similar therapeutic opportunities. The biological finding of the four main breast cancer subtypes caused by different subsets of genetic and epigenetic abnormalities raises the hypothesis that much of the clinically observable plasticity and heterogeneity occurs within, and not across, these major biological subtypes of breast cancer.



**Did anyone  
bring their own data?**





# Resources

- Official website with tutorials: <http://mixomics.org/>
- Bugs on Github: <https://github.com/mixOmicsTeam/mixOmics>
- Tutorials:  
[https://mixomicsteam.github.io/mixOmics-Vignette/id\\_06.html](https://mixomicsteam.github.io/mixOmics-Vignette/id_06.html)
- Ask questions: <https://mixomics-users.discourse.group/>
- 

The screenshot shows a software application window for 'mixOmics'. The sidebar on the left lists several sections: '5.2 Example: PLS-DA', '5.3 Example: sPLS-DA', '5.4 Take a detour: prediction', '5.5 AUROC outputs complement p...', '6 N-Integration' (which is currently selected), '6.1 Block sPLS-DA on the TCGA c...', '6.2 Load the data', and '6.3 Parameter choice'. The main content area has a title '6 N-Integration' and a descriptive paragraph: 'N-Integration is the framework of having multiple datasets which measure different aspects of the same samples. For example, you may have transcriptomic, genetic and proteomic data for the same set of cells. N-integrative methods are built to use the information in all three of these dataframes simultaneously.' At the bottom of the main area, there is a footer note: 'mixPLS is a novel mixomics framework for the integration of multiple data sets while explaining their relationships. It is based on Partial Least Squares (PLS) and related methods for Data Integration Analysis for omics data.'

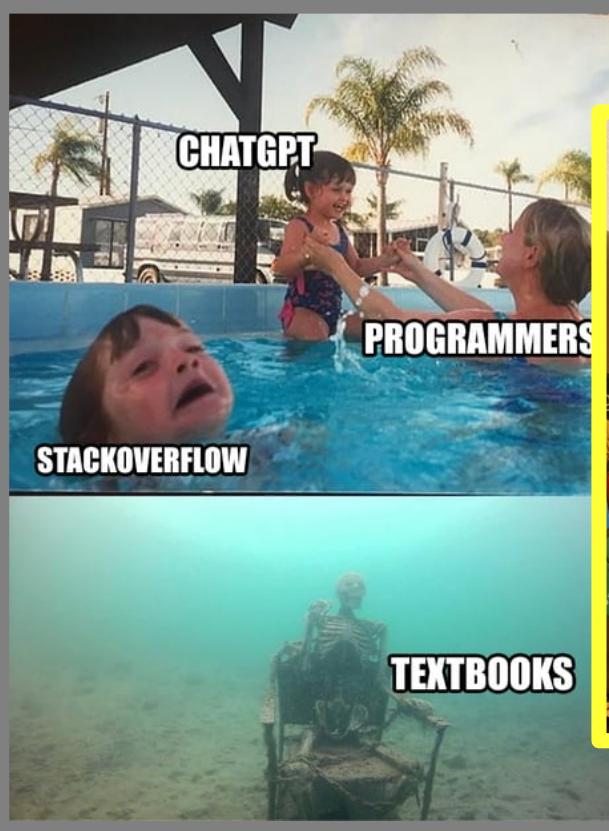
6 N-Integration

N-Integration is the framework of having multiple datasets which measure different aspects of the same samples. For example, you may have transcriptomic, genetic and proteomic data for the same set of cells. N-integrative methods are built to use the information in all three of these dataframes simultaneously.

mixPLS is a novel mixomics framework for the integration of multiple data sets while explaining their relationships. It is based on Partial Least Squares (PLS) and related methods for Data Integration Analysis for omics data.

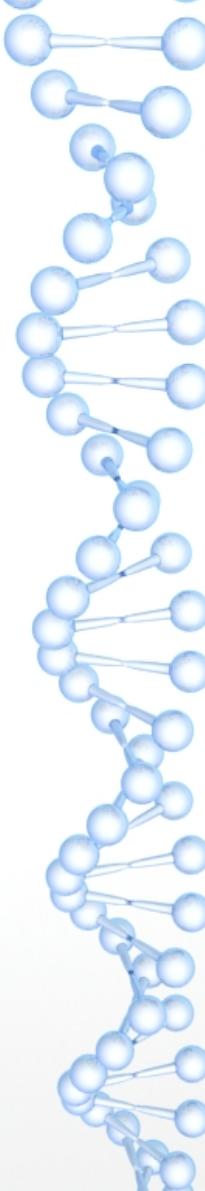


# LETS START SOME CODING!!!!



Learns random sentences from random people	✓	✓
Talks like a person but doesn't really understand what it's saying	✓	✓
Occasionally speaks absolute non sense	✓	✓
Is a cute little bird	✓	✗





# WRAP UP

- Different statistical methods:
  - **PCA, PLS-DA, sPLS-DA, DIABLO**
- Different visualisation tools:
  - **Arrow Plot, Variable Plot, Loadings Plot**
- How to work with own data:
  - **Normalisation, format, other considerations**

