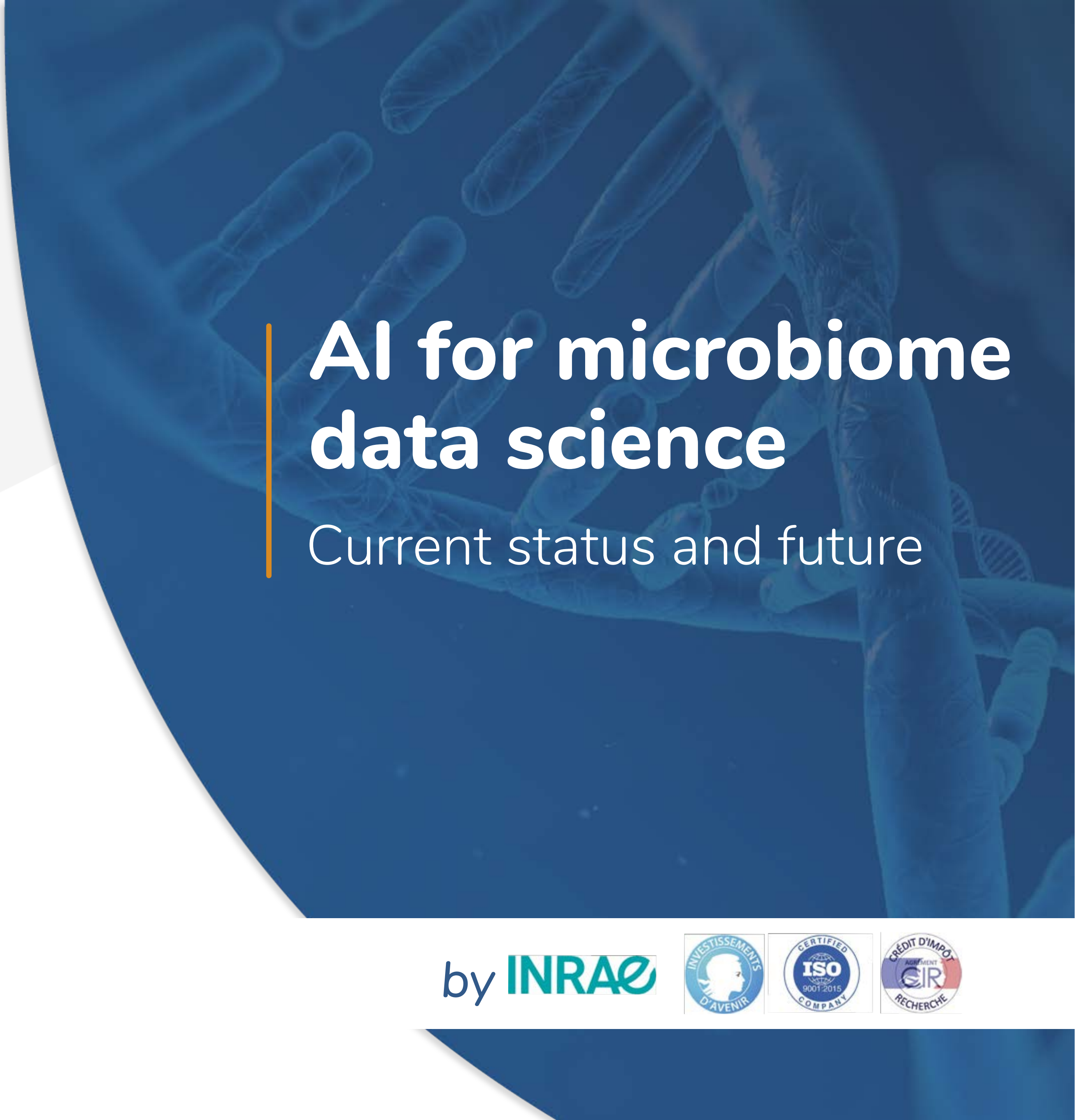
metagenopolis
mgps.eu



AI for microbiome data science

Current status and future

MetaGenoPolis

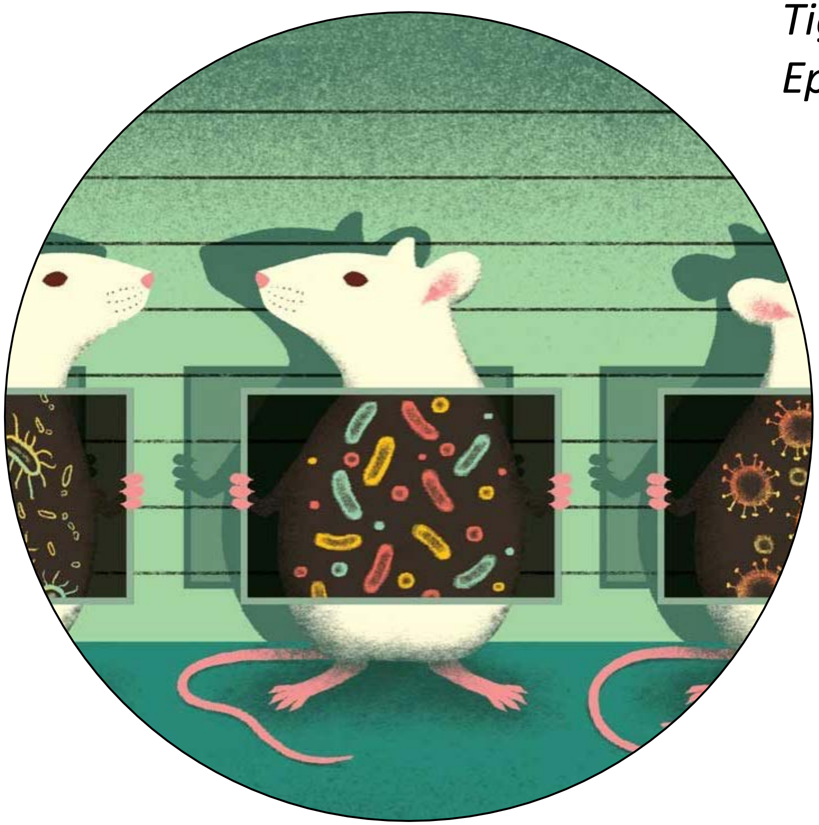
Centre de recherche INRAE de Jouy-en-Josas
Domaine de Vilvert, Bât.325
78 350 Jouy-en-Josas France

by INRAE

The gut microbiota

An association resulting from a long co-evolution:
– host
– microbiota
– genome
– metagenome

For humans:
- 23 000 human genes
- 500 000+ bacterial genes



Integrity of the mucosa
Tight junctions
Epithelium cellular renewal

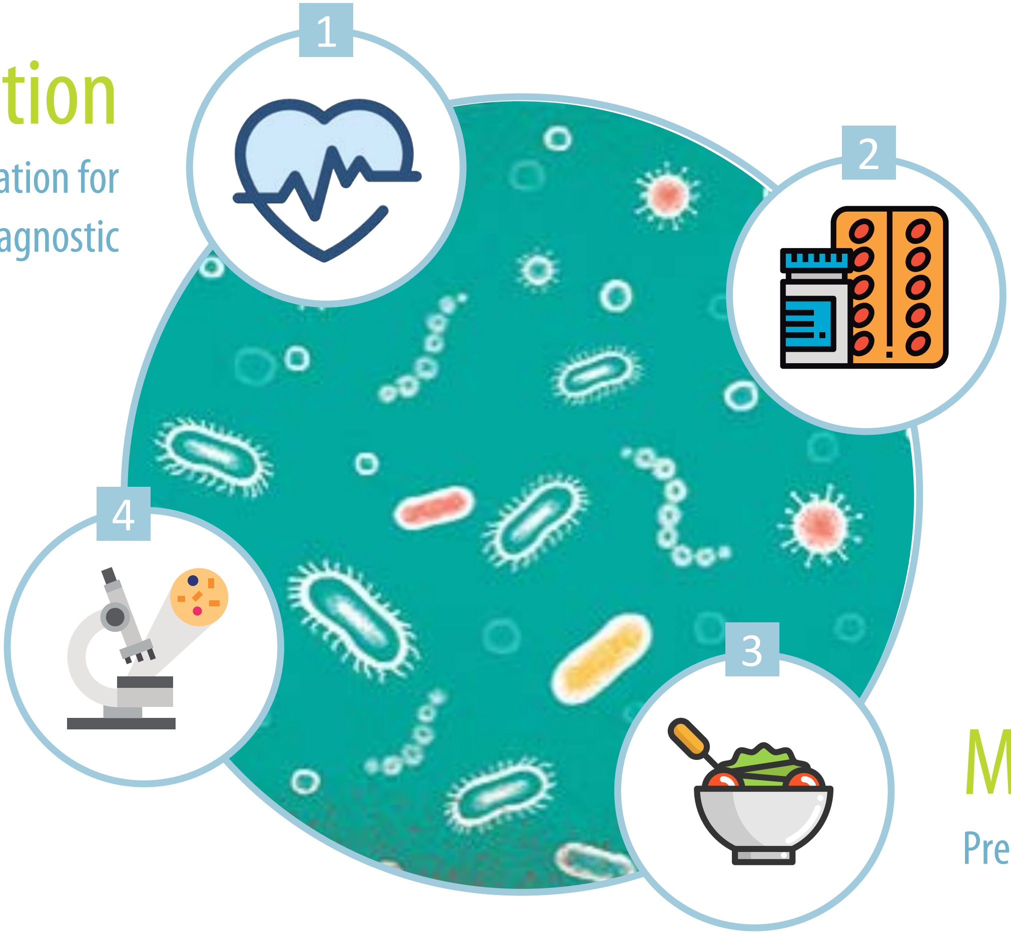
Barrier effect
Prevention of the proliferation of pathogens

Metabolism
Fiber degradation
Metabolites production
Bioavailability modulation
Energy extraction

Immune system
Stimulation and maturation

Gut-brain axis

Applications in nutrition, health and well-being



Stratification

Personalization for diagnostic

New treatments

New therapeutic targets

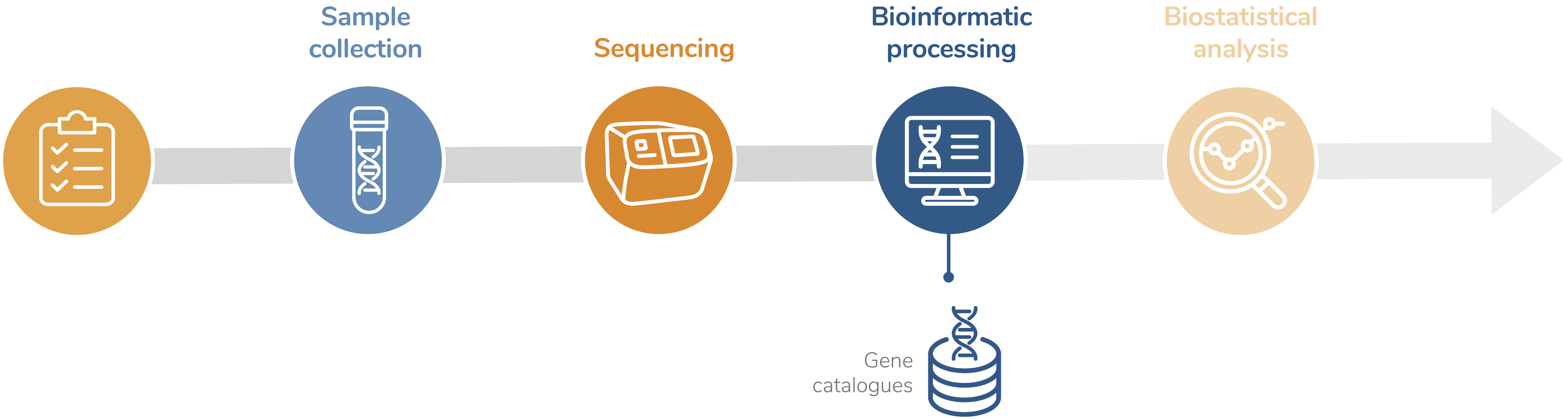
Modulation target

Preventive or curative

The microbiota, saviour organ

Microbiome transplantation

The MGP analysis pipeline



Sample collection

Sequencing

Bioinformatic processing

Biostatistical analysis

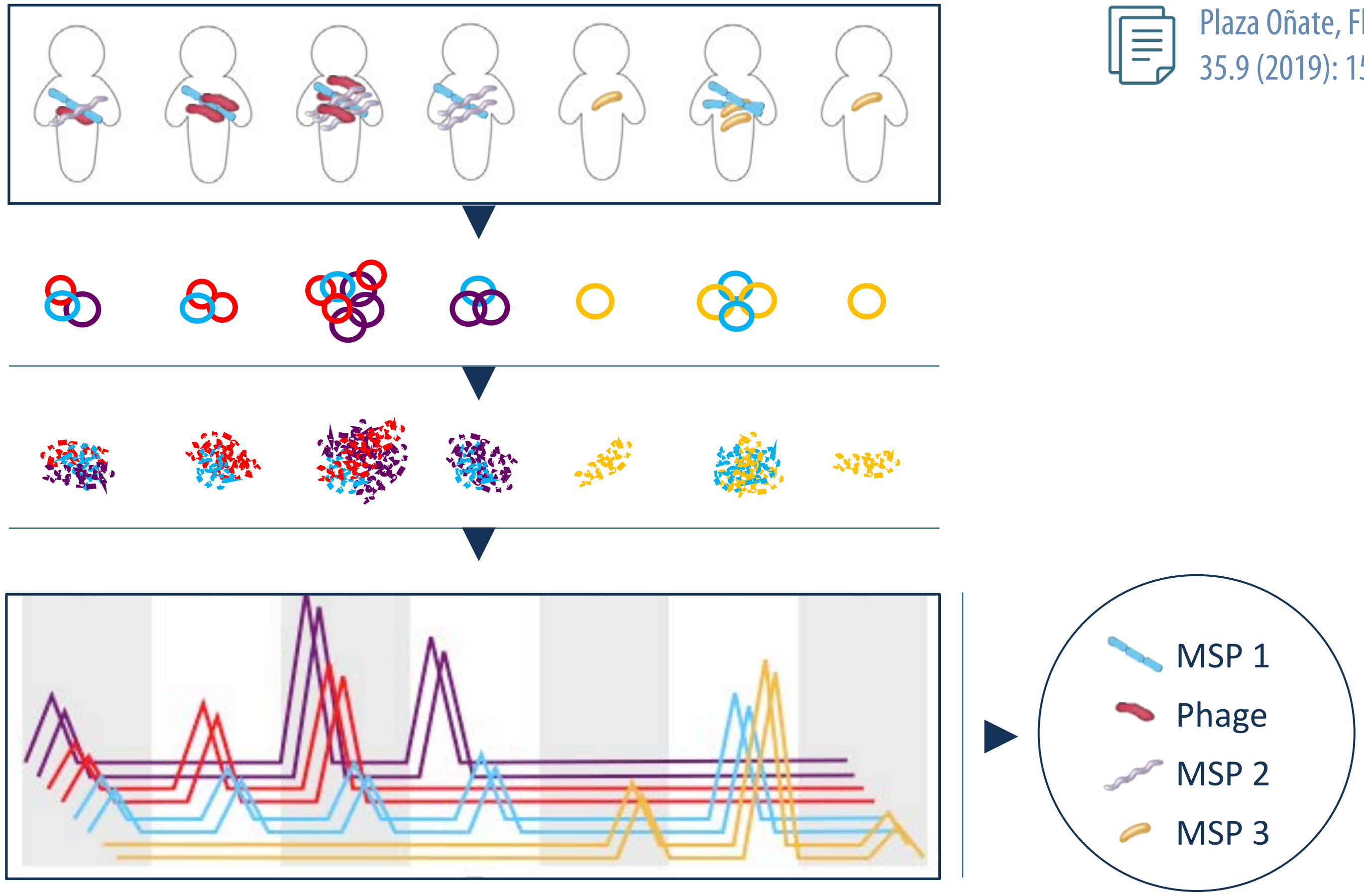
Gene catalogues

Number of variables:
Human: 10.4 M genes
Mouse: 2,6 M genes
Pig: 7,7 M genes
Cow: 13 M genes

Wen, Chengping, et al. Genome biology 18.1 (2017): 1-13.

Gene clustering in Metagenomic Species (MSP)

Plaza Oñate, Florian, et al. Bioinformatics 35.9 (2019): 1544-1552.



Intestinal gut bacteria

DNA extraction

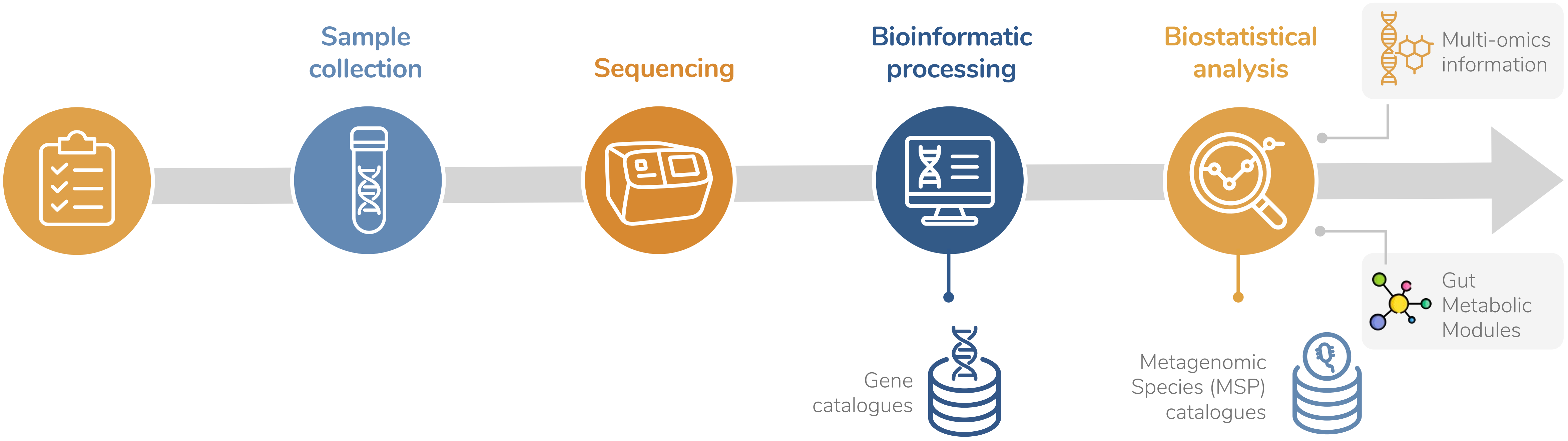
High throughput sequencing

Abundance profiles of gene catalogue

Organize the catalogue by co-variance

Table of abundance Metagenomic Species

The MGP analysis pipeline



Sample collection

Sequencing

Bioinformatic processing

Biostatistical analysis

Multi-omics information

Gut Metabolic Modules

Gene catalogues

Metagenomic Species (MSP) catalogues

Number of variables:
Human: 10.4 M genes
Mouse: 2,6 M genes
Pig: 7,7 M genes
Cow: 13 M genes

Number of variables:
Human: 1990 MSP
Mouse: 541 MSP

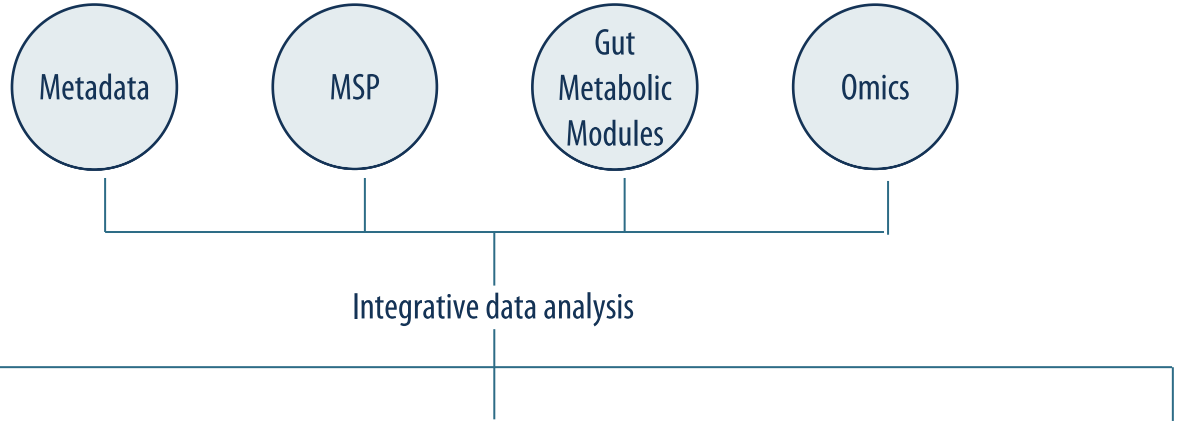
Input data for biostatistical analysis:
Abundance tables with

- Tens to hundreds observations
- Thousands to millions variables

Routine exploratory statistical analyses

metagenopolis mgps.eu

Know-how:
- Effect Size metrics
- Parametric & nonparametric tests
- Correlations
- Data visualization

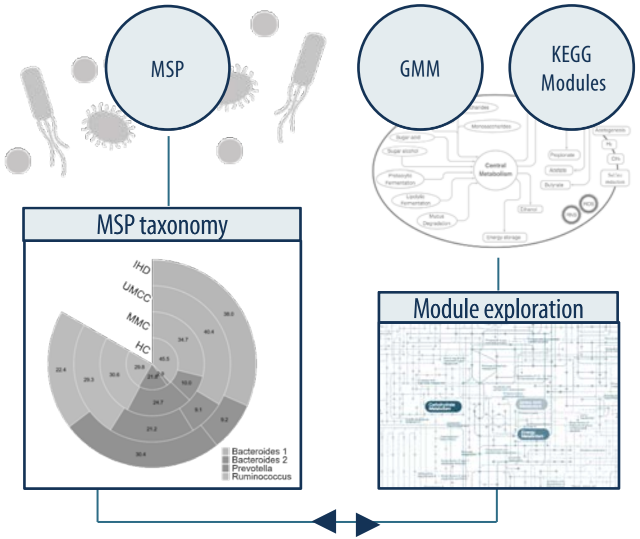


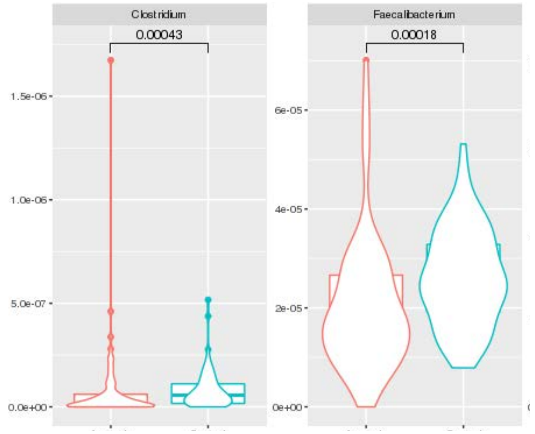
Tools: RStudio, tidyverse, Shiny, rmarkdown, dplyr, ggplot2, tidyr

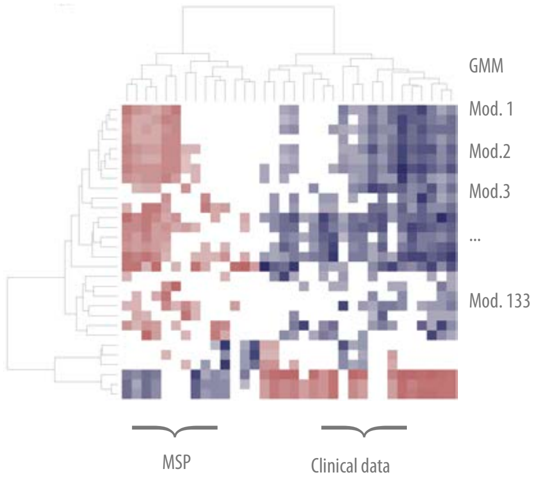
Taxonomic & functional composition

Identification of the changes in the microbiota

Associations between microbial features and clinical parameters







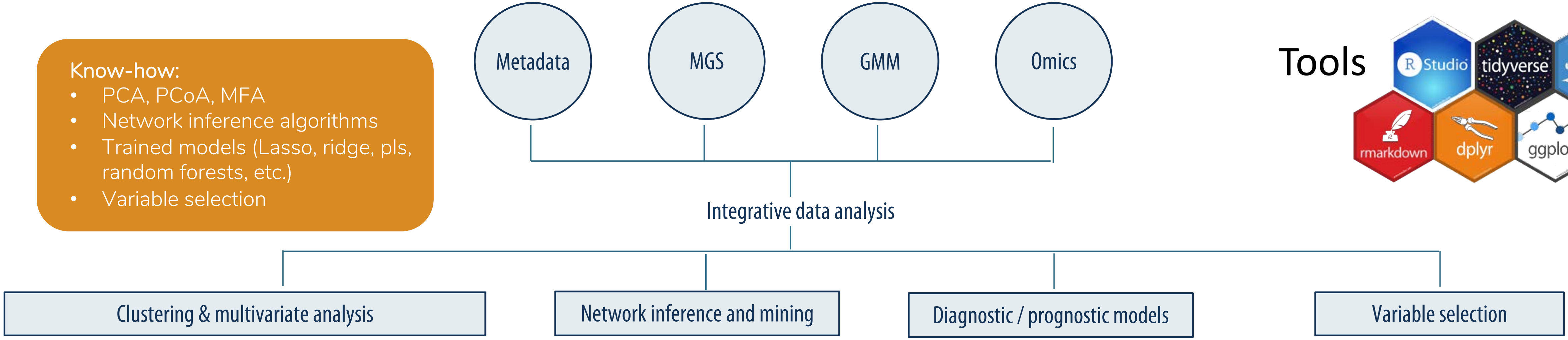
Routine AI tools for (un)supervised learning

Know-how:

- PCA, PCoA, MFA
- Network inference algorithms
- Trained models (Lasso, ridge, pls, random forests, etc.)
- Variable selection

Tools

RStudio, tidyverse, Shiny, rmarkdown, dplyr, ggplot2, tidyr



Metadata

MGS

GMM

Omics

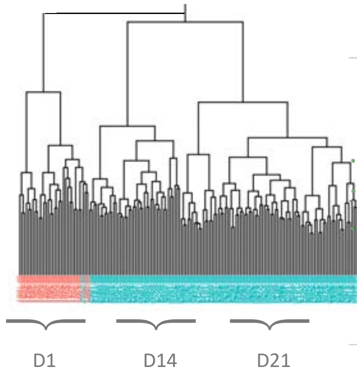
Integrative data analysis

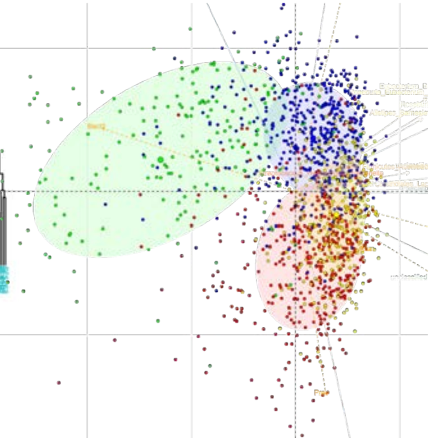
Clustering & multivariate analysis

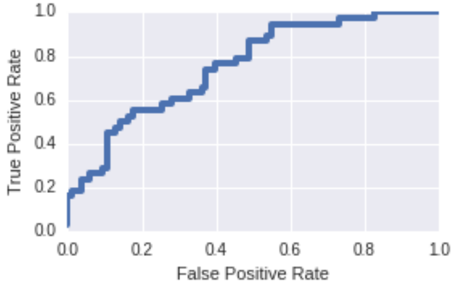
Network inference and mining

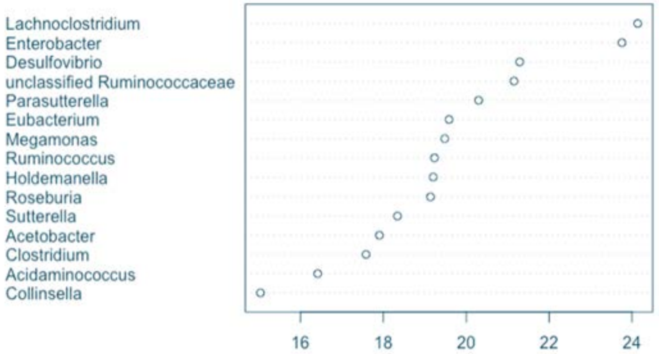
Diagnostic / prognostic models

Variable selection









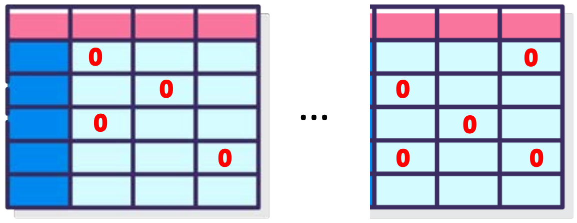
AI challenges for microbiome data science

1. Statistical specificity of the data
2. High inter-individual variability
3. Complex dependency structure

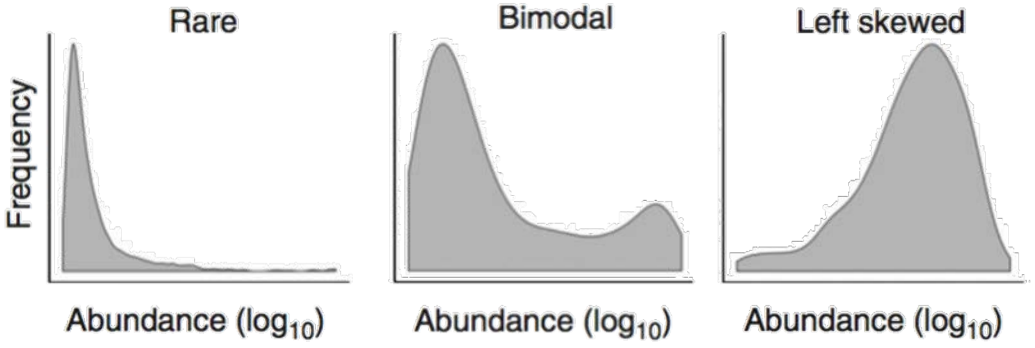
AI challenges for microbiome data science

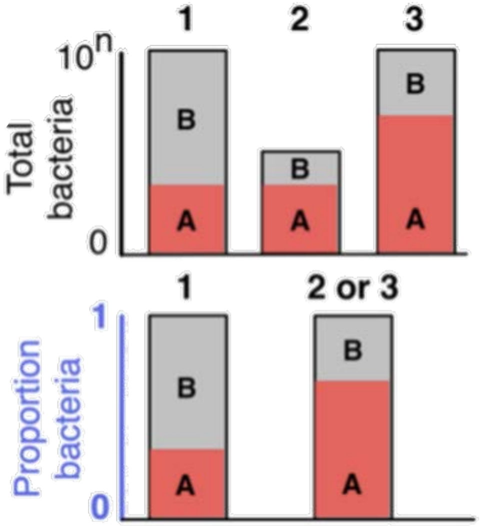
Statistical specificity of the data

- More variables than observations
- Sparse (structural or informational zeros)
- Do not follow a gaussian distribution
- Compositional



from 65 to 95 % of zeros

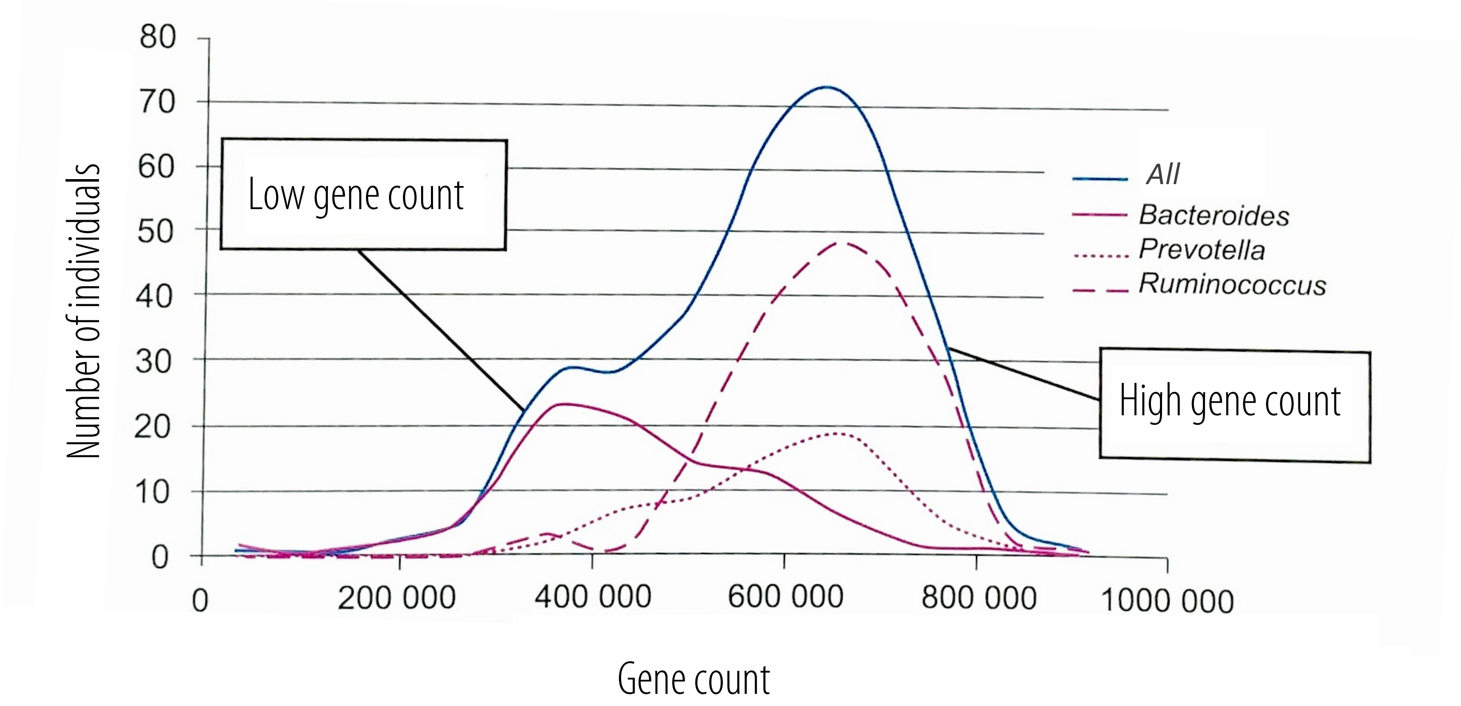


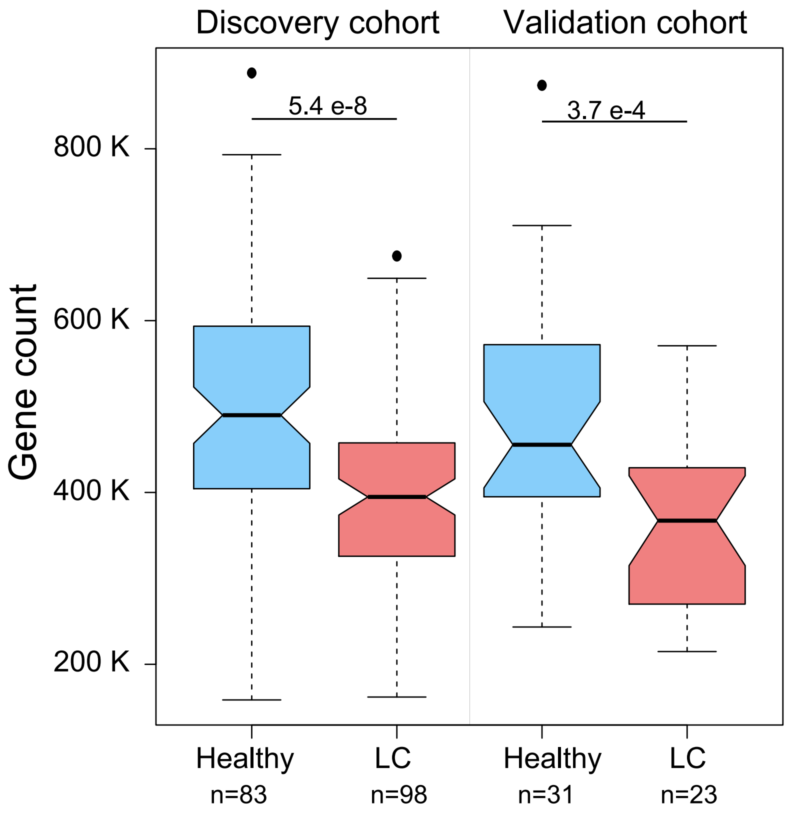


AI challenges for microbiome data science

High inter-individual variability

& limited data available



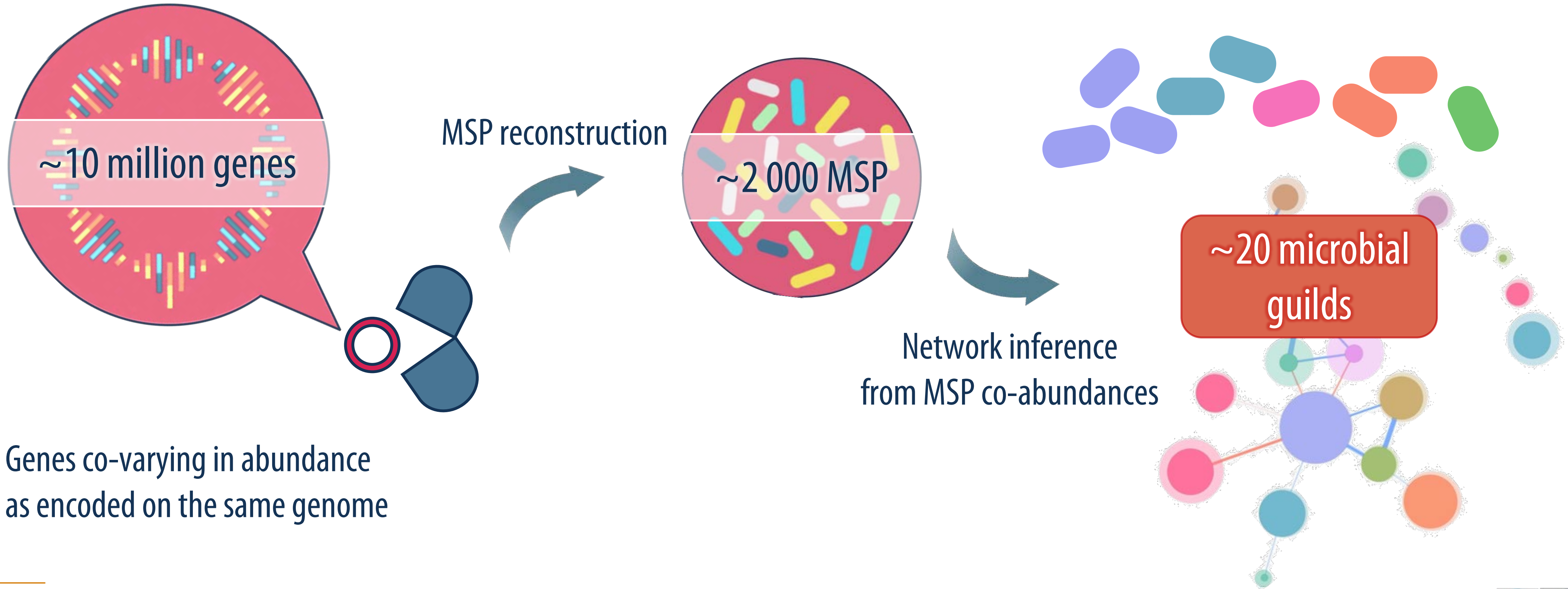


Marteau, Philippe, and Joël Doré. Ed John Libbey (2017).

Qin, Nan, et al. Nature 513.7516 (2014): 59-64.

AI challenges for microbiome data science

Complex dependency structure with high functional redundancy



~10 million genes

MSP reconstruction

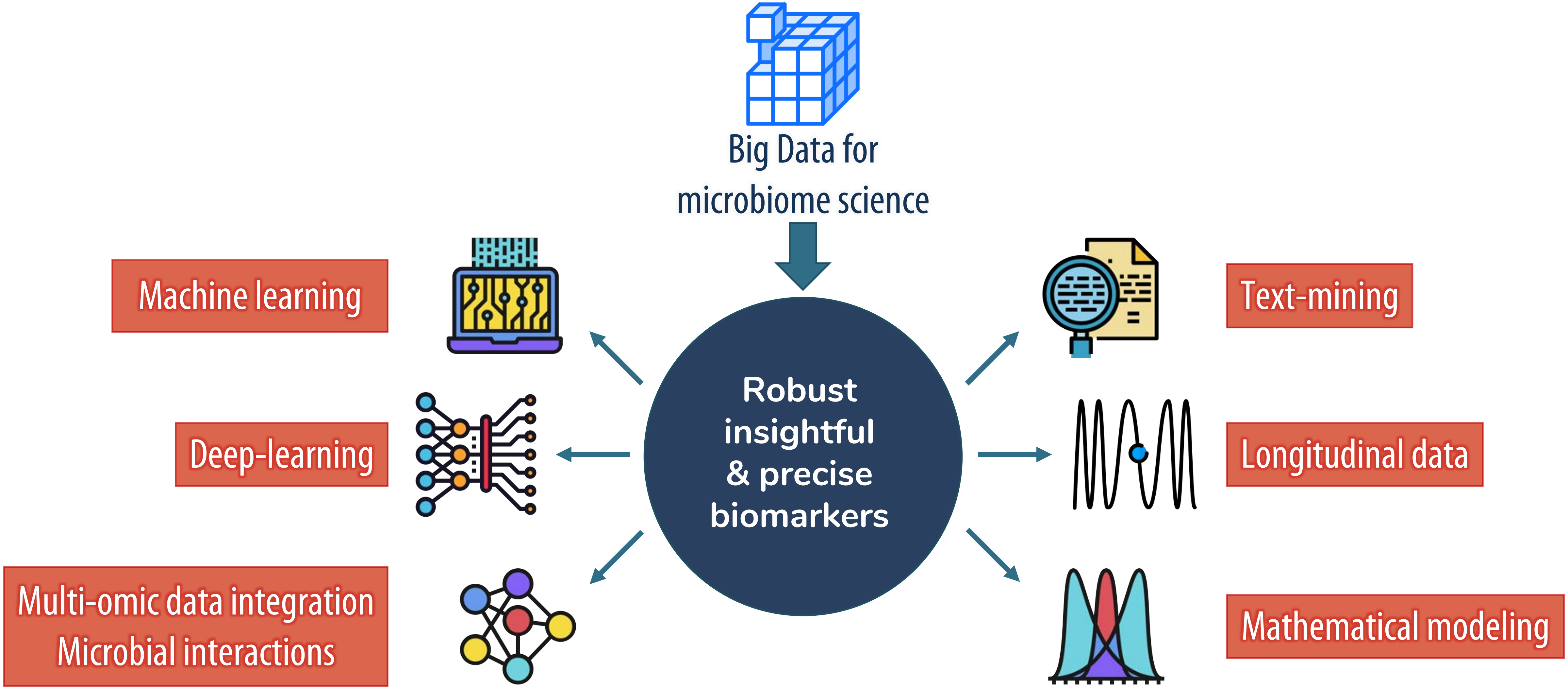
~2 000 MSP

Network inference from MSP co-abundances

~20 microbial guilds

Genes co-varying in abundance as encoded on the same genome

Future AI tools for microbiome data



Big Data for microbiome science

Robust insightful & precise biomarkers

Machine learning

Deep-learning

Multi-omic data integration
Microbial interactions

Text-mining

Longitudinal data

Mathematical modeling

FRENCH GUT PROJECT

A citizen science project at the national level with the ambition to better define the heterogeneity of the French healthy gut microbiome and its deviations in chronic diseases

French participation to the
Million Microbiome of Humans Project (MMHP)

French Gut
(100 000 gut metagenomes)
INRAE consortium with
public institutions,
industry & foundation

French Gut - Model

Funding

- Public funding
- Industrial co-financing
- Donation via the Microbiome Foundation

MICROBIOME FOUNDATION

Open science

Public annual release for metagenomic data and basic information

Duration: 5 years

Phase 1: data acquisition

Recruitment strategy
Evolutive & balanced for health status & age

- Citizens Volunteers
- Patient networks & cohorts

Questionnaires

- **Basic information:**
  Age, sex, BMI, health status, country
- **Precise phenotypic profiling:**
  (60+ questions)
  health, life style, diet

Shotgun metagenomic sequencing at MGP

Phase 2: data exploitation

Data analysis
Integrating metagenomic data and phenotypic profiling

MGP's expertise in bioanalysis

- Public partnership projects [€]
- Pre-competitive projects [€€]
- Competitive projects [€€€]

New questionnaires
Hypothesis driven

Shotgun metagenomic
for a specific group of individuals

Thanks

© Pictograms from flaticon.com

@MgpsLab magali.berland@inrae.fr