Navigating Big and Bigger Complexity to uncover the Secrets of Health Data

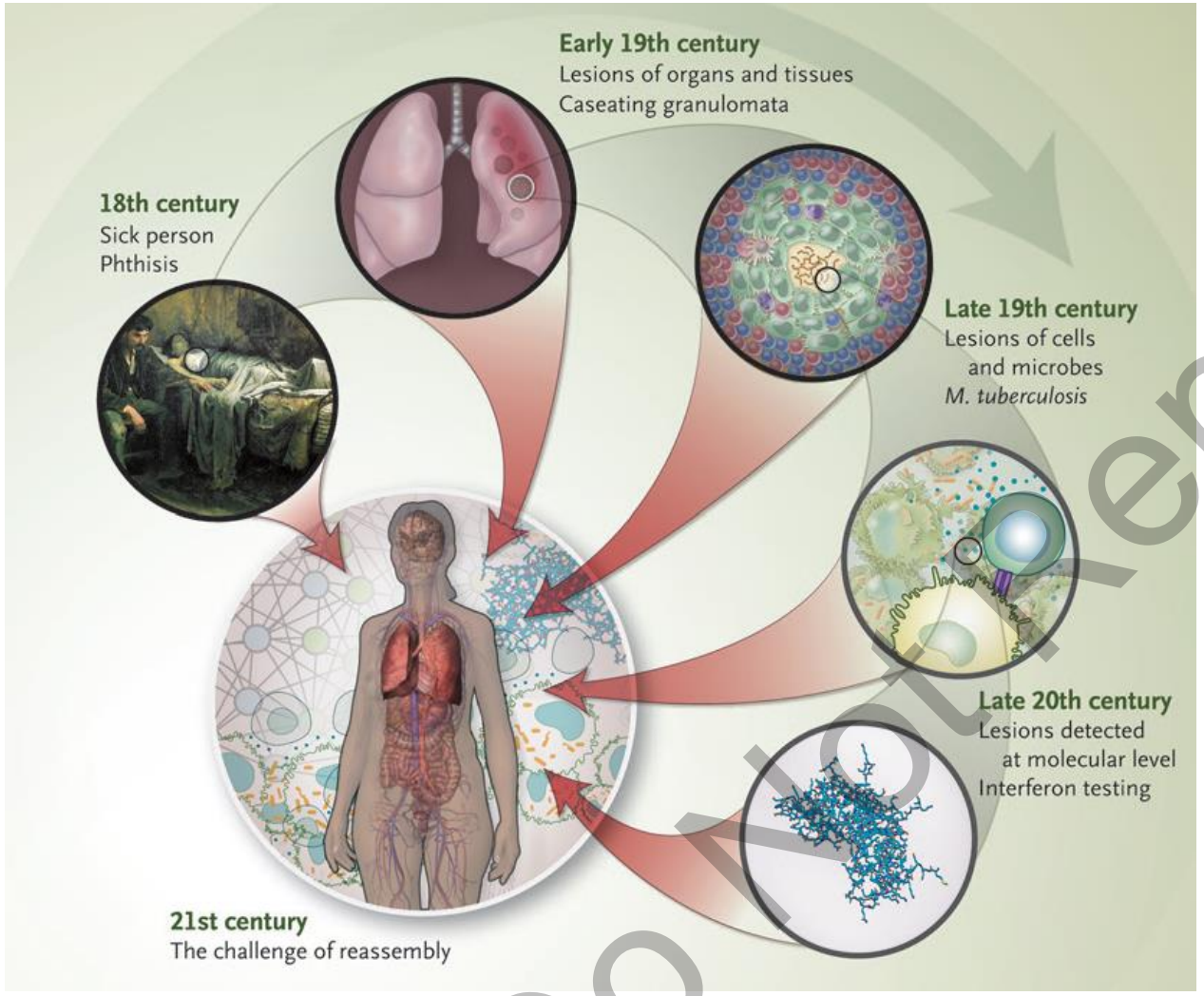
J-L Balligand, MD, PhD

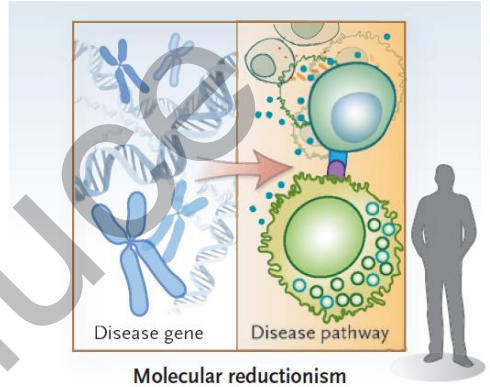
Pharmacology and Therapeutics (FATH), Institut de Recherche Experimentale et Clinique (IREC), UCLouvain (Brussels)

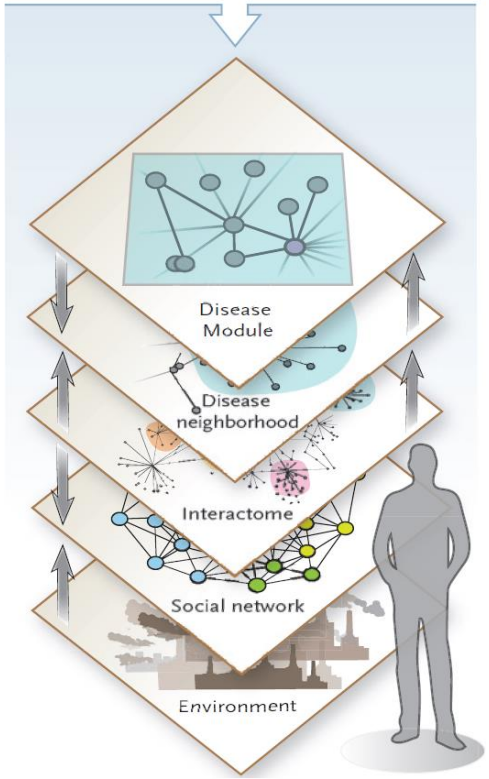
Disclosures

- UCLouvain is part of the Network Medicine Alliance and Institute; *https://www.network-medicine.org/*
- No other relevant C.I.

From reductionism…



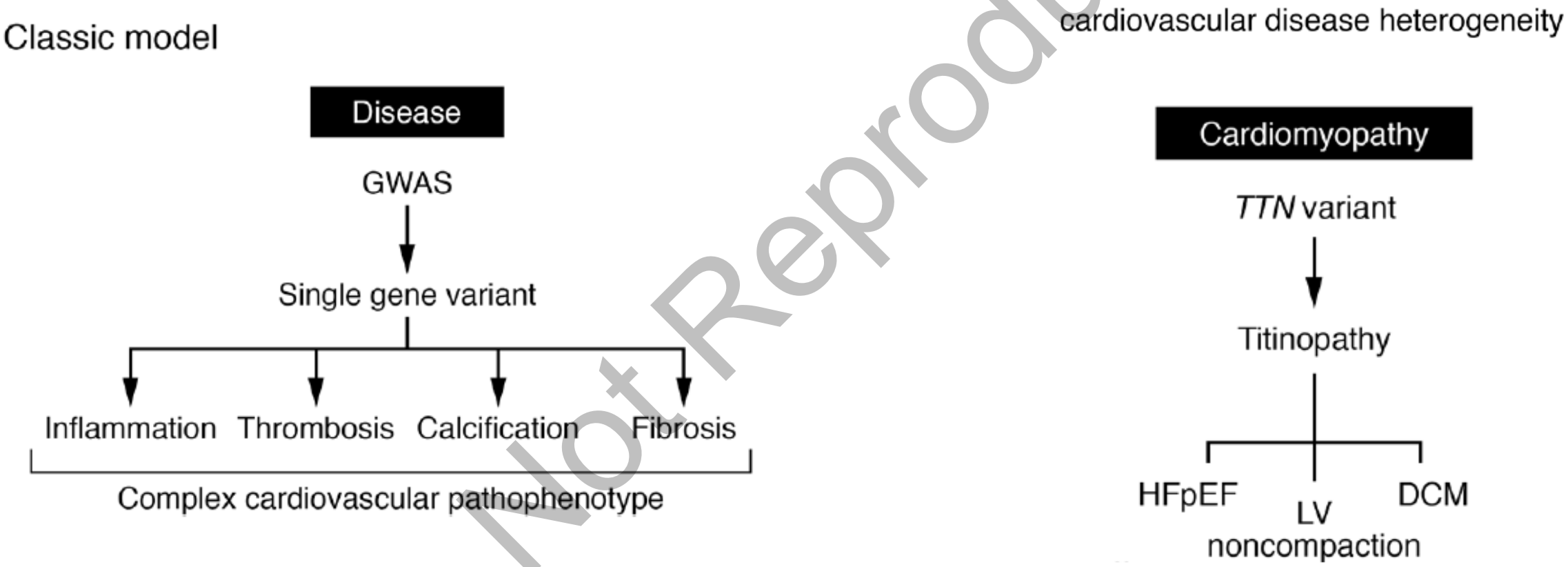




…to post-genomic holism

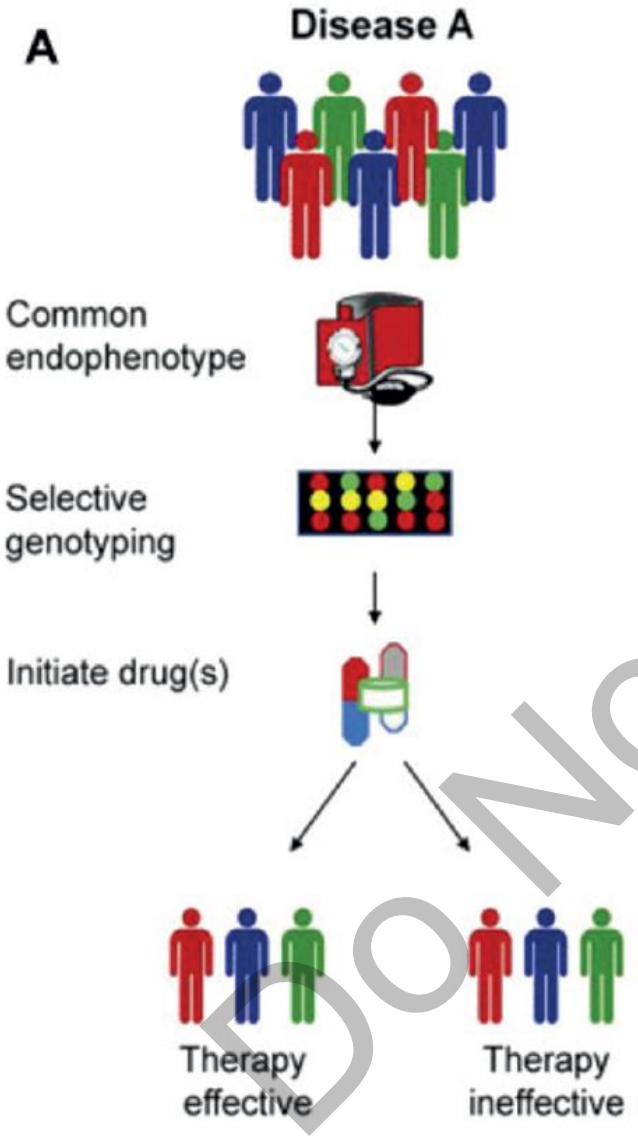
From Greene JA et al. *N Engl J Med* 2017

Limitations of the genetic approaches



From Leopold J et al . *J Clin Invest* 2020

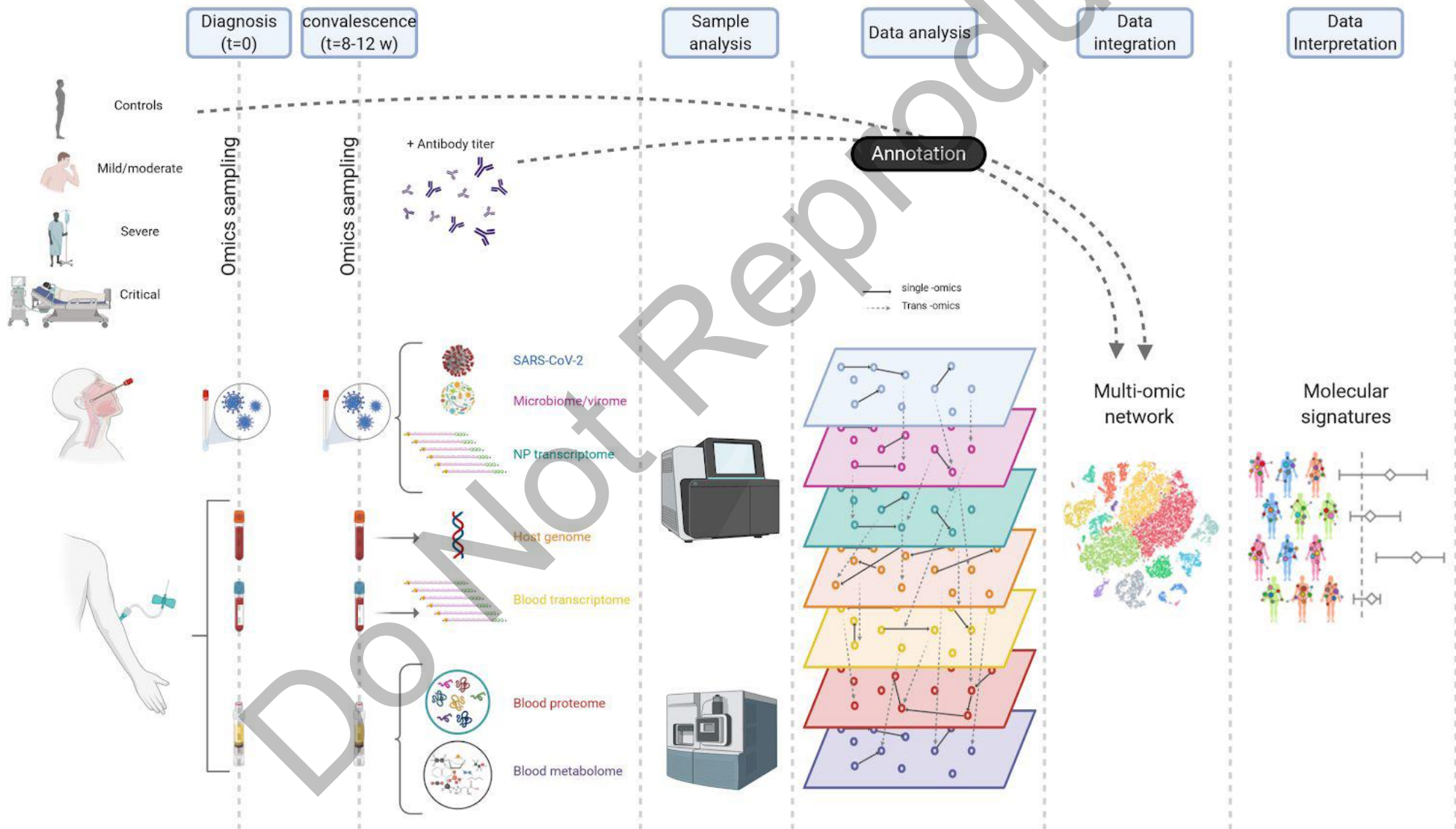
The reductionist approach



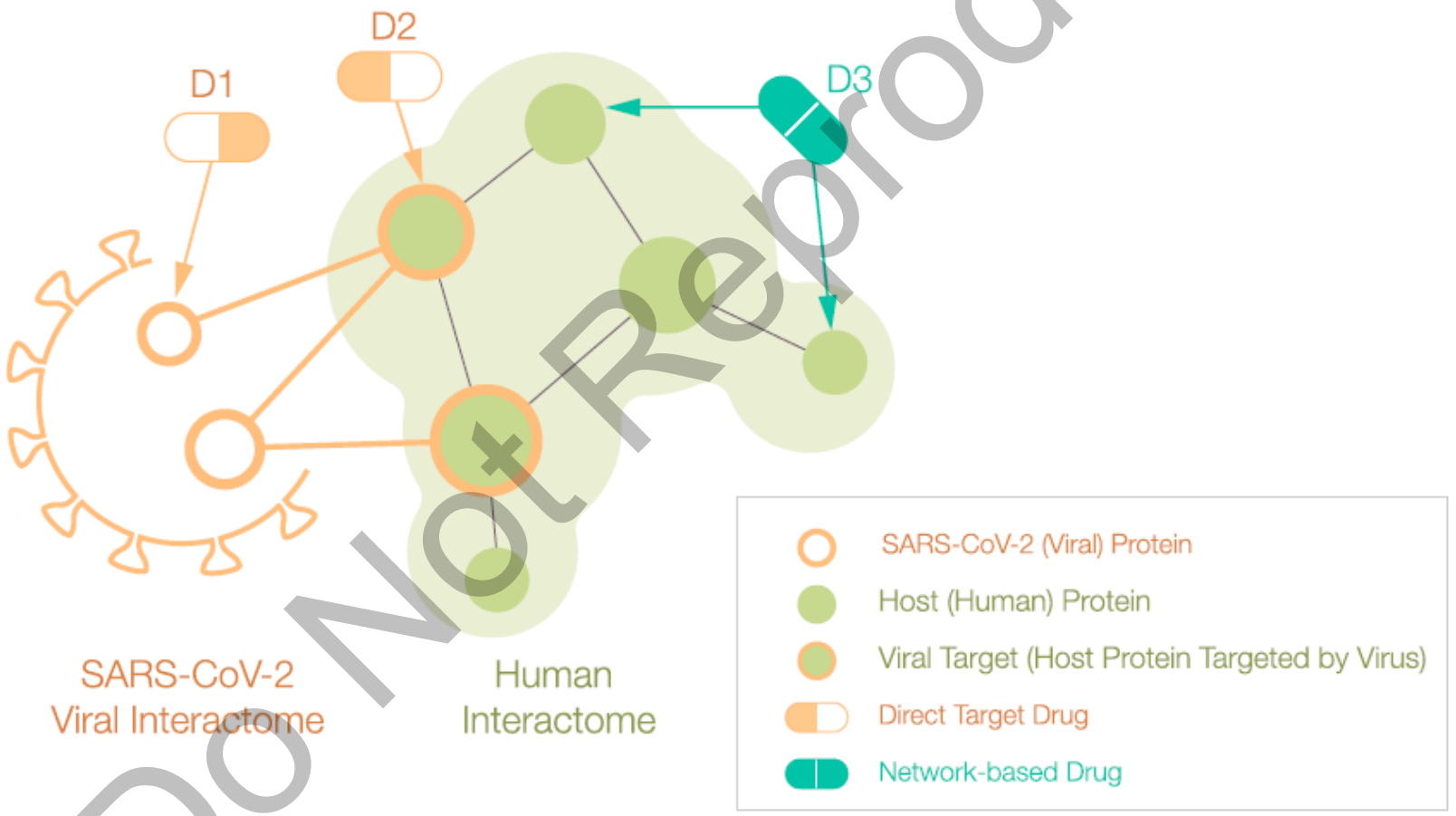
HYGIEIA

Supported by Sofina Covid Solidarity Fund
Impulsed by Fondation Saint-Luc

UCLouvain
BRUXELLES

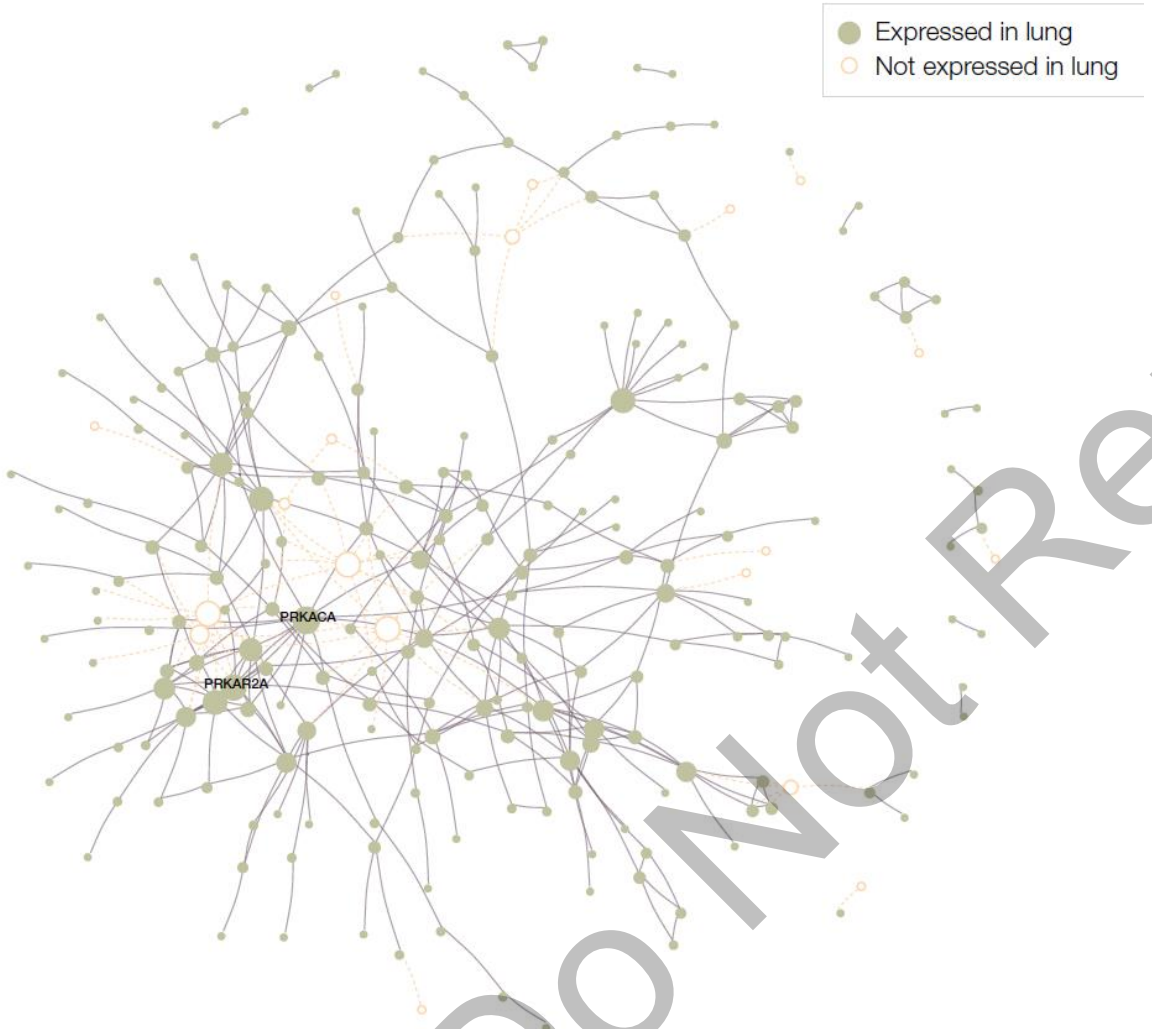


Network-based drug repurposing: COVID19 disease module and interactions (*Proximity hypothesis*)



From Gysi et al. *Proc Natl Acad Sci* 2021

COVID19 disease module



332 human proteins to which 26 SARS-coV2 proteins bind

Repurposing candidate drugs are selected if they bind target proteins in the network vicinity of the disease module

Top-ranked drugs screened in human cell lines: 62% success rate (vs. 0.8% hit rate from non-guided screens)

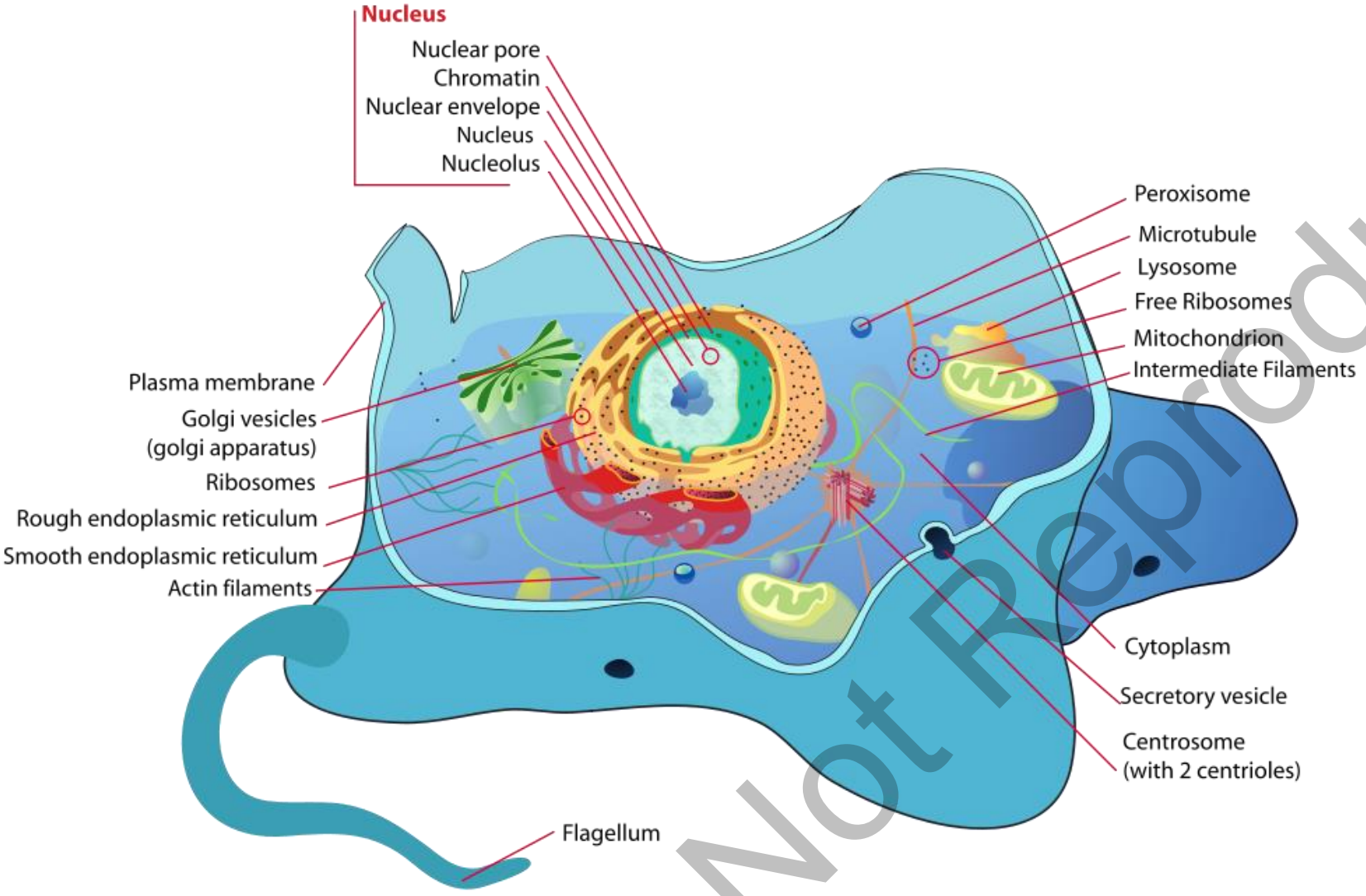
! Almost all successful drugs do NOT bind human proteins directly targeted by SARS-CoV-2 and would not have been predicted from docking-based strategies

= « Network drugs »

From Gysi et al. *Proc Natl Acad Sci* 2021

Some caveats

- Quality of data, standardization; avoid «batch » effect
- Quantity of data, statistical power
- Need critical appraisal of input data, e.g. many PPI databases are polluted by « noise » (e.g. if constructed from literature data)
- PPI often neglect post-translational modifications and subcellular locations



Courtesy of L. Gatto, *UCLouvain*

Spatial proteomics is the systematic study of protein localisations.

Localisation is function: Localisation and sequestration of proteins within sub-cellular niches is a fundamental mechanism for the post-translational regulation of protein function.

Integration of PPI and localisation data: interactions between proteins requires compatible locations.

Other caveats

- Need to integrate broader context, e.g. ethnicity, « exposome »
- Whole organ/tissue data vs. Single cell data ?
- Normal controls ?

Potential solutions and future directions (1)

- Refine data with more biospecimens as good surrogates for organ/tissues, e.g. liquid biopsies
- Obtain control specimens
- Integrate epigenome (incl. non-coding RNA), PTMs and sub-cellular locations in PPI networks

Potential solutions and future directions (2)

- Integrate metabolome (as elements of communication between networks)
- Take microbiome metabolism into account
- Time trajectories (dynamic networks)

Potential solutions and future directions (3)

- Mining the electronic health records (EHR) (real world data)
- Interoperability of IT systems
- Accommodation of data protection regulations ?
- Federated analysis

UCLouvain

icteam

Petabytes of unexploited data in hospitals

Could be used for research or to improve clinical workflow

Courtesy of S. Jodogne, *UCLouvain*

UCLouvain

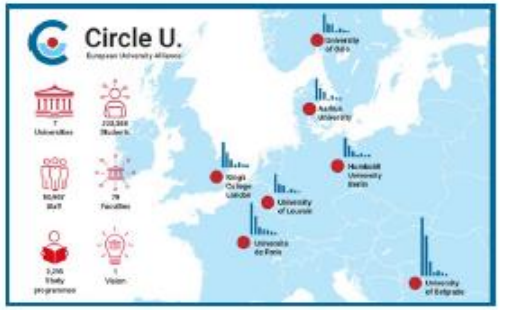
icteam

This is not only about EHR!

- **Task-specific servers**: Medical imaging, operating rooms, pathology, dentistry, medical laboratories…

- **Ad-hoc files**: Excel on shared network drives or e-mails, ZIP on Dropbox, raw text on my own computer…

- **Multi-centric studies**: How can I work on data from various sources? How can I access Réseau Santé Wallon or Abrumet?

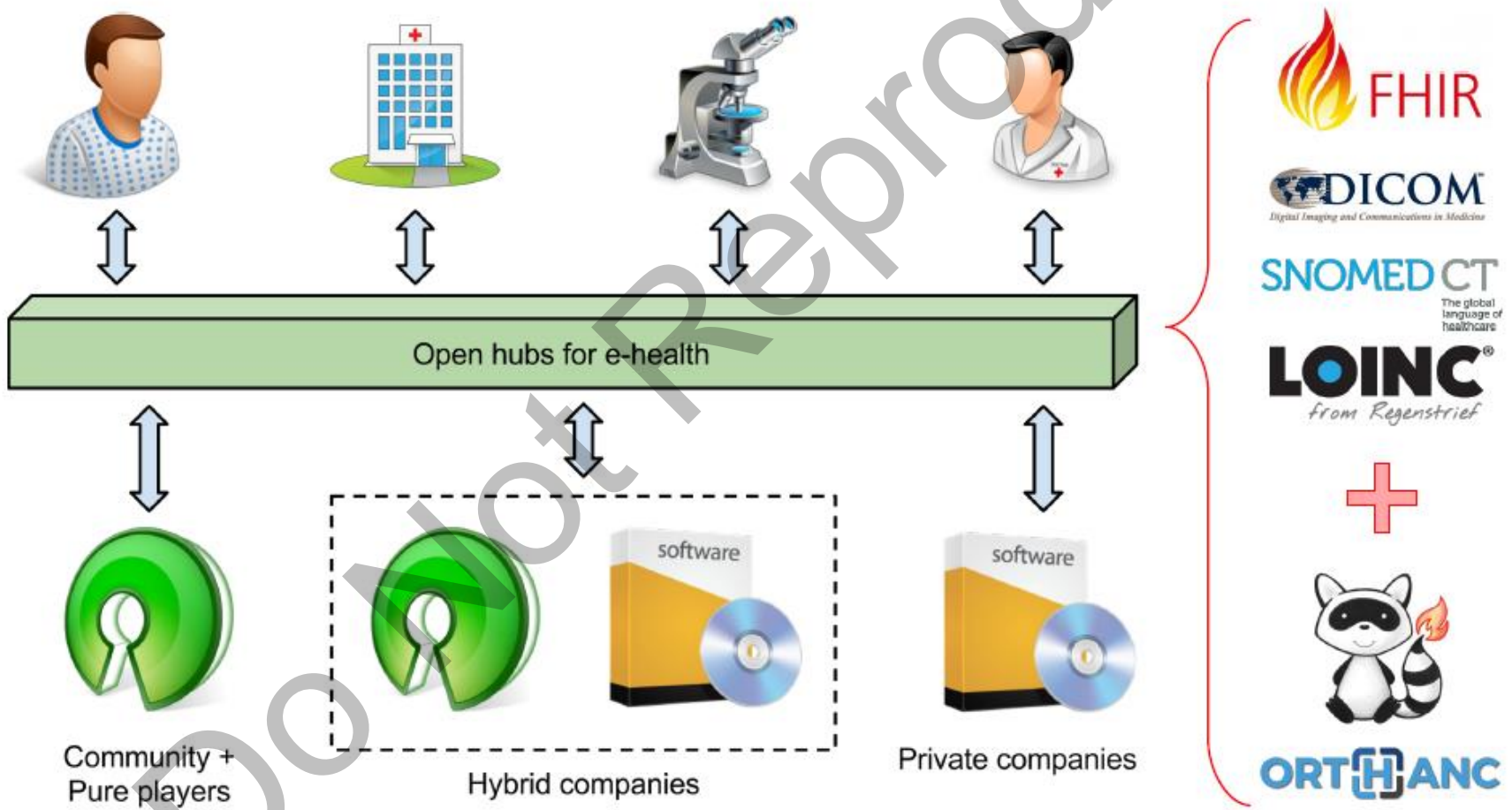


Courtesy of S. Jodogne, *UCLouvain*

UCLouvain

icteam

Big picture – Opening health ecosystems

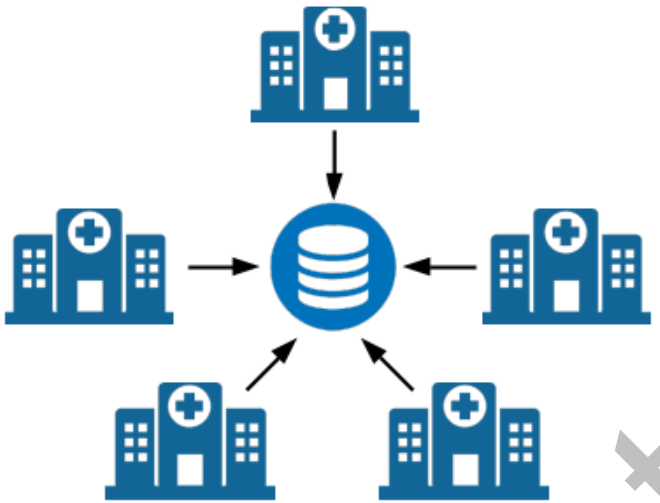


Courtesy of S. Jodogne, *UCLouvain*

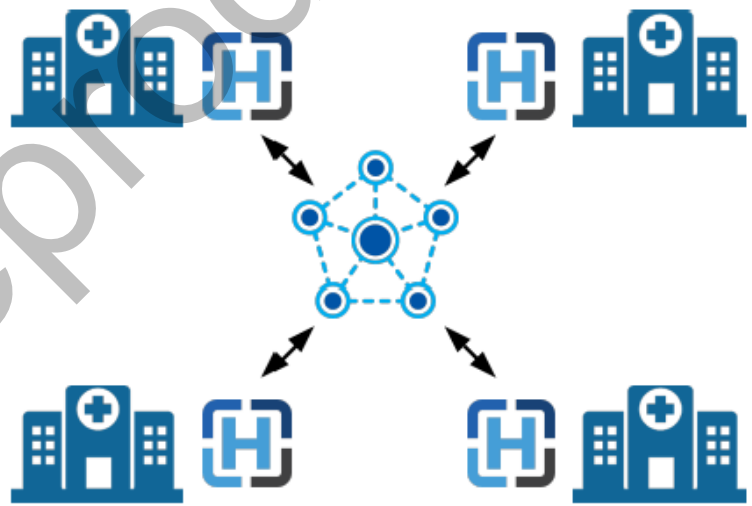
UCLouvain

icteam

Multi-centric studies → Federated learning?



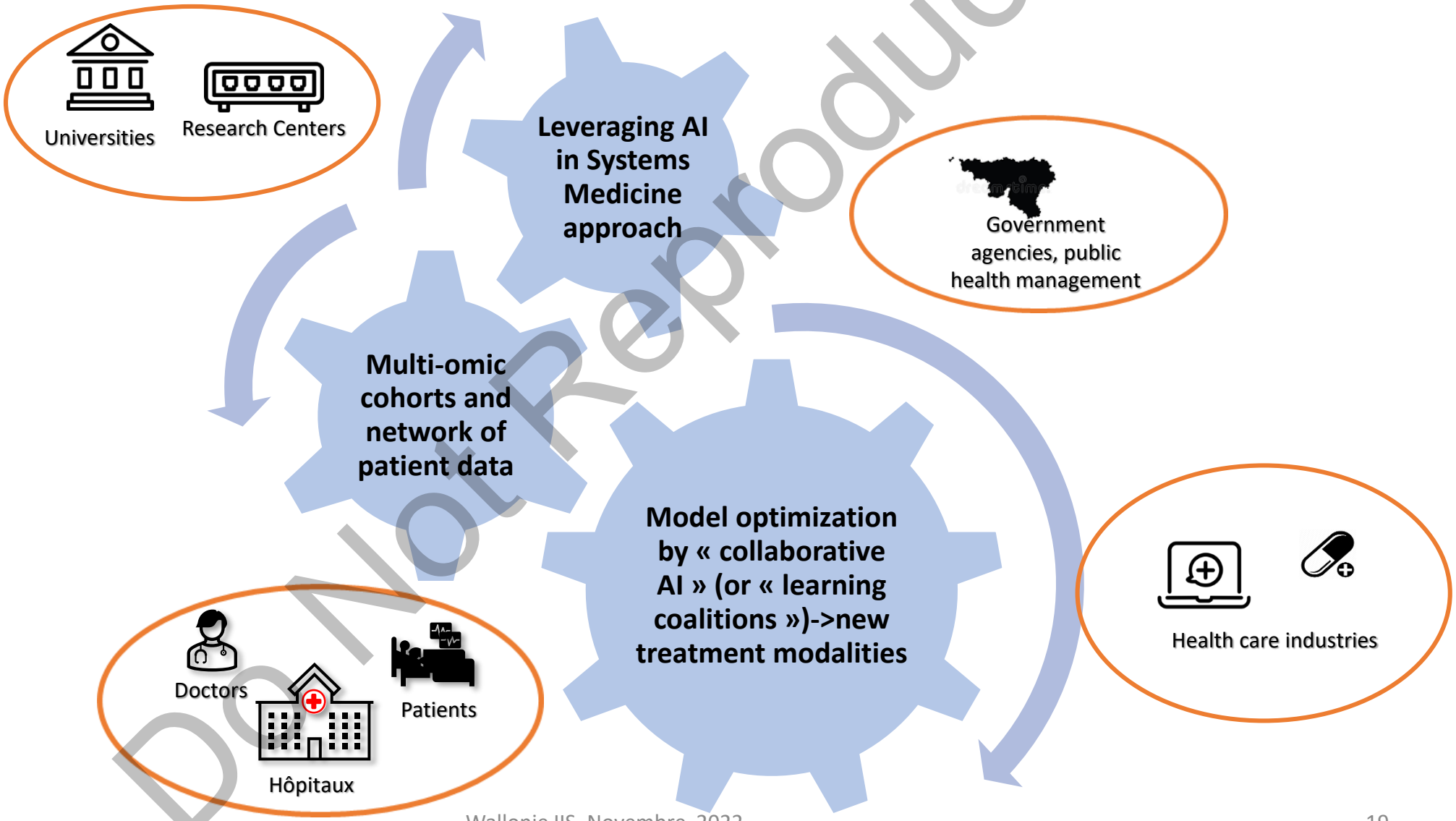
GDPR? Privacy?
Re-identification?



Only exchange the **weights** of the model, not the data

Courtesy of S. Jodogne, *UCLouvain*

MedReSyst: implementation of Systems Medicine in Wallonia



Acknowledgments

Laure Elens

Laurent Gatto

Jean-Cyr Yombi

Leila Belkhir

Julien Degreef

Dimitri Vanderlinden

Jean-François Collet

Didier vertommen

Benoit Kabamba

Joseph Dewulf

Vincent Haufroid

Sebastien Jodogne

Benoit Macq

UCLouvain

icteam

Is there an open interoperability standard for EHR?

- "Fast Healthcare Interoperability Resources"
- Work by HL7 (the same consortium than HL7v2 and HL7v3)
- First draft in 2011, first normative version in October 2019
- Developer-friendly (REST API, JSON...)
- Belgium e-Health: *"In support of 'Plan d'actions e-Santé 2019-2021' the HL7 FHIR standard is the preferred standard to use."*

Courtesy of S. Jodogne, *UCLouvain*

UCLouvain

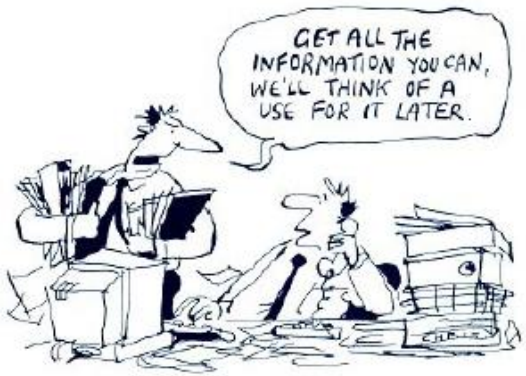
icteam

Machine learning requires data

Often AI researchers start with open data from challenges

But AI for clinical research need:

- A technical, **formal description** of the clinical problem
- High-quality, **real-world "input"** data extracted from the hospital information systems
- **Expected "output" labels** as defined by the clinical team
- Possible deidentification (**anonymization**) of data if AI research is done out of the hospital
- Deep learning requires *much* data



Integration of the epigenome in Network Medicine

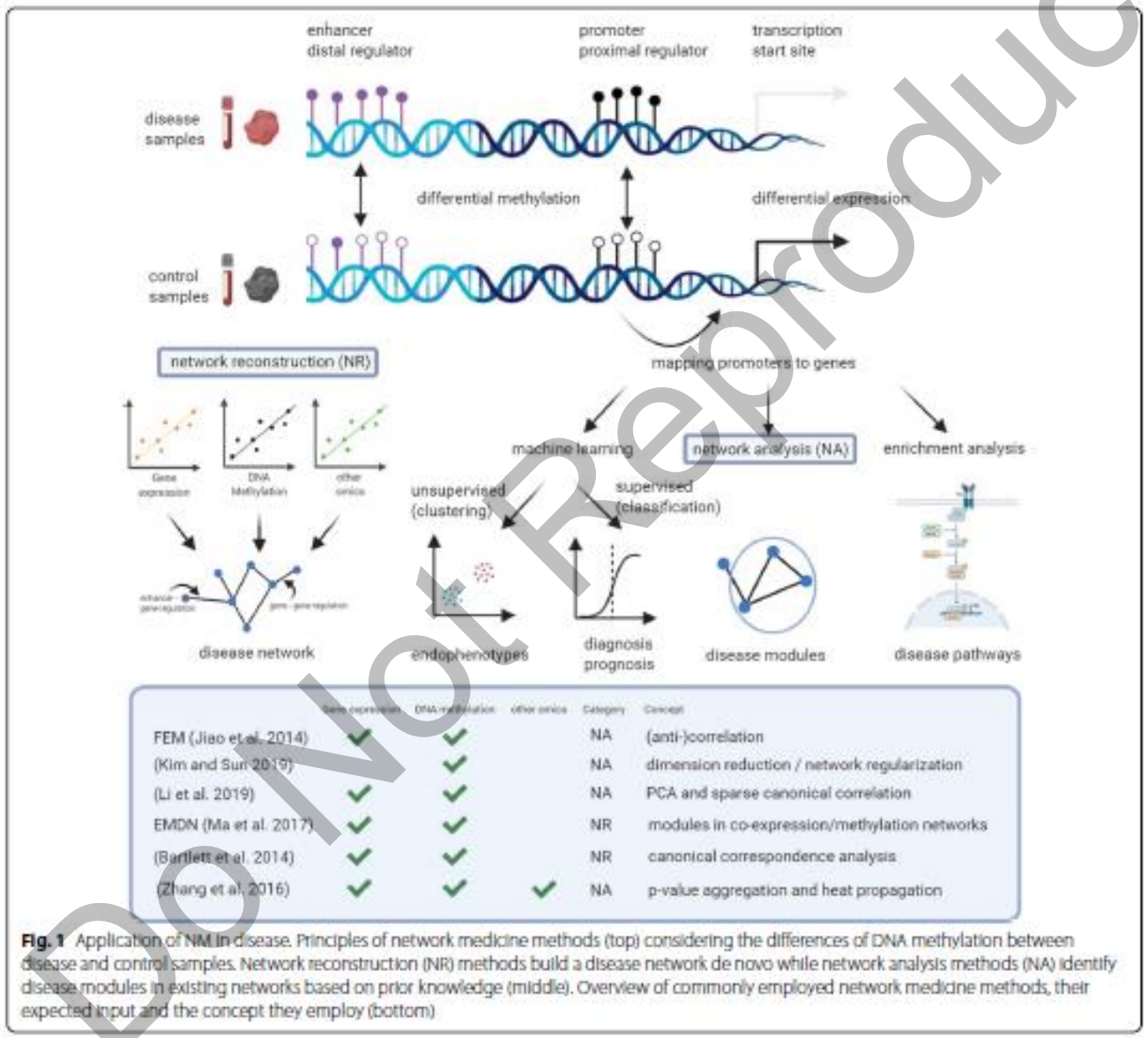


Fig. 1 Application of NM in disease. Principles of network medicine methods (top) considering the differences of DNA methylation between disease and control samples. Network reconstruction (NR) methods build a disease network de novo while network analysis methods (NA) identify disease modules in existing networks based on prior knowledge (middle). Overview of commonly employed network medicine methods, their expected input and the concept they employ (bottom)

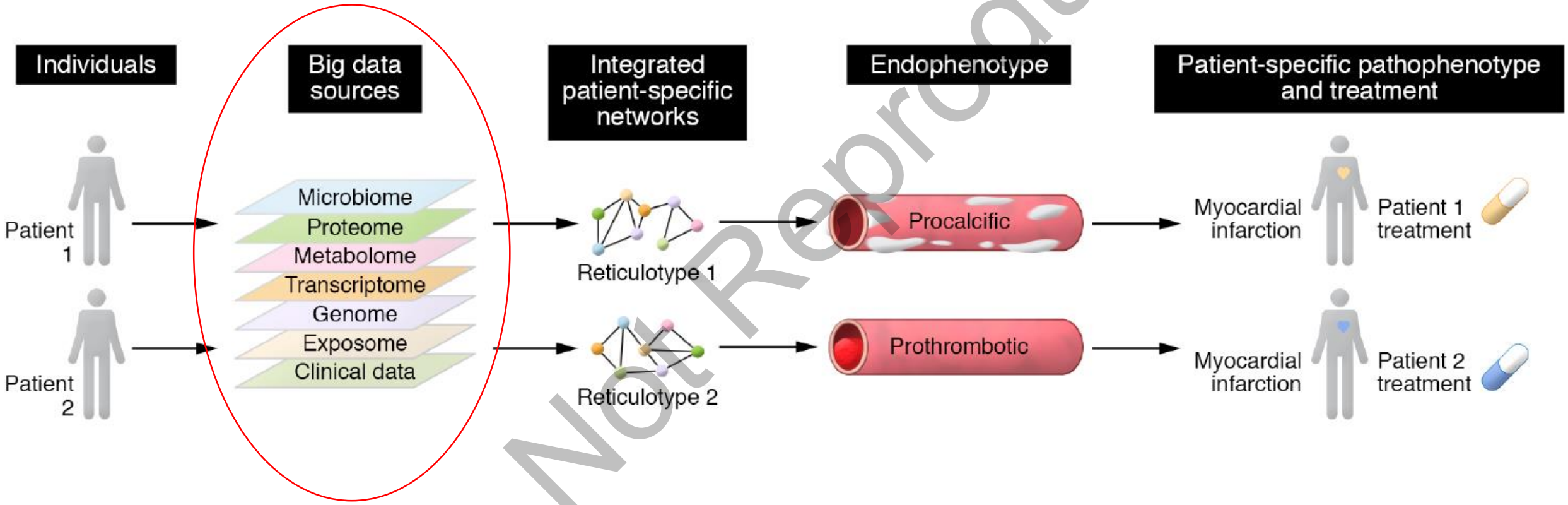
From Sarno et al. *Clin Epigenetics* 2021

Disease modules in the interactome

Curation of all validated PPI in the human cell that form the human interactome

Binary (Y2H)
Regulatory (TRANSFAC)
Metabolic (CORUM)
Kinase and Signaling Networks
Literature surveys (IntAct, MIND, BioGrid, HPRD)

Challenges and future directions



From Leopold J et al . *J Clin Invest* 2020

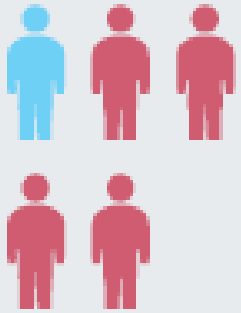
Using PPI networks to facilitate drug target identification and guide drug repurposing

(beyond the « magic bullet » theory)

IMPRECISION MEDICINE

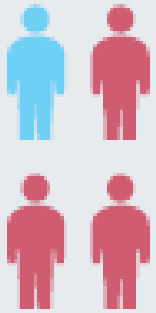
For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

1. ABILIFY (aripiprazole)
Schizophrenia



2. NEXIUM (esomeprazole)
Heartburn

3. HUMIRA (adalimumab)
Arthritis



4. CRESTOR (rosuvastatin)
High cholesterol

From NJ Schork, *Nature* 2015

Drugs have many targets

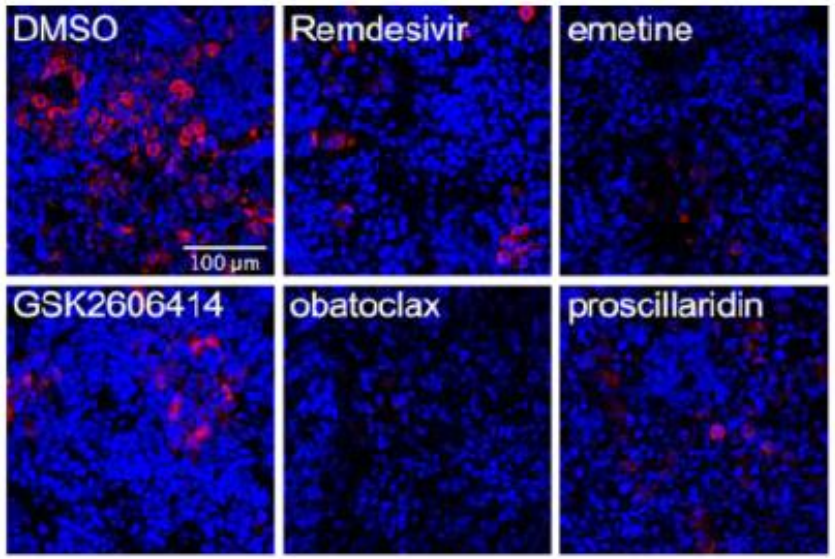
- From 3-5 to a thousand different targets have been assigned to individual drugs (average: 32.4 targets/drug among the >470 FDA-approved drugs (from Chatrier et al. *BMC Pharmacol Toxicol* 2017)

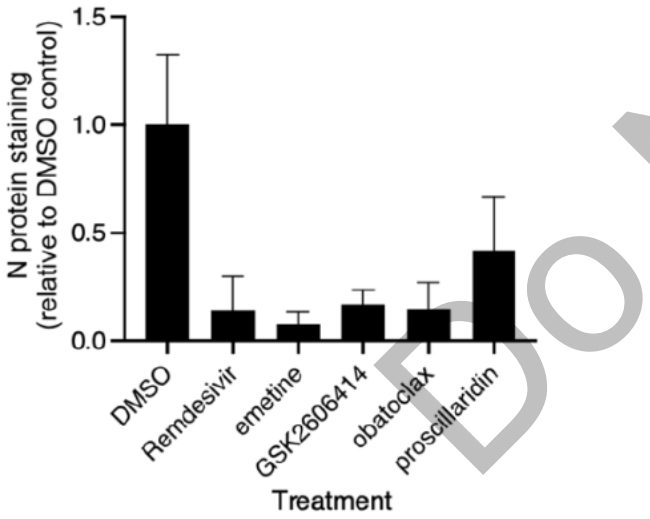
- This is a plausible explanation for frequent side-effects

- But also an opportunity for drug re-purposing

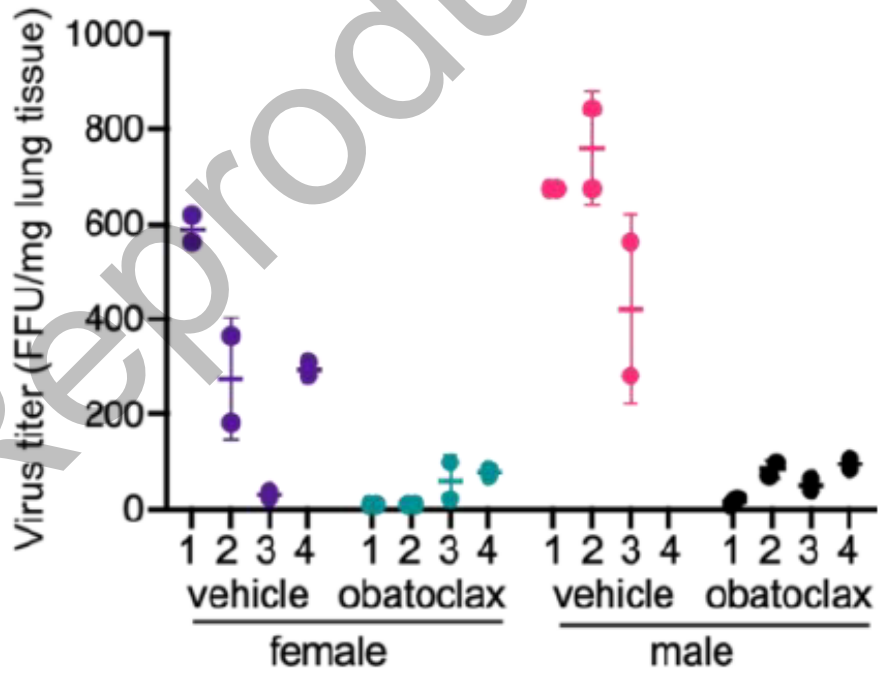
Drug repurposing screen of 6,710 compounds (Broad library): *in vitro* and *in vivo* validation

Primary epithelial cells





hACE2-overexpressing mice



From Patten et al. *BioRxiv* 2021
(Courtesy of J. Loscalzo)

Protein localisation data:

- ➔ Targeted microscopy-based (**Human protein Atlas**): 2,390 proteins across 32 subcellular locations.
- ➔ Global protein localisation map (**MS-based spatial proteomics**)
- ➔ Annotations (**Gene Ontology**)

Protein-protein interactions:

- ➔ **STRING** database of known and predicted protein-protein direct (physical) and indirect (functional) interactions. 5,879,727 pairwise interactions across 19,354 unique proteins.
- ➔ **Bioplex** (large experimental resource): 118,162 pairwise interactions between 13,689 unique proteins.

Applications:

- ➔ **Transfer learning**: to inform one with the other
- ➔ **Deep learning**: large scale integration

Courtesy of L. Gatto, *UCLouvain*

Interactome (PPI) hypothesis: do disease genes (gene products) cluster in discrete modules in the interactome ?

Principles linking the interactome to human disease:

-local hypothesis: proteins involved in the same disease tend to interact

-disease module hypothesis: proteins involved in the same disease tend to cluster in connected subnetworks (*disease module*)

-Functional coherence hypothesis: proteins in a disease module are often involved in the same biological process

-Shared components hypothesis: related diseases are located in the same interactome neighborhood from which unrelated diseases are separated

Paci et al. *NPJ Syst Biol Appl* 2021
Courtesy of J. Loscalzo

Different people

UCLouvain

icteam

AI for healthcare requires multidisciplinary approach

Algorithms irrelevant to clinical practice

Unrealistic idea on the technical point of view

Cross-fertilization

UCLouvain

icteam

Possible solutions

- **Clearly state the needs**: Think using the *input/output* model (what data is available, where is it stored, what answer do I want?)
- **Teamwork**: Hire *actual* software engineers within clinical departments (not just for helpdesk or computer configuration)
- **Research projects**: Plan budget for software engineer (cf. data managers)
- **"Living labs"**: Create places where engineers and physicians can exchange
- **Education**: Nowadays, software engineers "learn by doing" and physicians "learn if geek" → need for *transversal courses*

Interoperability

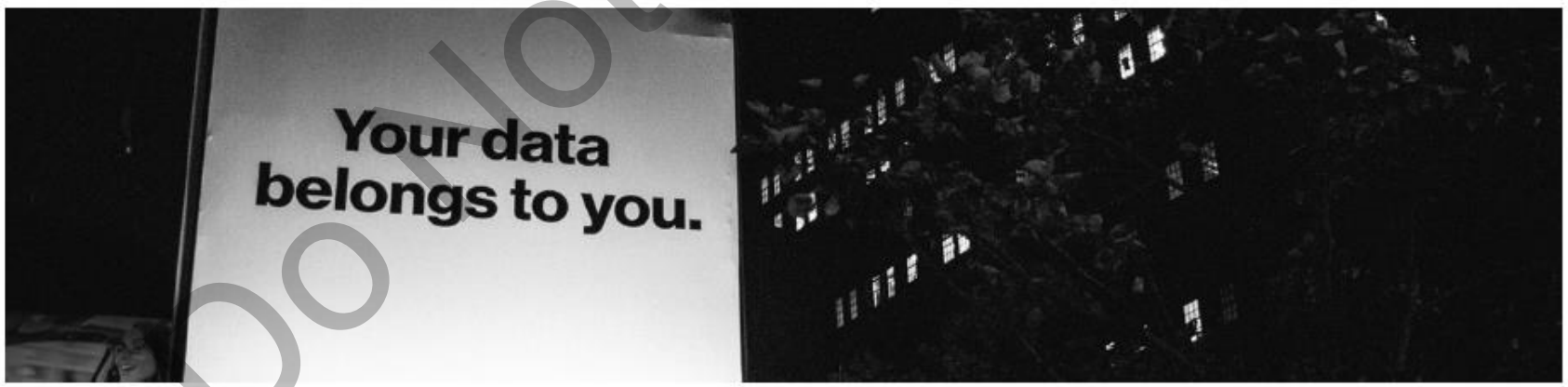
UCLouvain

icteam

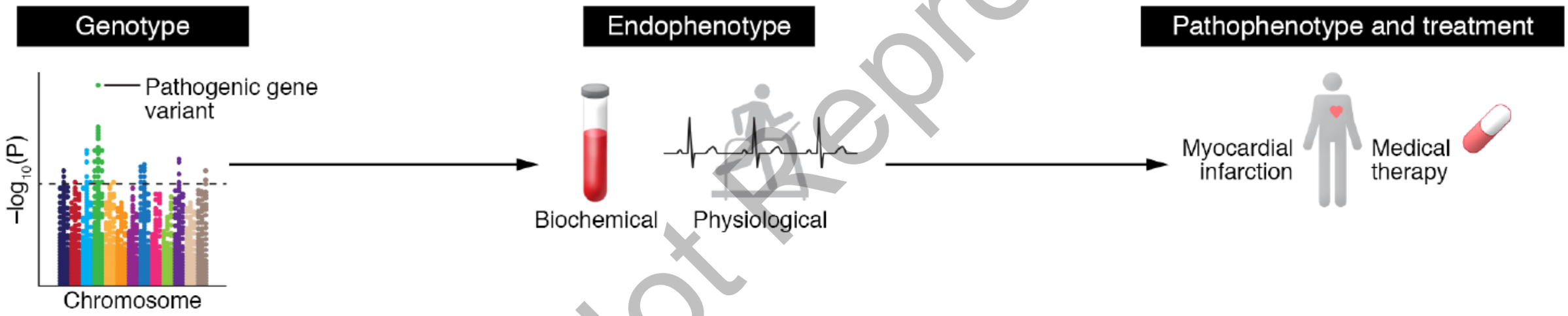
Locked data

Good reasons: Security of clinical data, ethics...

Bad reasons: EHR vendor saying: *"You must pay for me to develop an access"* (**lock-in**)

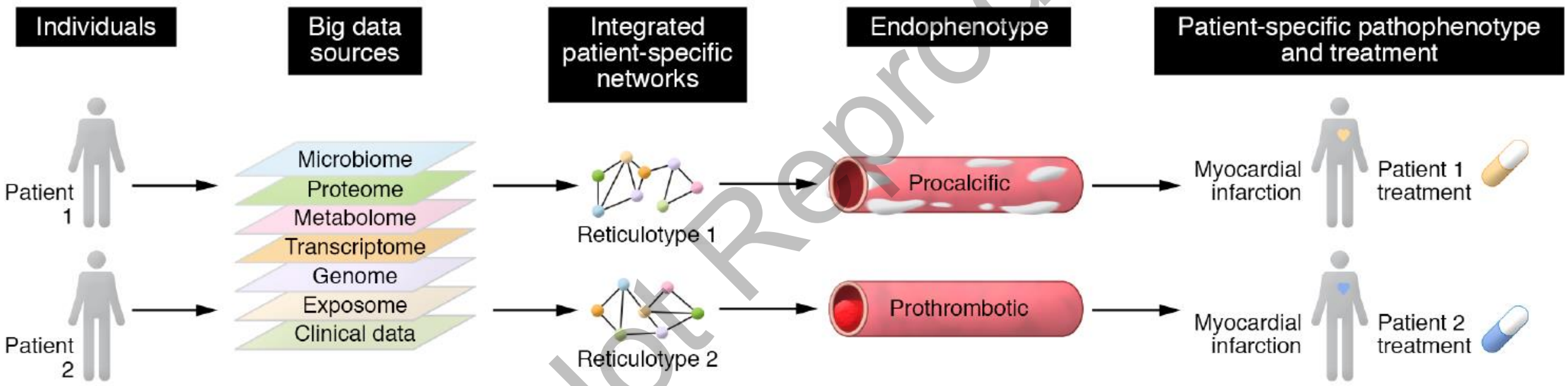


The reductionist approach



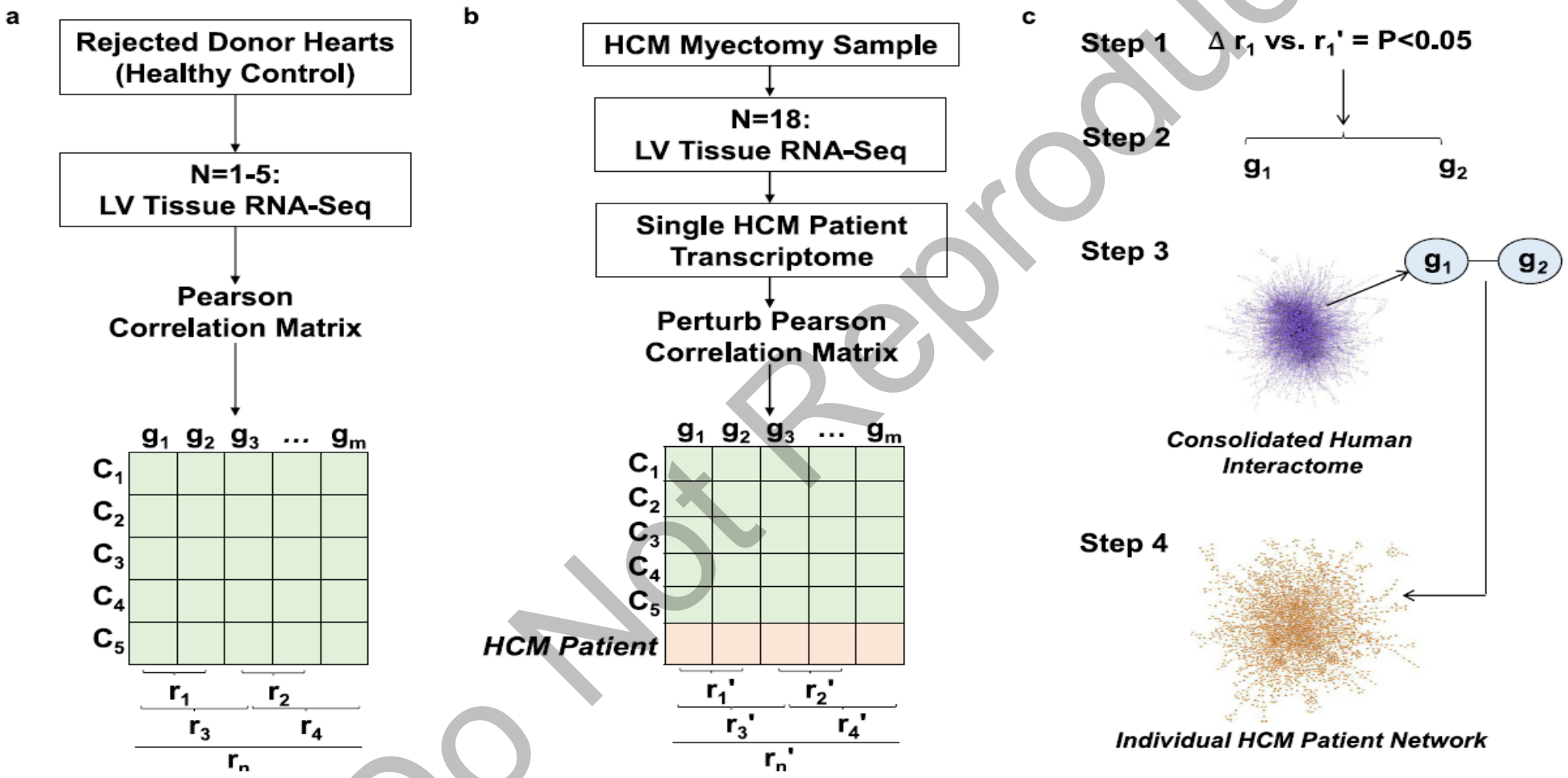
From Leopold J et al . *J Clin Invest* 2020

Network medicine uses « omic » data to define the genetic context that explains individual phenotypes



From Leopold J et al . *J Clin Invest* 2020

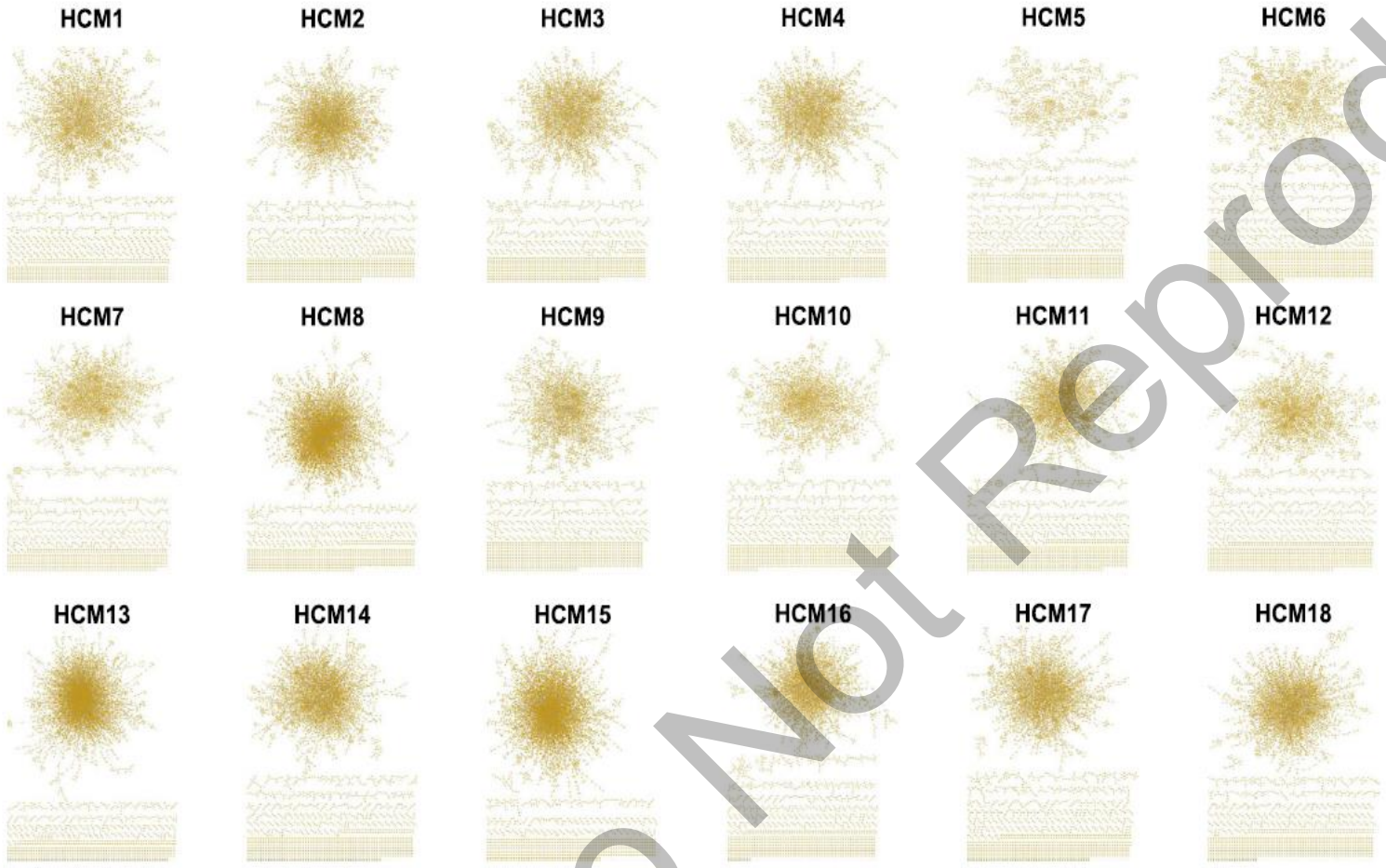
Derivation of individual disease module in hypertrophic cardiomyopathy



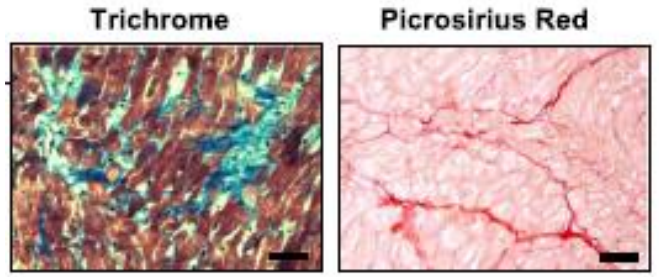
(*pairwise perturbation correlation analysis*)

From Maron et al. *Nat Commun* 2021

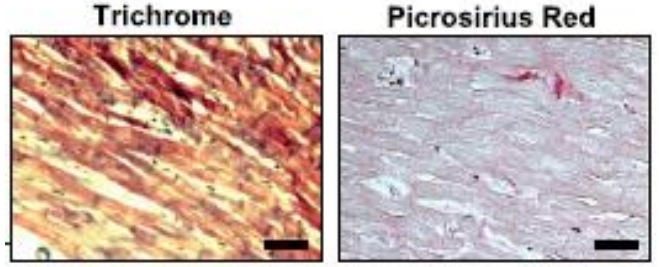
Individual disease modules in HCM inform pathophenotype



HCM8



HCM7



JAK2

From Maron et al. *Nat Commun* 2021