



Adopting Metabolic Modeling Tools in the Biopharmaceutical Industry

Dr. Anne Richelle

Process Systems Biology &
Engineering Center of Excellence
GSK Vaccines, Rixensart, Belgium

anne.x.richelle@gsk.com



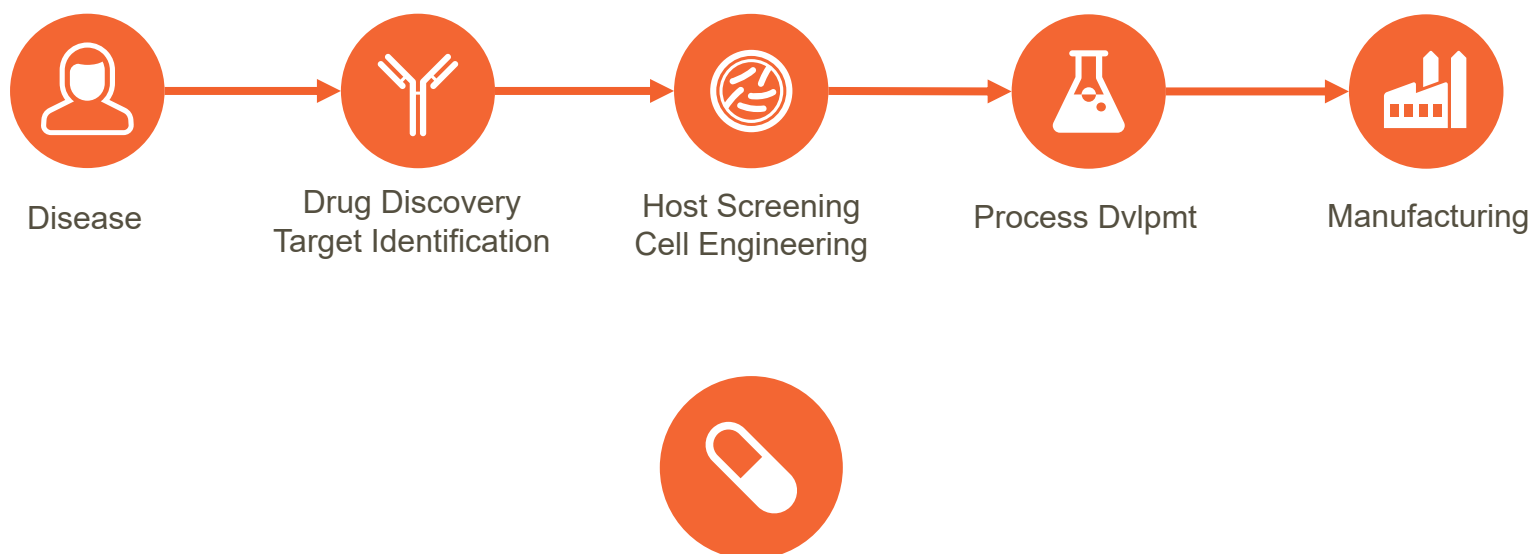
Disclosure statement

Anne Richelle is an employee of the GSK group of companies

Biopharmaceutical Industry



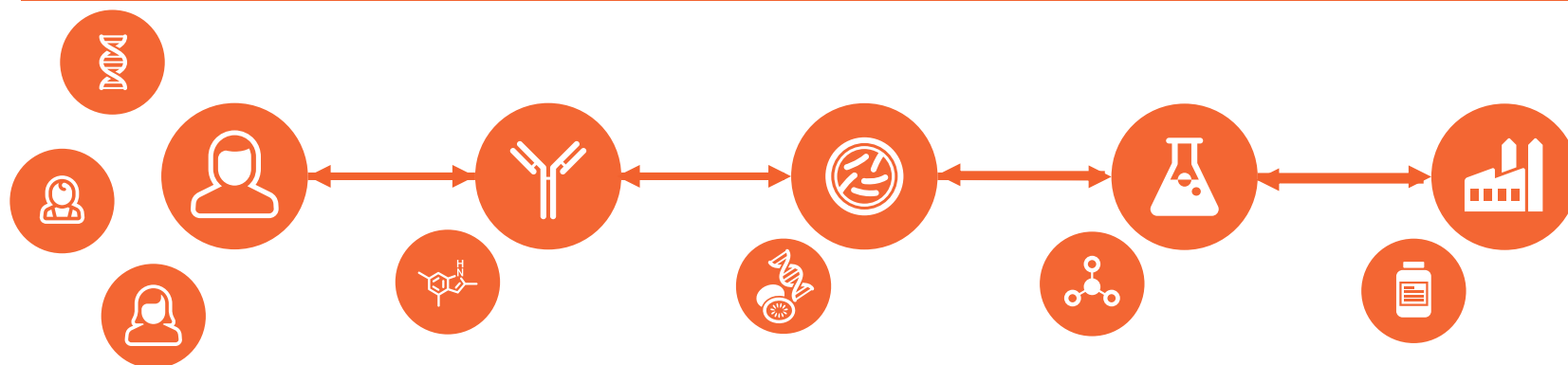
Timeline of research activities



Biopharmaceutical Industry



Numerous influence factors



Data acquisition

Big Data?



- Biopharma = low and expensive data
- databases not often in machine-readable format*
- lack of standardization and hidden variables

➔ The problem is the integration of diverse data sources and the extraction of meaningful information

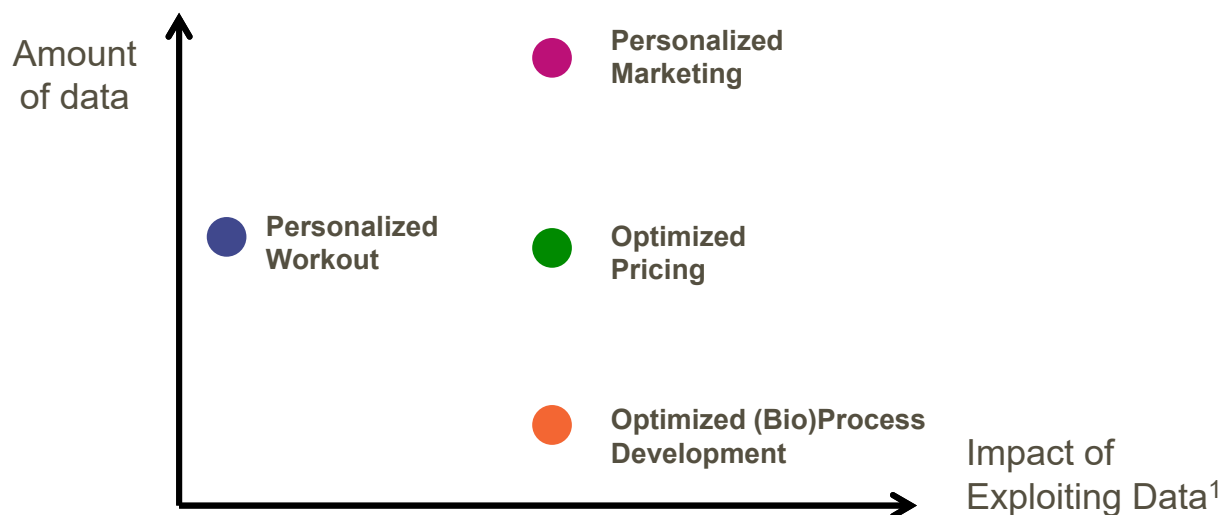
*This is potentially subject to change in a near future thanks to the work of C. Malamud at JNU in New Delhi (cf. Pulla. *The plan to mine the world's research papers*. [Nature](#). 2019 Jul;571(7765):316-318)

Machine Learning



The discovery driver for large datasets ?

Very powerful tool that help researchers in all fields to extract information from data



1. SOURCE: McKinsey Global Institute analysis

BUT...



ML tools can also turn up fool's gold

- Can really large databases “speak for themselves”?
 - ➔ Big Structures are full of spurious correlations due to non-causal coincidences, hidden factors, nature of big randomness¹
- “Many of the algorithms are so complicated that it is impossible to inspect all the parameters or to reason about how exactly the inputs have been manipulated”²
- Typical ML approaches are black boxes
 - ➔ you can build predictive models from your data but you can't provide an explanation for their results

Need to take into account the empirical knowledge that
have been achieved over decades of research in biology

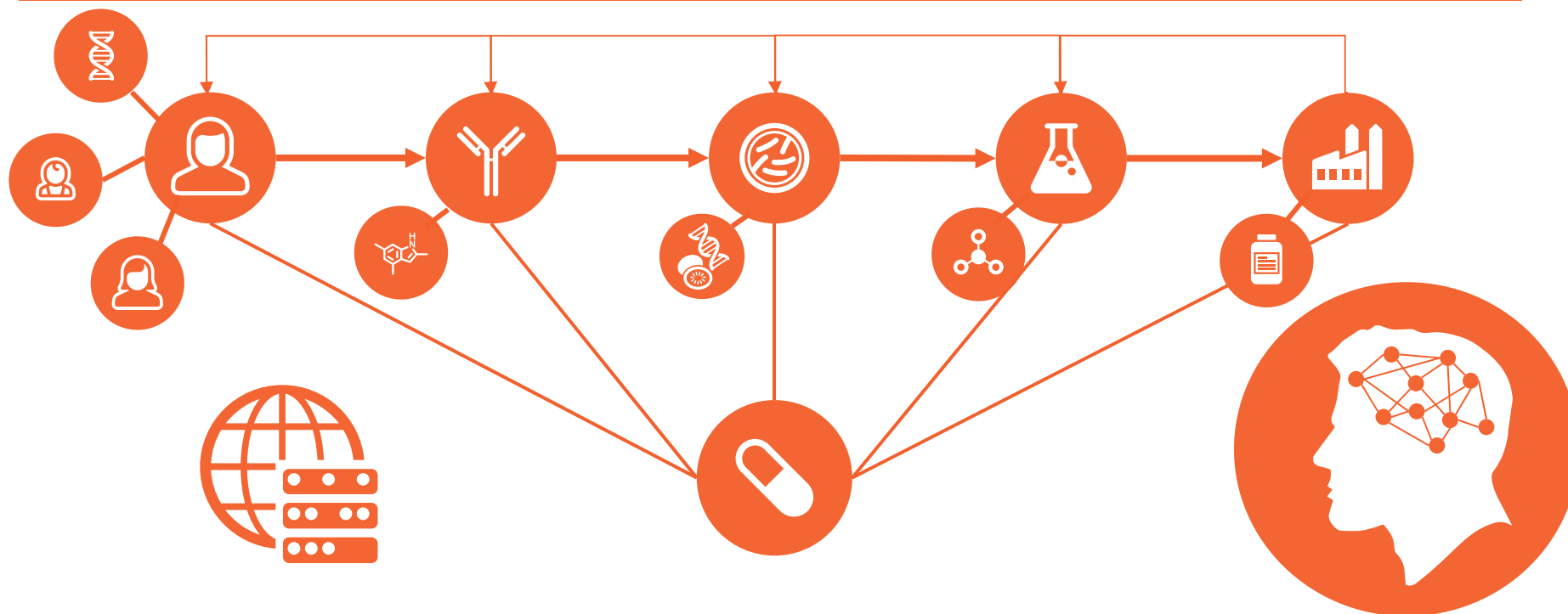
1. Cristian S. Calude, Giuseppe Longo, (2016), The Deluge of Spurious Correlations in Big Data, [Foundations of Science](#)

2. Patrick Riley. Three pitfalls to avoid in machine learning. Nature. 2019 Aug;572(7767):27-29. doi: 10.1038/d41586-019-02307

Systems thinking



The influence factors are related and dependent

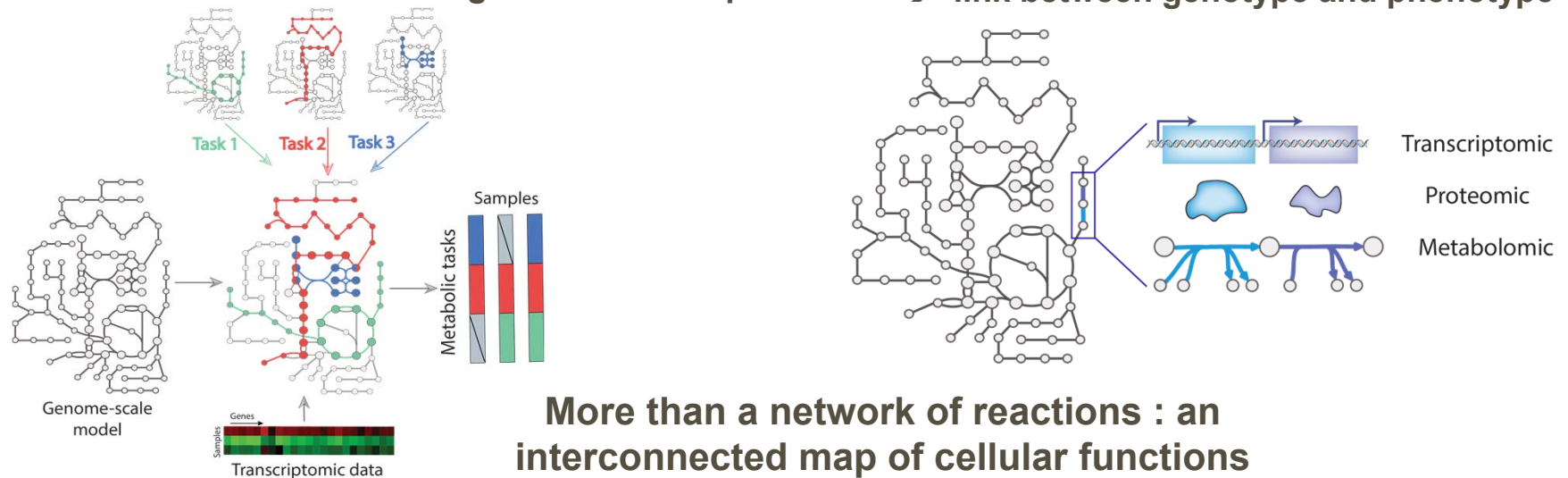


Systems biology



A network-based approach for the description of complex biological systems

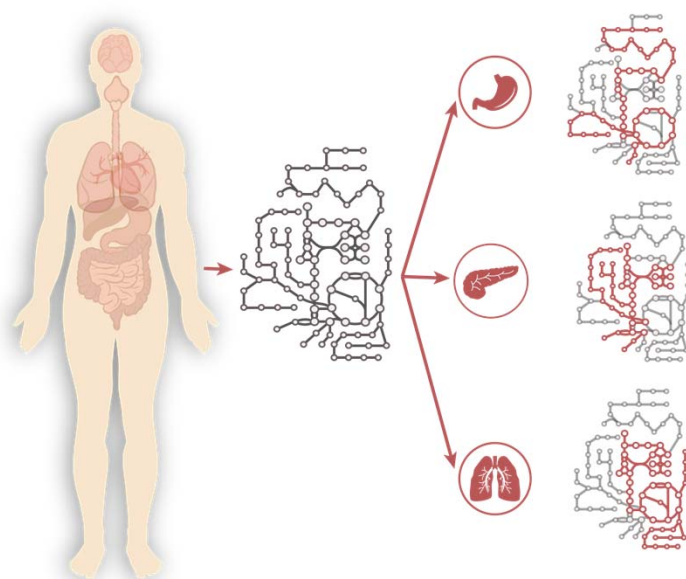
- Coherent organization of large datasets into biological networks
→ provide non-intuitive insights on biological systems that *in vivo* experiments alone cannot provide
- Platform for –omics data **integration** and **interpretation** → **link between genotype and phenotype**



Biological networks



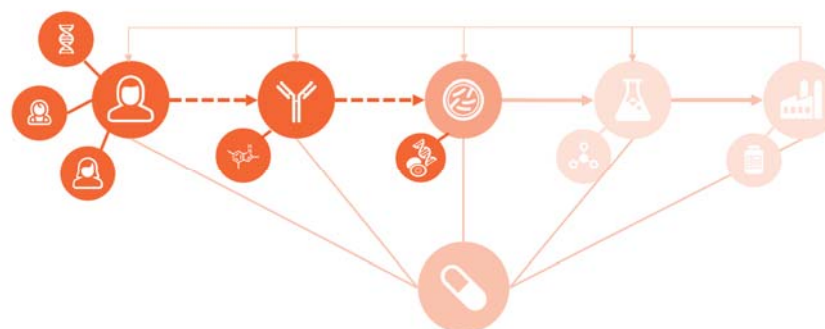
A library for cell- and tissue-specific model building



Development of algorithms to recapitulate the metabolism of specific cell and tissue types → useful insights in the metabolism under these conditions¹

Systems tools have been proven to be invaluable at the level of preclinical research

→ e.g., design of new drugs by informing target selection, engineering of cells by rewiring metabolism toward the production of a product of interest

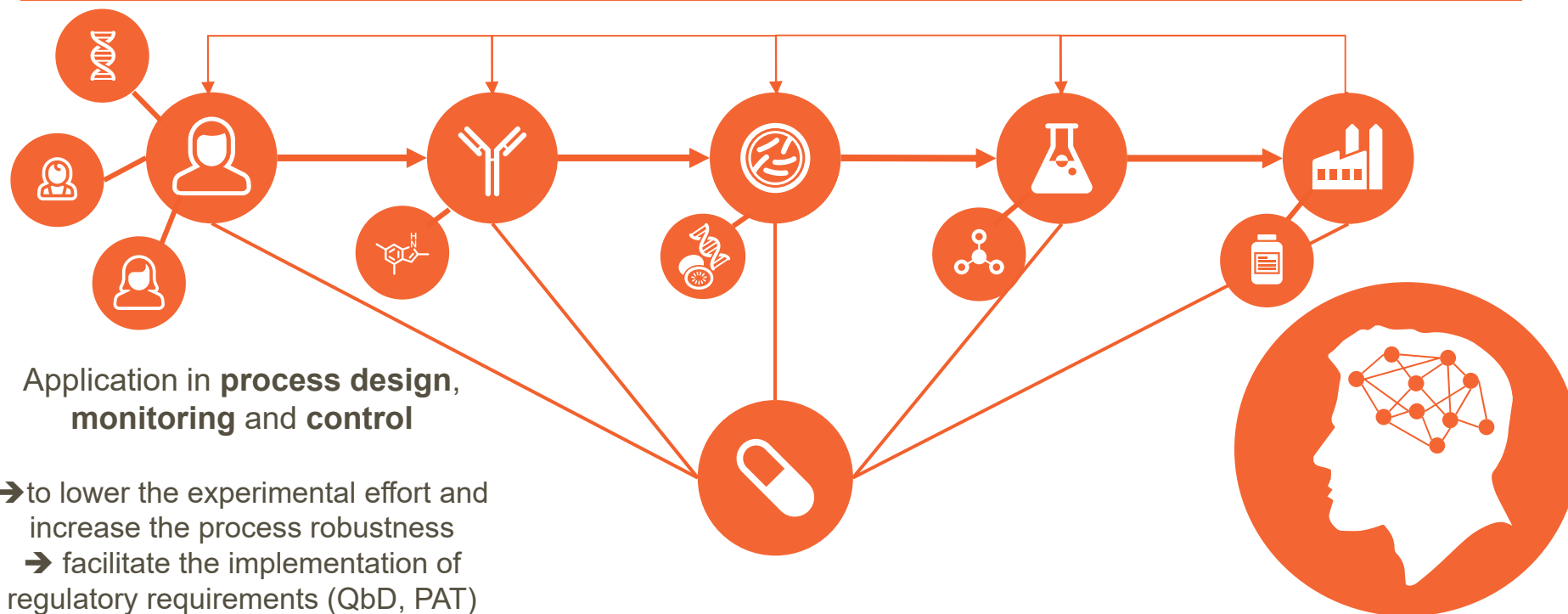


BUT these approaches could be used for much more than this !

1. Opdam et al. Systematic evaluation of methods for tailoring genome-scale metabolic models. [Cell Syst.](#) 2017 Mar 22;4(3):318-329.e6. doi: 10.1016/j.cels.2017.01.010

Systems biology

Beyond drug discovery and cell design



Gaps in our knowledge and technical limitations



Key aspects restraining the routine implementation of SB tools in Biopharma

1. Real-Time Monitoring : Track to Better Control

- Real-time monitoring technologies = critical aspect of the automatization
- On-line in-situ measurements of metabolites and cell concentrations limit our capacity to have a live snapshot of the cell metabolism
- Increasing interest for spectroscopic methods (e.g., NIR, FTIR, mid-IR, Raman, Fluorescence)
 - ➔ provide a “molecular fingerprint” of samples
 - ➔ UNFORTUNATELY the capacity to systematically extract accurate quantitative data is, **for the moment**, limited to a handful number of metabolites
- Advances in the chemometric modeling field will be the key for the effective extraction of the maximum information from these type of spectra
- FURTHERMORE the rise of on-line single-cell probing opens the door for new generation of high-throughput -omics technologies
 - ➔ near real-time measurements of cell transcriptome and proteome available in a close future?

Gaps in our knowledge and technical limitations



Key aspects restraining the routine implementation of SB tools in Biopharma

1. Real-Time Bioprocess Monitoring : Track to Better Control

2. Complexity of Metabolic Networks : Reduce to Better Predict

- Complexity of large networks also hinders their utility in various practical applications
 - Underdetermined systems = multitude of alternate solutions
 - Too complex to be handle for the development of utilitarian tools (e.g., feed-back control loop)
- Different approaches have been proposed for tailoring these metabolic networks based on *a priori* knowledge and/or available experimental data
 - How to define what phenotype should be “protected” during the reduction?
 - Currently no quantitative description of the GPR relationship → use of strong assumptions to link the GPR expression and the metabolic reaction activity
- “Network tailoring” approaches generally do not completely leverage the problem of system underdetermination → choice of adequate strategy to solve the system will always be required to achieve an instantaneous picture of the flux distributions in the cell

Gaps in our knowledge and technical limitations



Key aspects restraining the routine implementation of SB tools in Biopharma

1. **Real-Time Bioprocess Monitoring : Track to Better Control**
2. **Complexity of Metabolic Networks : Reduce to Better Predict**
3. **Modeling with Hybrid Approaches : Combine to Effectively Implement**
 - Significant increase of AI and ML applications in bioprocess engineering
 - ➔ e.g., automated counting of cell colonies grown on petri-dishes using digital imaging¹
 - Could be used to effectively establish the link between metabolism and operating conditions in a way that can be used to implement control and optimization strategies
 - HOWEVER very few applications combine AI/ML with systems biology tools ...
 - ➔ Absence of dedicated software and interdisciplinarity of the expertise required
 - ➔ Several major open question such as the effective generation of sufficiently informative experimental data and the joint parameter identification of the mechanistic and data-driven parts of the model

1. Ferrari et al.. (2017). Bacterial colony counting with convolutional neural networks in digital microbiology imaging. *Pattern Recognition*, 61:629-640.

Biopharmaceutical Industry

Towards the systematic application of systems thinking

