Integration and analysis of heterogeneous biological data modelled with multilayer graphs for a better understanding of feed efficiency

Metaprogram DIGIT-BIO
- ✓ **Axe 1 : Multiscale deciphering of living functions**
- ❑ **Axe 2 : Prediction of phenotypes**
- ❑ **Axe 3 : Trasnsfer and generalisation**

PhD (01/12/2020 – 30/11/2023)

Supervisors
Florence GONDRET, INRAE, UMR PEGASE
Emmanuelle BECKER, Univ Rennes 1, UMR IRISA

Context – What is feed efficiency in swine?

The relative ability of the animal to convert feed into weight gain
- Reducing feed costs and wastage
- While maintaining physiological functions and health

⇒ Profitability and sustainable of the animal production

Key but **complex** phenotype summarizing multiple biological routes

⇒ Determining the biological basis for inter-individual by analyzing the dependency structure of the data : key regulators and important networks

Objectives

1. Starting hypothesis

Key drivers of the phenotypic divergence can be better defined by considering the different levels of biological organization between biological entities

2. Research questions

- How can we link high-throughput data (transcriptomics and metabolomics) collected in different tissues to animal phenotype?
- What is the plus-value of adding open-source knowledge on biological entities?

3. Objectives

- Identify important sub-networks at different –omics levels
- Determine the interdependence between entities through reactions in which they participate
- Evaluate the most efficient scenarios through path lengths of reactions between entities

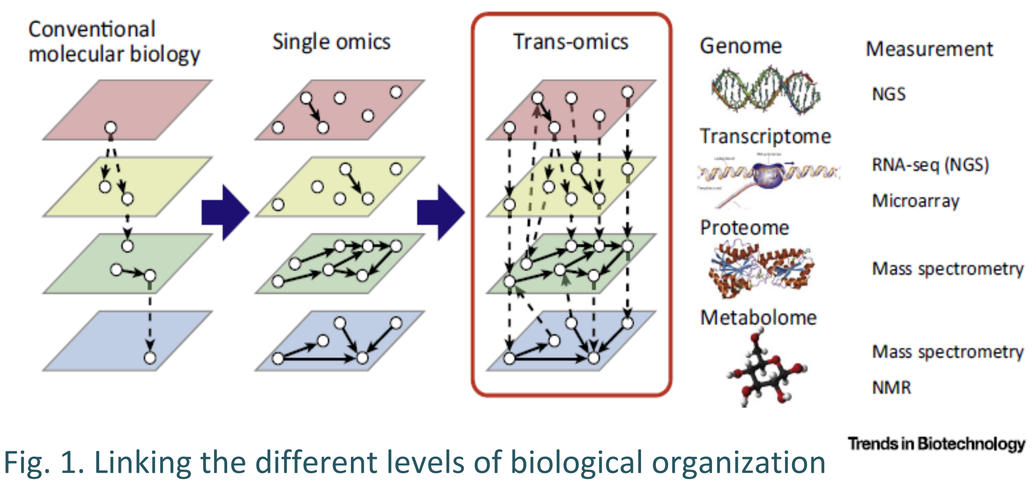

Fig. 1. Linking the different levels of biological organization allows for a holistic view of biological entities (source: K. Yuri et al.)

Methods

1. Statistical approach

1. WGCNA on microarray data from blood
- Analyze pair-wise correlations to find modules of genes = co-expression modules
- Link these modules to feed efficiency measures, such as Feed Conversion Ratio (FCR)

2. PCA on ^{1}H-NRM data and target metabolites from blood and muscle
- Identify new coordinates (=profiles) summarizing metabolomic data
- Link these profiles to co-expression networks of genes

2. Multilayer network approach

1. Data integration : Interactions connecting transcriptome and metabolome
- Biological Pathway Exchange format (BioPAX) used to represent pathways at cellular and molecular levels
- Reference ontologies (UniProt, ChEBI) and Semantic Web technologies (SPARQL) to retrieve molecules from the modules in graphs of interactions

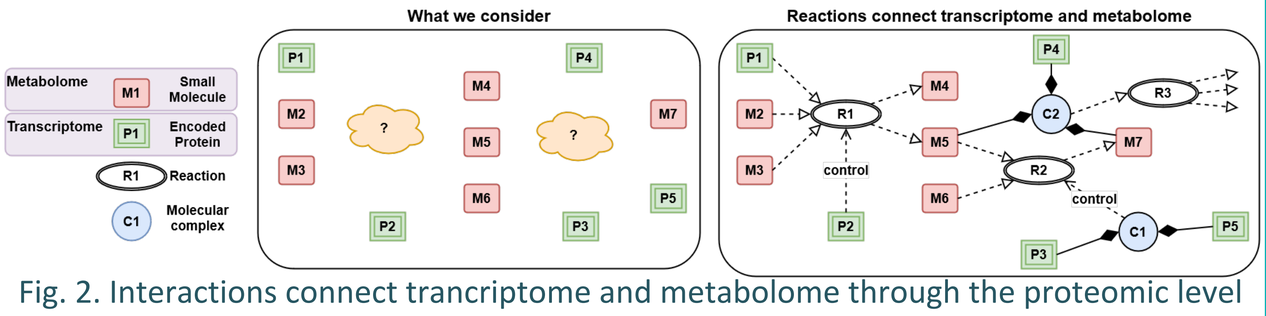

Fig. 2. Interactions connect trancriptome and metabolome through the proteomic level

2. To be continued : graph analysis

Path-finding algorithms to describe the interactions between entities within a module and connecting gene networks to metabolic profiles

Results (1)

- 33 modules identified in microarray data (from 13 to 2,490 unique genes)
- 6 of which are significantly correlated to FCR
- Functional annotation bioinformatics with enrichment scores made biological sense to these modules
- 5 modules of genes ⇔ metabolic profiles in blood

Results (2)

- SPARQL queries allowed identifying and fixing non-compliance with BioPAX specifications for complexes of molecules, to avoid indirection between reactions during path finding

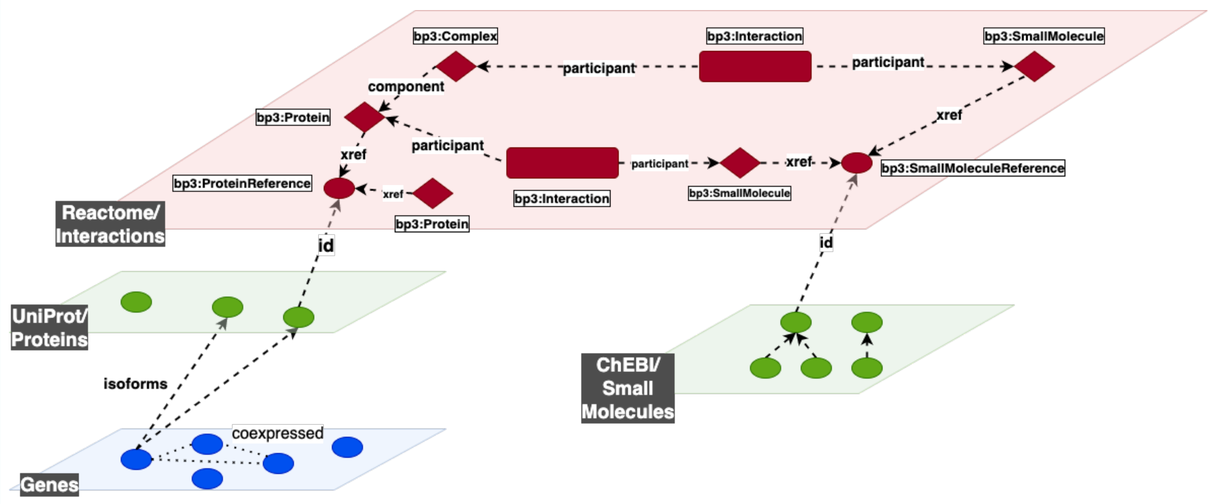

Fig. 3. Integration schema. Each layer represents a type of data from a different data source linked to the others layers by different properties

Centre
Normandie - Bretagne

Camille Juigné
camille.juigne@inrae.fr
cjuigne.github.io/cjuigne/