Genetics and Bioinformatics

Kristel Van Steen, PhD[2]

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

ULg

kristel.vansteen@uliege.be

Precision Medicine at the interface of translational science and systems medicine

- Precision Medicine
- The Individual as a System
- Need for a Paradigm Shift
- Take-home messages

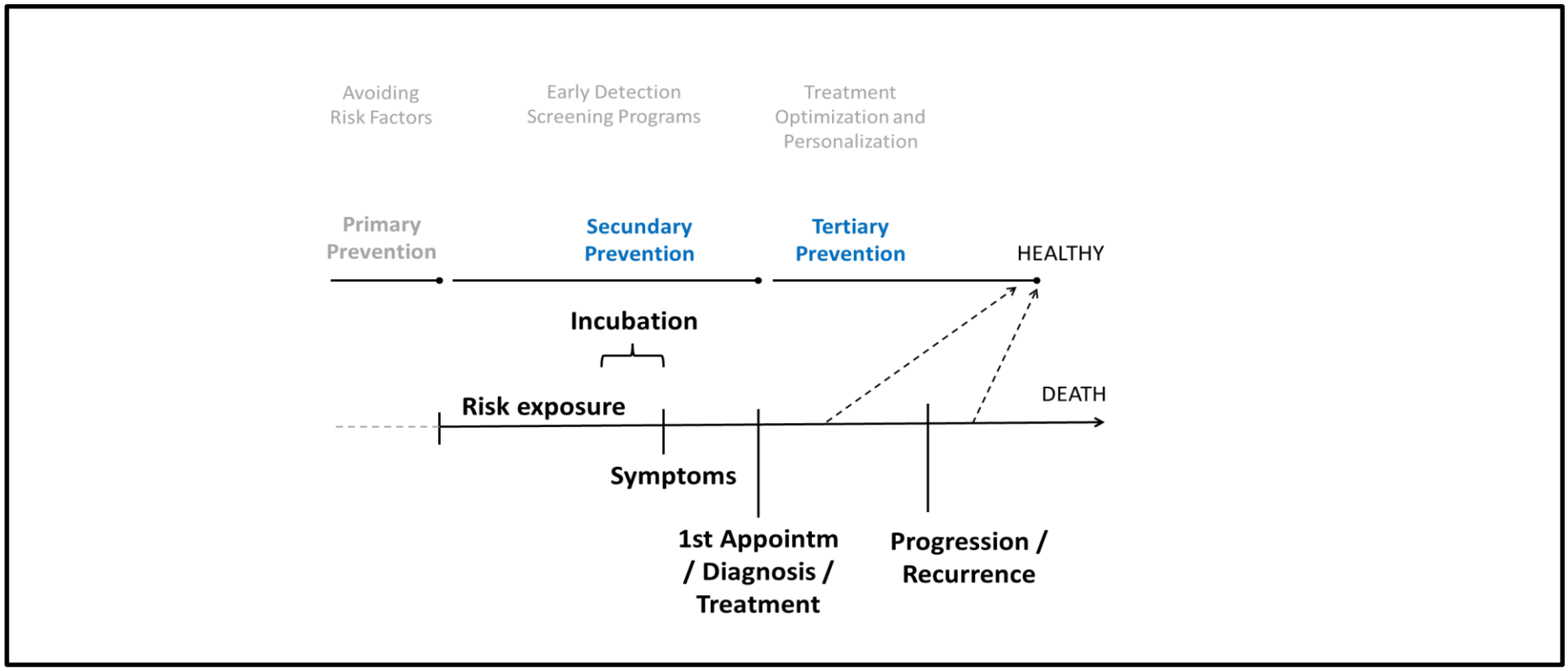
Precision Medicine

Precision Medicine

"a medical model using characterization of individual's phenotypes and genotypes (e.g., molecular profiling, medical imaging, lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely and targeted prevention."

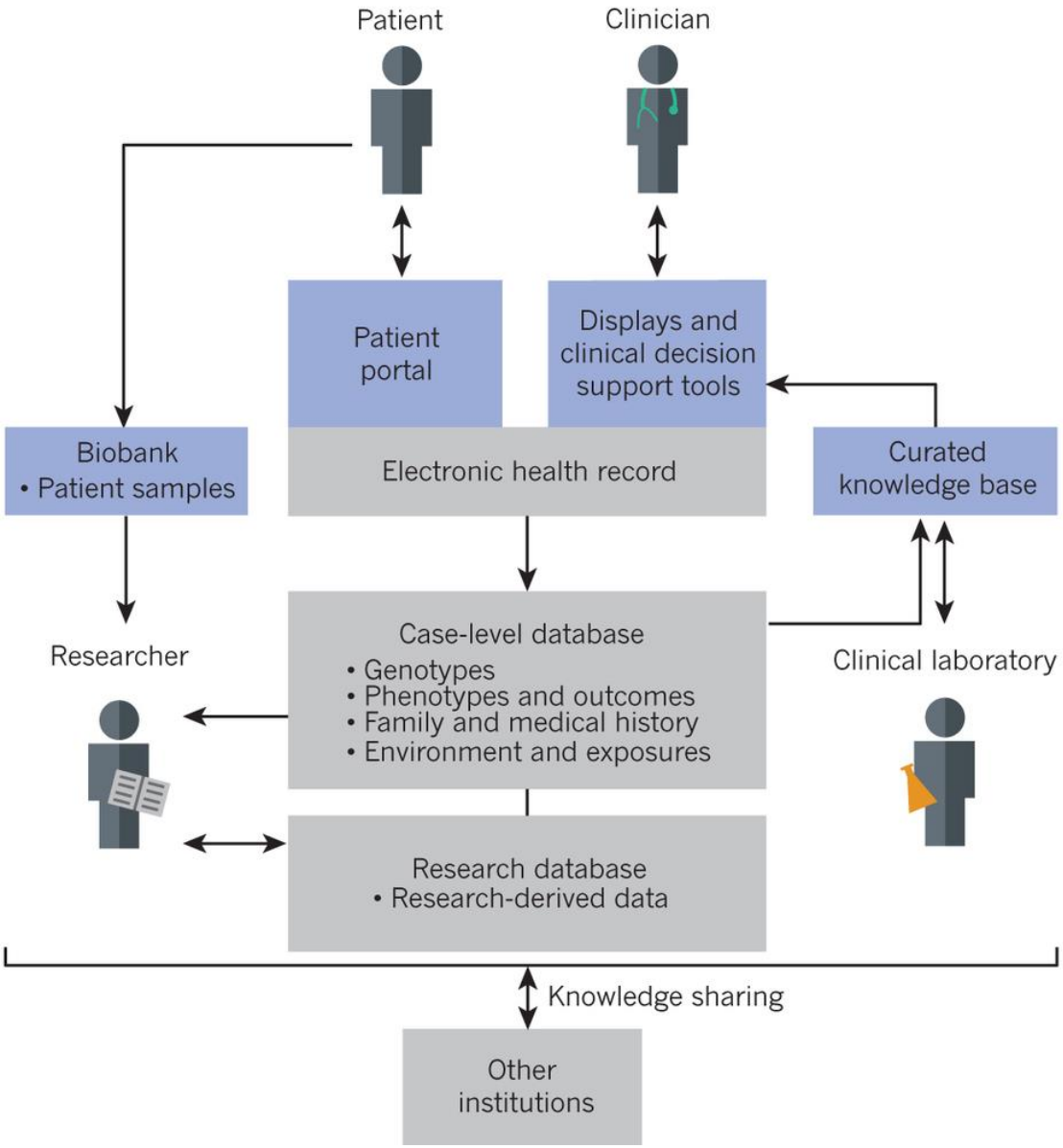
(HORIZON2020 Advisory Group;
EU Health Ministers – December 2015)

PM & Prevention: a dynamical process



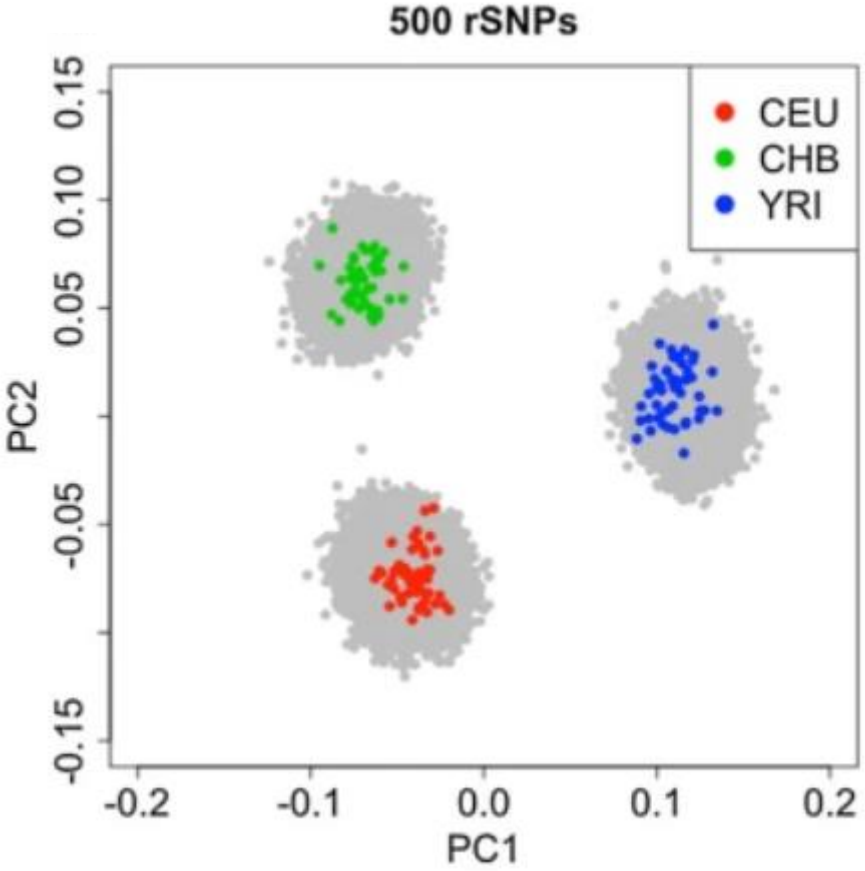
(K Van Steen & N Malats)

PM & Individual Characterization: patient eco-system

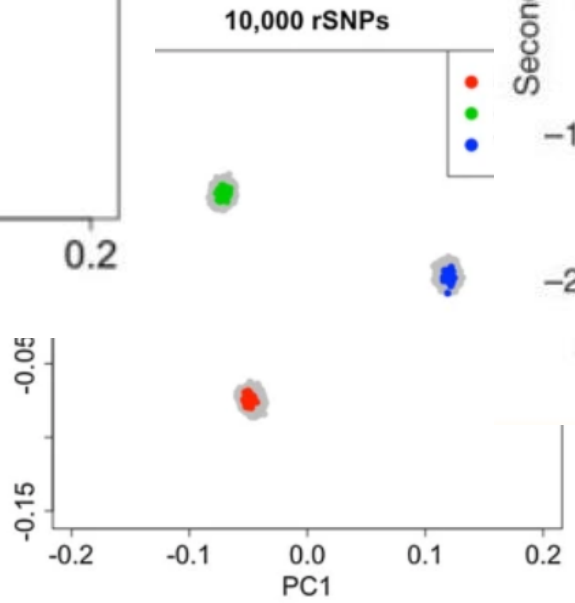


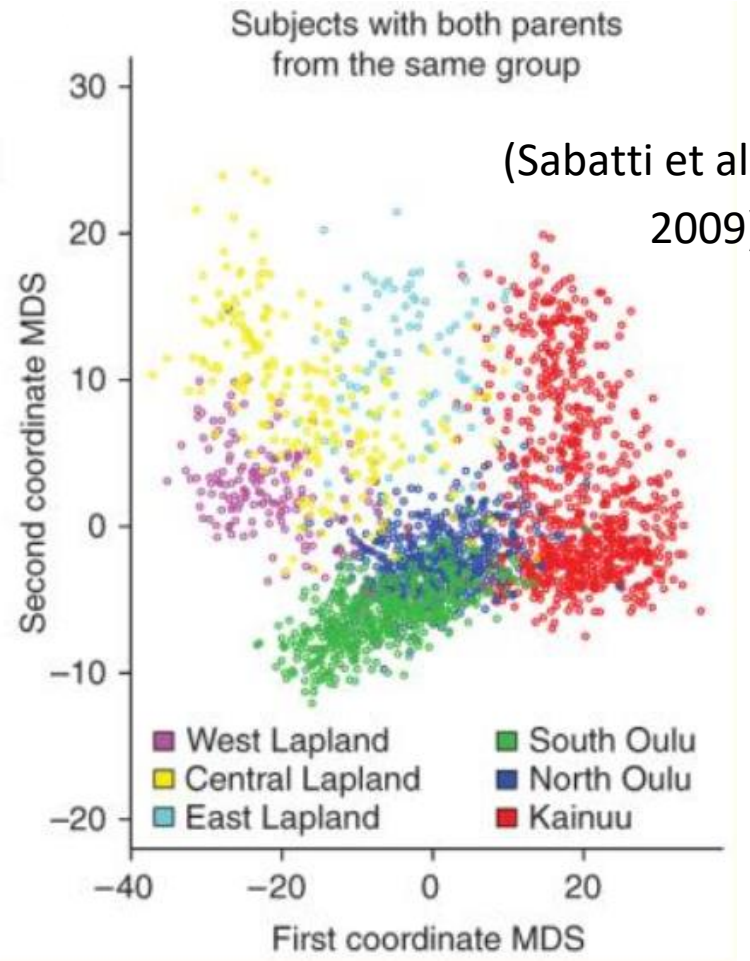
(Aronson & Rehm 2015)

PM & Data acquisition: the more the better?



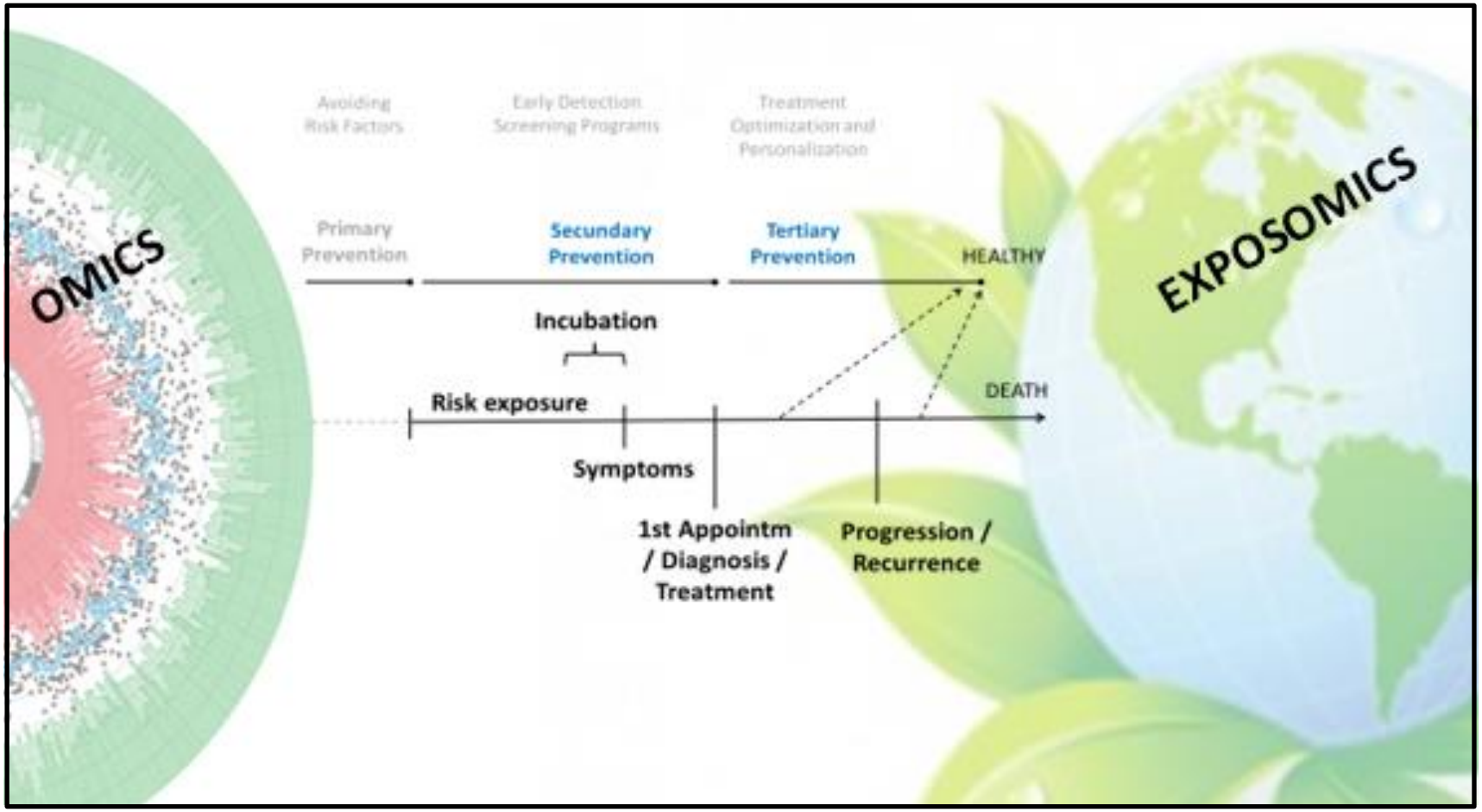
(Pardo-Seco et al. 2014)



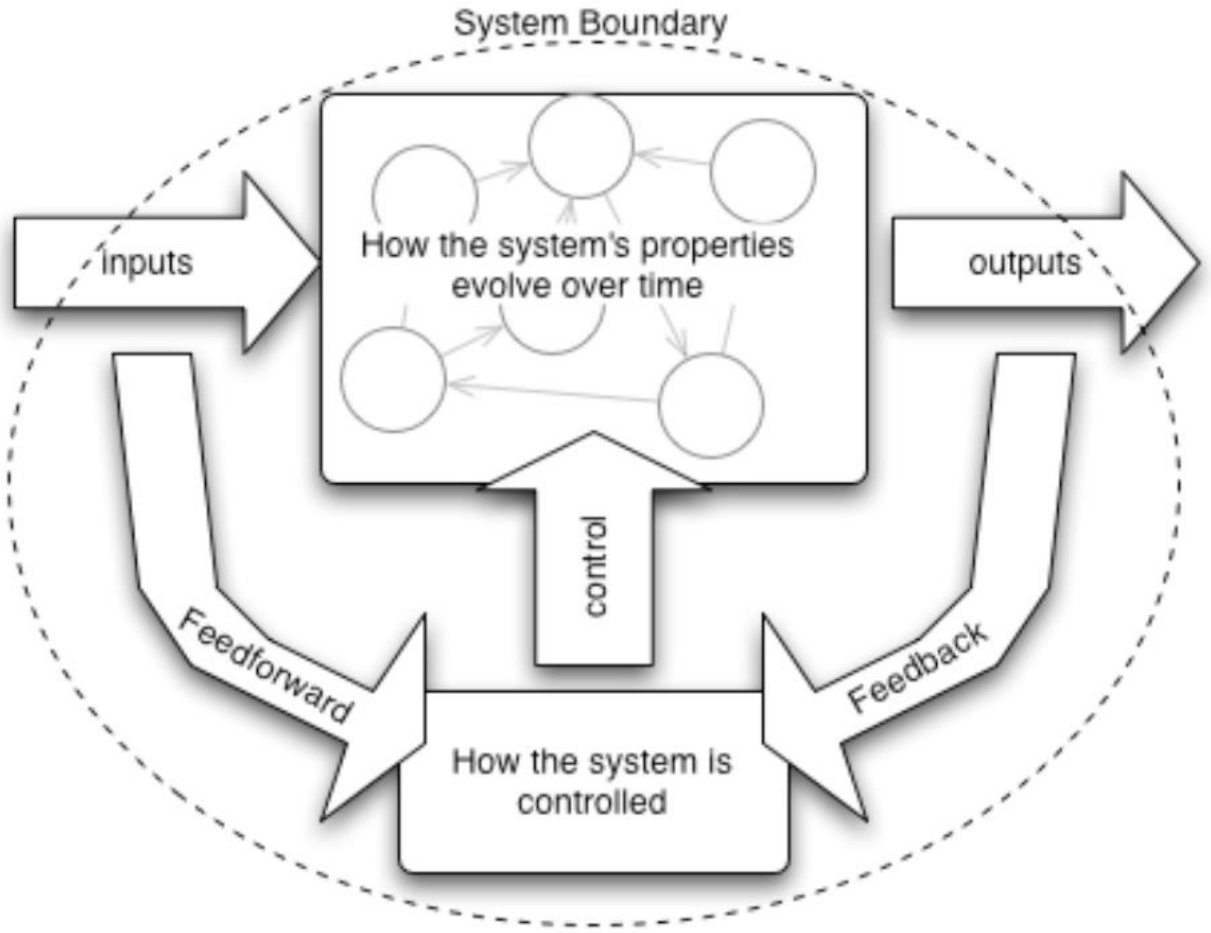


(Sabatti et al. 2009)

Data acquisition: the more the better?



Systems: defining boundaries



(@2004-5 Steve Easterbrook)

PM: the Individual as a System

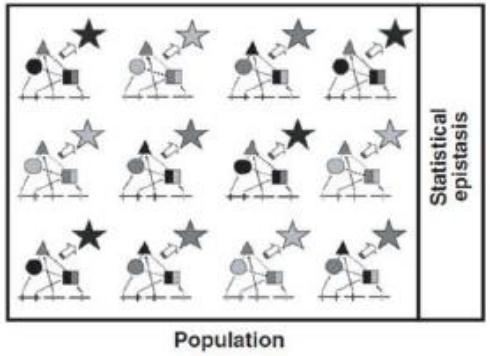
Interactions & Integration

Buzz words: Interaction information, synergy, interactome, …

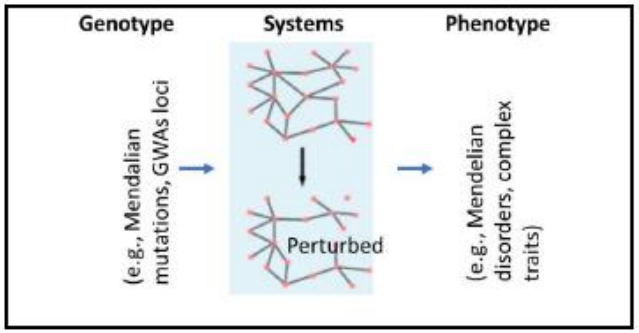
The **interactome** refers to the entire complement of interactions between DNA, RNA, proteins and metabolites within a cell.
These interactions are influenced by genetic alterations and environmental stimuli.
Mind the context!

Perturbations link to individual-to-individual heterogeneity

- Perturbations to the interactome are individual-specific.



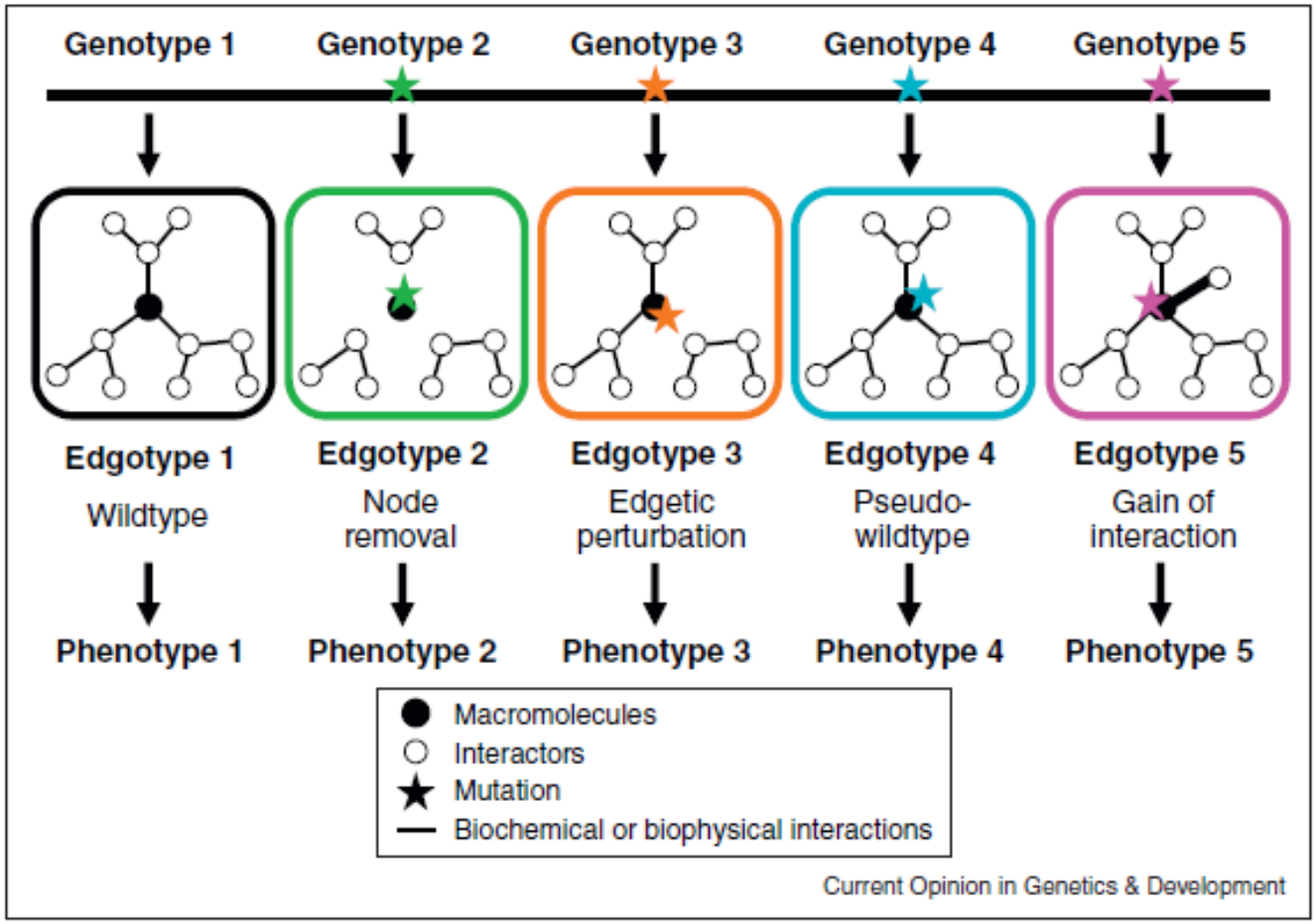
(Moore 2005)



(adapted from Vidal et al. 2011)

- For Precision Medicine, the link between perturbations in network and systems properties and phenotypes might be as important as that between genotypes and phenotypes.

Example 1: perturbations of PPI networks



(Sahni et al. 2013: *edgotypes*)

Example 2: environmental interactions with human SNPs

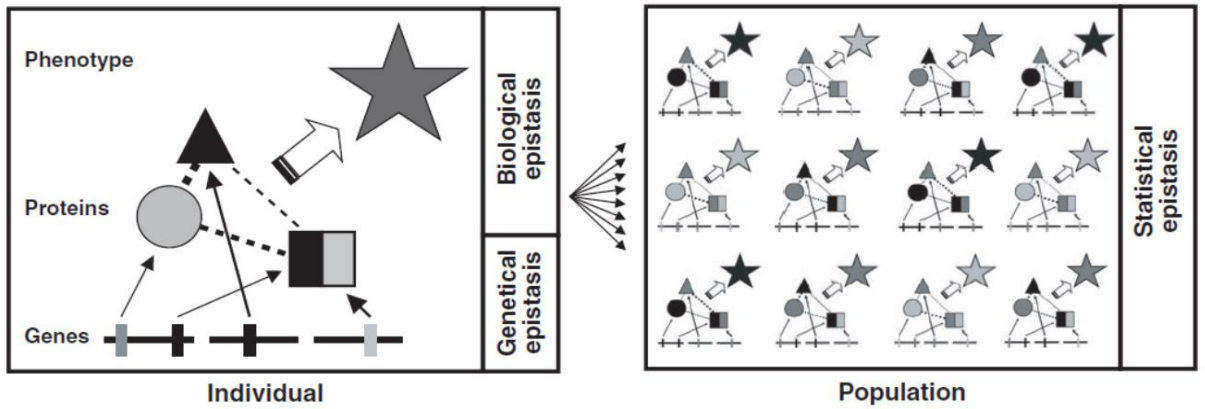
- … where environment may not be predictive for phenotype (Morrison et al. 2007; Dudbridge et al. 2017)
- … where genotype may not be predictive for phenotype (Cooper et al. 2013) – *Incomplete penetrance*, a *common phenomenon*
  - What: The expected phenotype of a particular genotype is not expressed
  - Why: Environmental factors as well as the effect of other genes may alter the phenotypic expression of a particular genotype
- Compensatory genes contribute to variation in natural and human populations; it is intuitive to assume that they should not be a limited phenomenon

Example 3: human SNP-SNP interactions

- Few success stories – define “success”
- A widely accepted protocol to perform a Genome-Wide Association Interaction Study (GWAIS) is still lacking:
  - many difficulties / attention points (technical, statistical, computational, the big picture – data integration) involved in performing large-scale SNP-SNP interaction screening
  - and in inferring biological evidence from statistical findings
- Role in Personalized Medicine: Alzheimer’s disease (**Gusareva et al. 2014**) with a discussion about biological translation and repercussions of the identified *epistasis* in Ebbert et al. 2015.

Epistasis

- The original definition (driven by biology) refers to a variant or allele at one locus preventing the variant at another locus from manifesting its effect (William Bateson 1861-1926).



(Moore 2005)

- Grown into a more **general theory and applications framework** for the analysis of interactions across and between -omics strata.

Lessons learned: moving from DNA networks to interactomes

Human Genetics (2019) 138:293–305
https://doi.org/10.1007/s00439-019-01987-w

REVIEW

How to increase our belief in discovered statistical interactions via large-scale association studies?

K. Van Steen[1,2] · J. H. Moore[3]

Received: 26 July 2018 / Accepted: 20 February 2019 / Published online: 6 March 2019

Abstract
The understanding that differences in biological epistasis may impact disease risk, diagnosis, or disease management stands in wide contrast to the unavailability of widely accepted large-scale epistasis analysis protocols. Several choices in the analysis workflow will impact false-positive and false-negative rates. One of these choices relates to the exploitation of particular modelling or testing strategies. The strengths and limitations of these need to be well understood, as well as the contexts in which these hold. This will contribute to determining the potentially complementary value of epistasis detection workflows and is expected to increase replication success with biological relevance. In this contribution, we take a recently introduced regression-based epistasis detection tool as a leading example to review the key elements that need to be considered to fully appreciate the value of analytical epistasis detection performance assessments. We point out unresolved hurdles and give our perspectives towards overcoming these.

Interactions & Integration

Buzz words: Fusing, merging, combining, …

- **Analytic level:**

  A **trans-disciplinary approach** should provide generic frameworks and should provide organizing principles for the interaction of diff. types of analytics

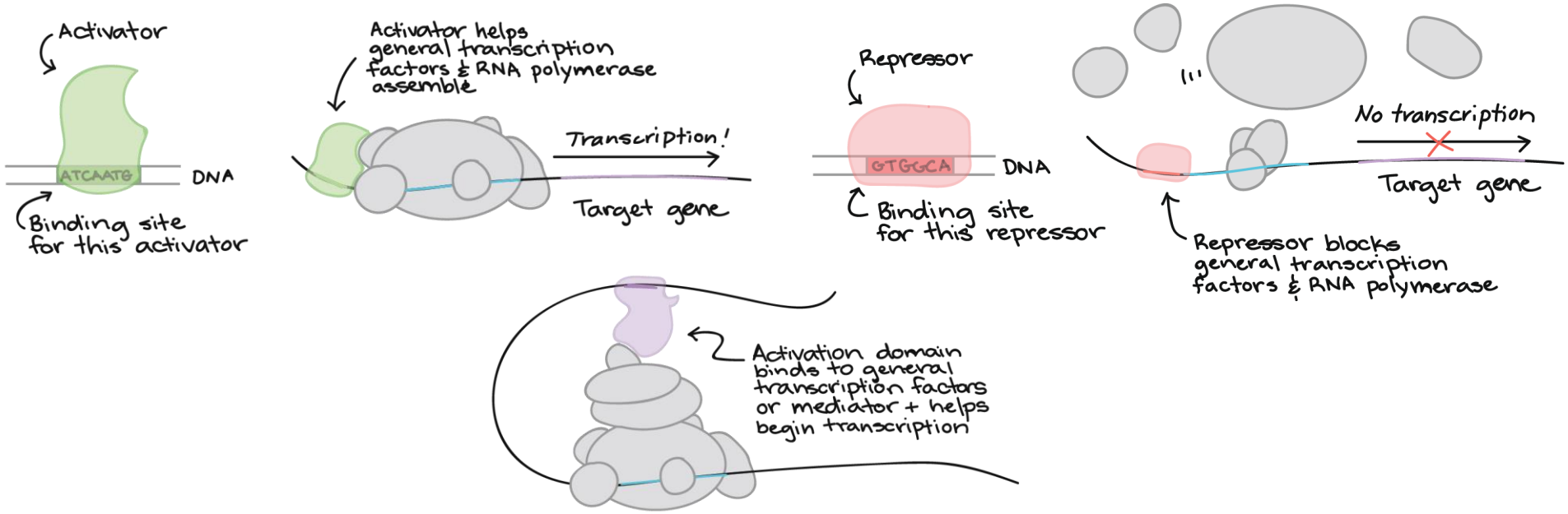
  (Van Steen, Cluj, 2015)

- **Data level:**

  Integration is the process of connecting systems (which may have fusion in them) into a larger system (Oxley & Thorsen, 2004) → accounting for interactions

“Message passing” between network layers of information

- Example: Transcription Factors (TFs)



TF interaction network (TranSignalTM TF ProteinArray) ⟷ TRRUSTv2 (Han et al. 2018) ⟷ Gene co-expression network

Need for a paradigm shift

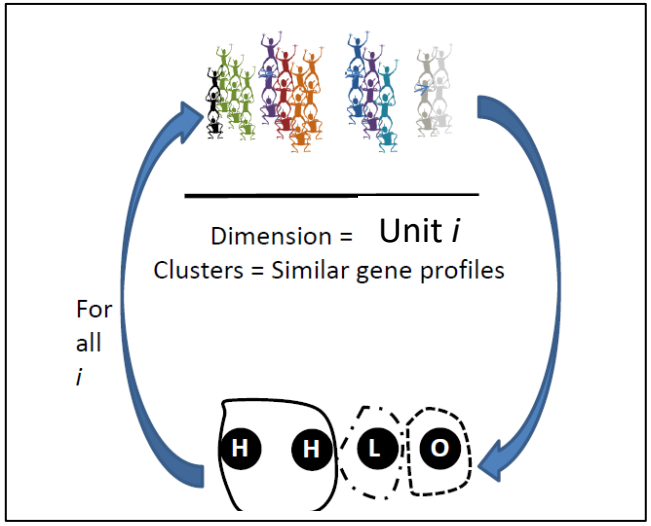
Paradigm shift at the level of integration: systems and granularity

- **Analytic integration** (modelling paradigms) – INFANCY

- **Data integration** (heterogeneous data types) – WELL PROGRESSING

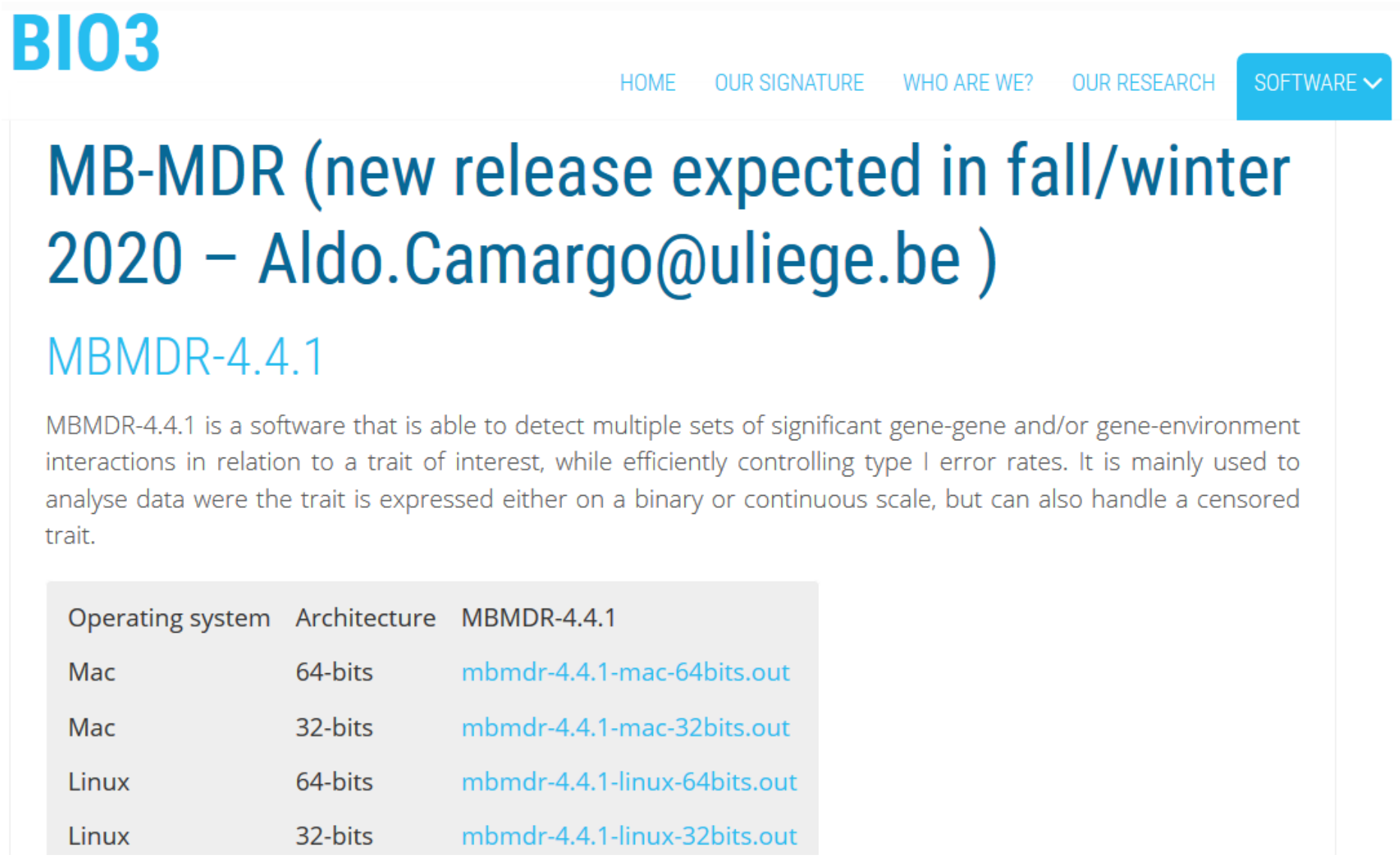
Example: revised units for MB-MDR ≥ 1D analyses

while being as comprehensive as possible within mini-systems



- Component-based
- Kernel-based
- Network-based

(Fouladi et al. 2015-2018;
Ocira et al. - DEST*in*CT grant ;
Walakira et al. 2020)



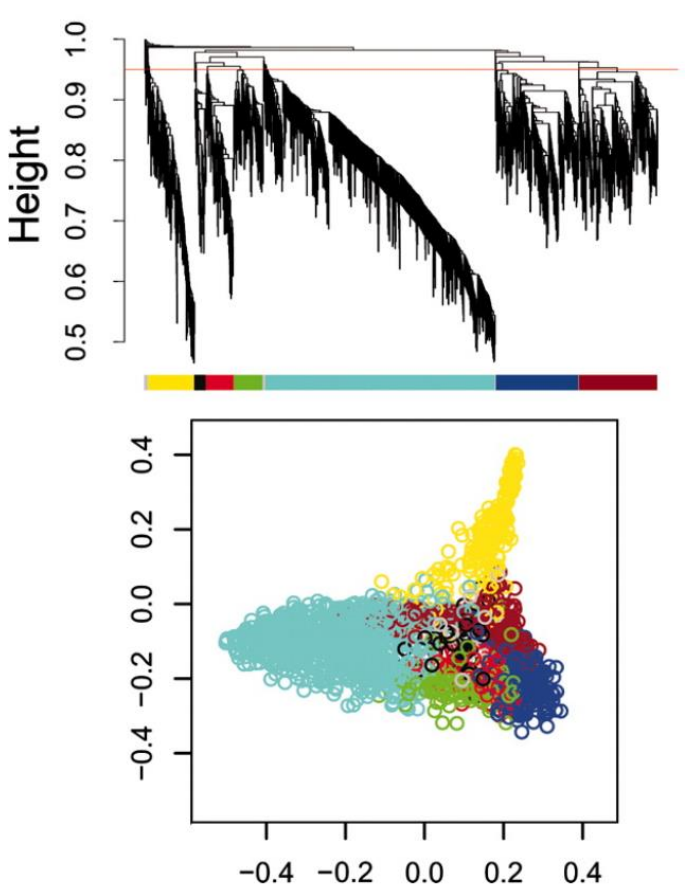
(http://bio3.giga.ulg.ac.be)

Expected benefits in the context of PM

1. Increased replicability & robustness when used in models for
    - prediction
    - diagnosis
    - treatment management
2. Less granularity, yet based on deep levels of detail
3. Well suited to create individual-specific (gene interaction) networks

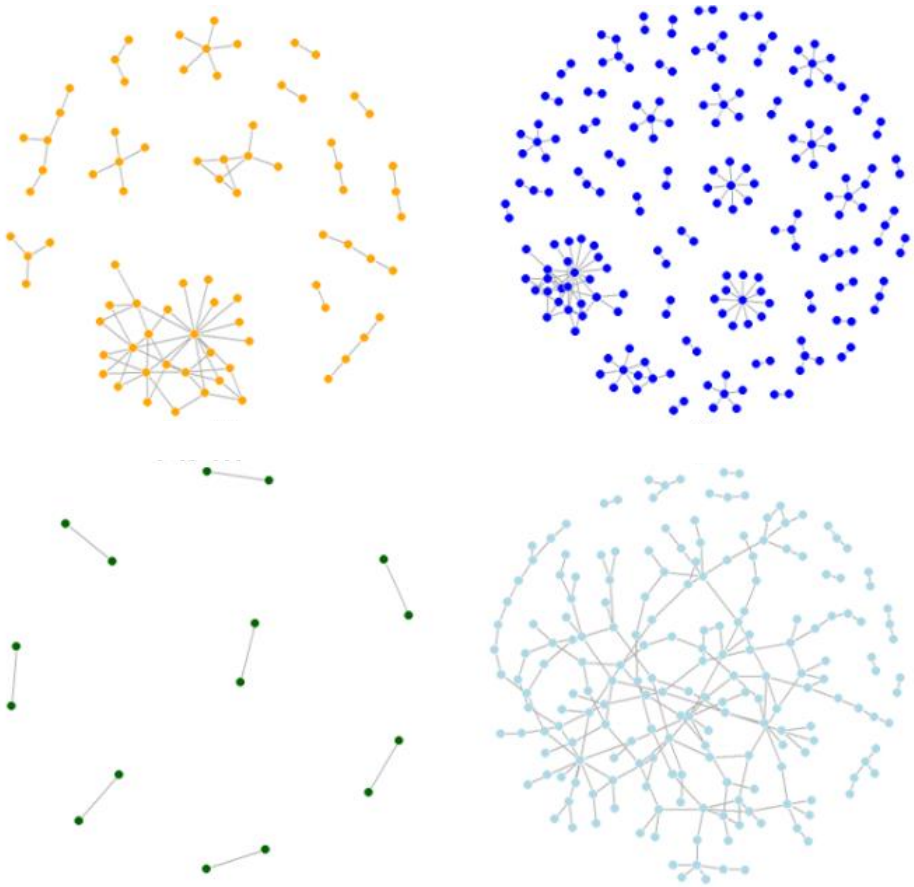
Paradigm shift at the level of interactions: individual-specific edges

group-based



(Oldham et al 2006)

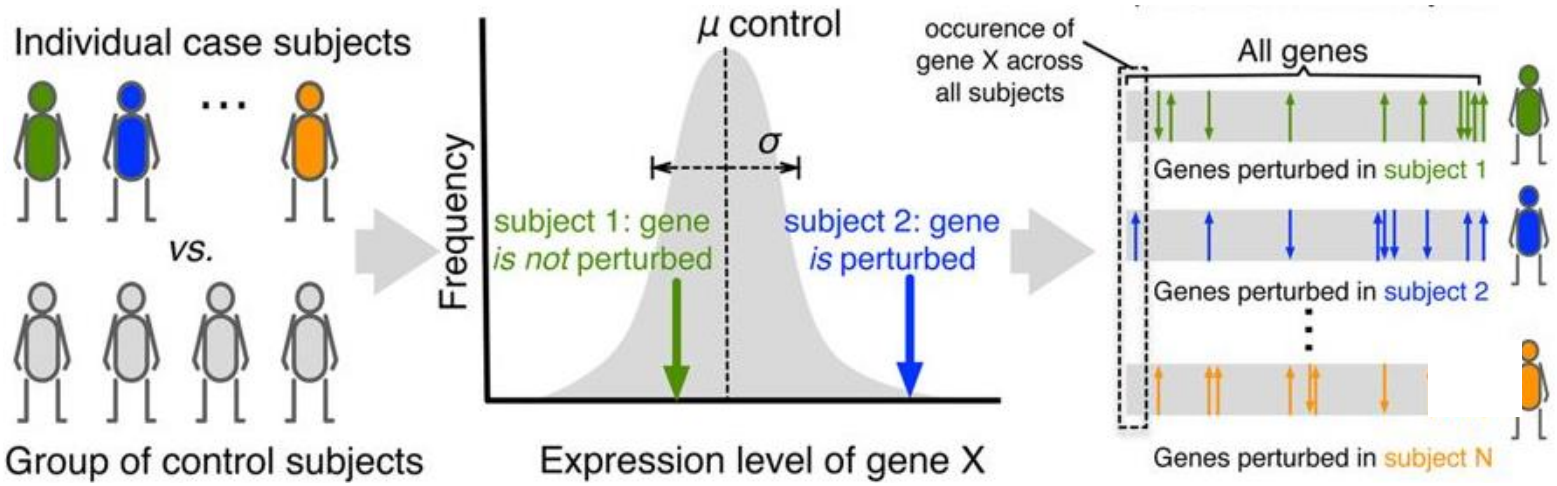
individual-based



(**Duroux et al 2020** – use of bio knowledge)

Types of individual-specific networks

- Omics-driven individual-specific “significant” nodes (Menche et al. 2017; **Bhardwaj et al 2020 application to PDAC**)



- Individual-specific “significant” edges via leave-one-out (Van Steen 2015 - @ ETHZ; Kuijjer et al. 2018)

↓

“Truly” individual-specific “significant networks”

Expected benefits in the context of PM

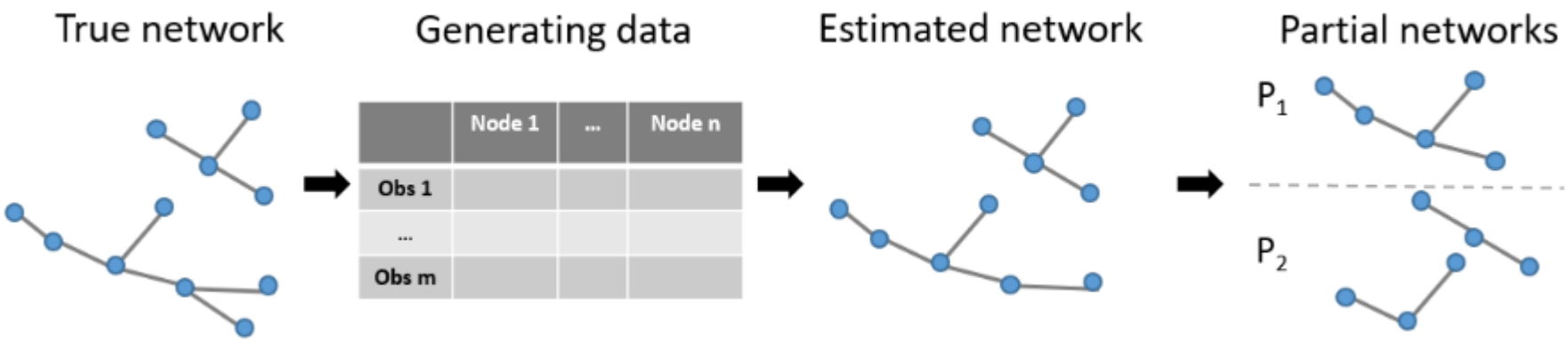
1. Assessing individual-specific (systems level) properties
2. Comparing individuals in a systemic way, enhancing
    - patient subtyping for stratified medicine or
    - molecular reclassification of disease
3. Comprehensive views on “outlying” individuals

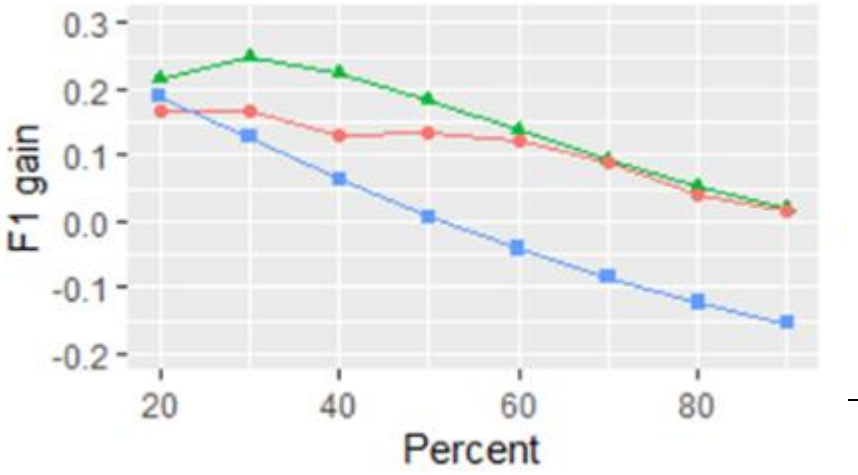
Major hurdles?

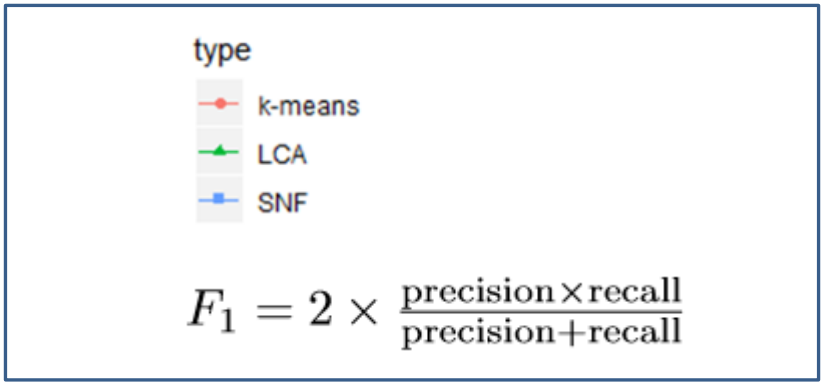
- Computation time
  - F.i. Leave-one out requires following the same underlying analytics that generated the source network
  - In more detail:
    - Create a network based on thousands of individuals
    - Leave one out and rebuild the network
    - Use the developed statistical tool to assess significance and to rebuild an individual-specific network
- Robust networks
- Contributing data
  - Informativity versus redundancy
  - Samples contributing to the “reference”

Example: SENs & robustness

- Aggregating networks that are partial representations of a true network (**Duroux et al. 2020**)







$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Example: SENs & computation time

- Multiple testing correction via “gammaMAXT” in MBMDR-4.2.2:

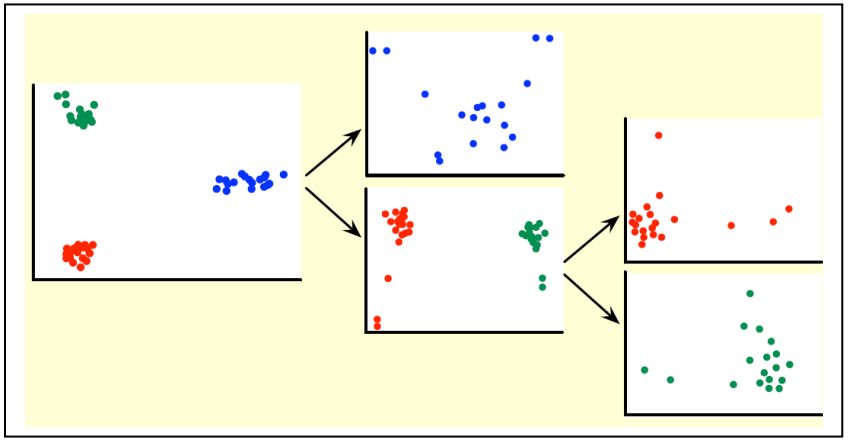
SNPs	Sequential version Binary trait	Parallel workflow Binary trait	Sequential version Continuous trait	Parallel workflow Continuous trait
10^3	13 min 33 sec	20 sec	13 min 18 sec	18 sec
10^4	52 min 15 sec	1 min 05 sec	56 min 14 sec	53 sec
10^5	64 hours 35 min	22 min 15 sec	70 hours 03 min	20 min 28 sec
10^6	$\approx$ 270 days	25 hours 12 min	$\approx$ 290 days	24 hours 06 min

The parallel workflow was tested on a 256-core computer cluster (Intel L5420 2.5 GHz 1333 MHz FSB). The sequential executions were performed on a single core of this cluster. The results prefixed by the symbol “$\approx$” are extrapolated.

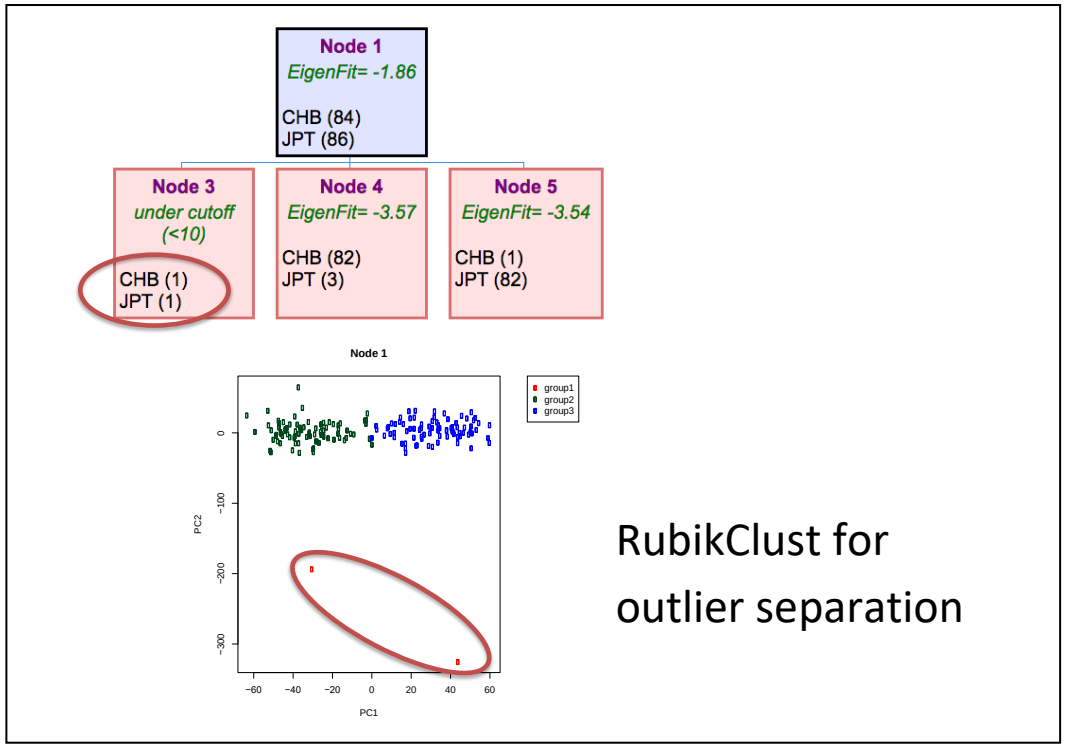
(Van Lishout et al. 2015)

Example: SENs & fine-scale structure

- Identifying influential or extreme patients
- Build on ipPCA (Intarapanich et al. 2009) for population samples:
  - Performs PCA with genotype data
  - If substructure exists in PC space individuals are assigned to one of two clusters (2-means algorithm / fuzzy c-means)
  - Iteratively performs test for substructure and clustering on nested datasets until stopping criterium is satisfied
    (no substructure)

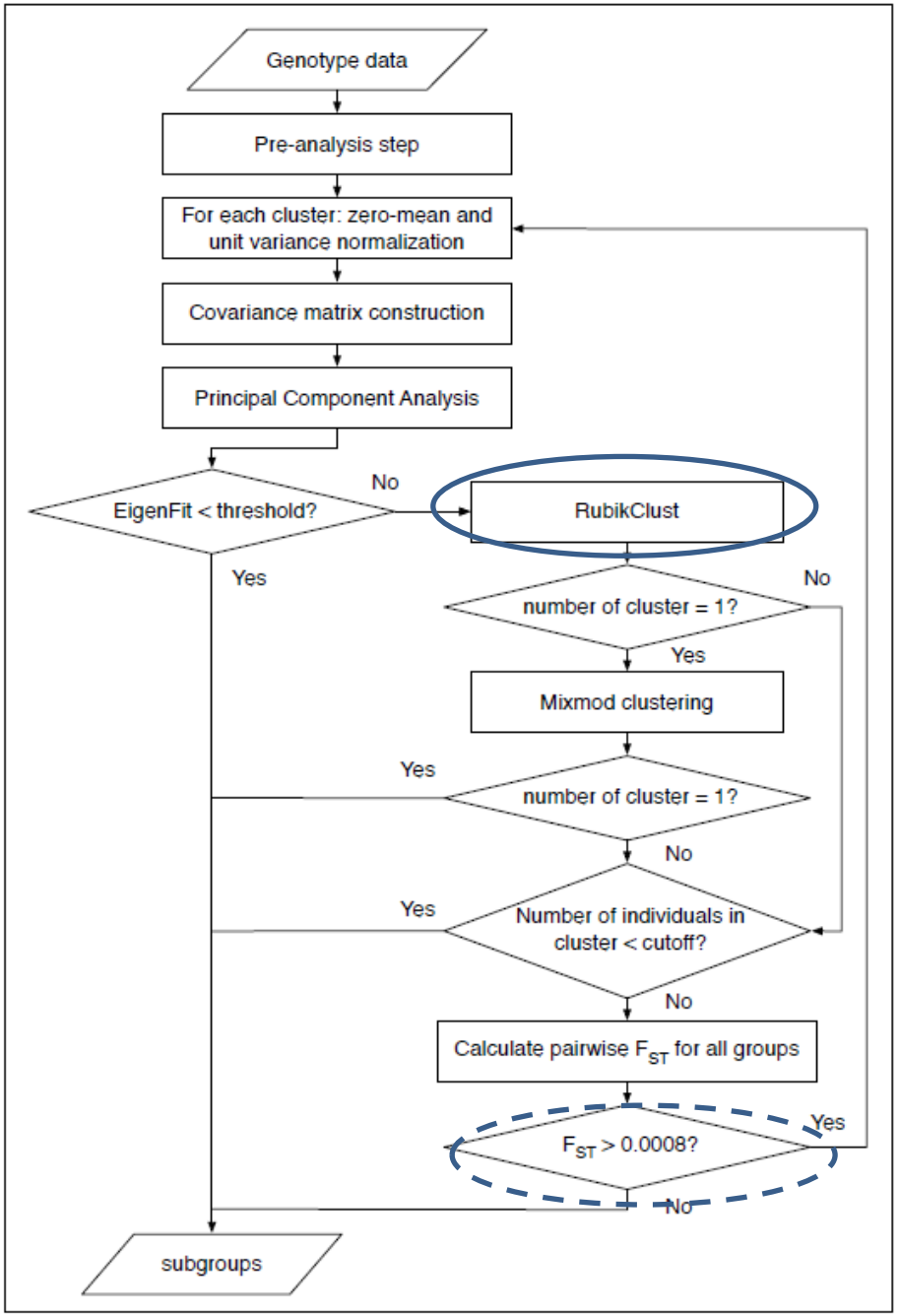


IPCAPS

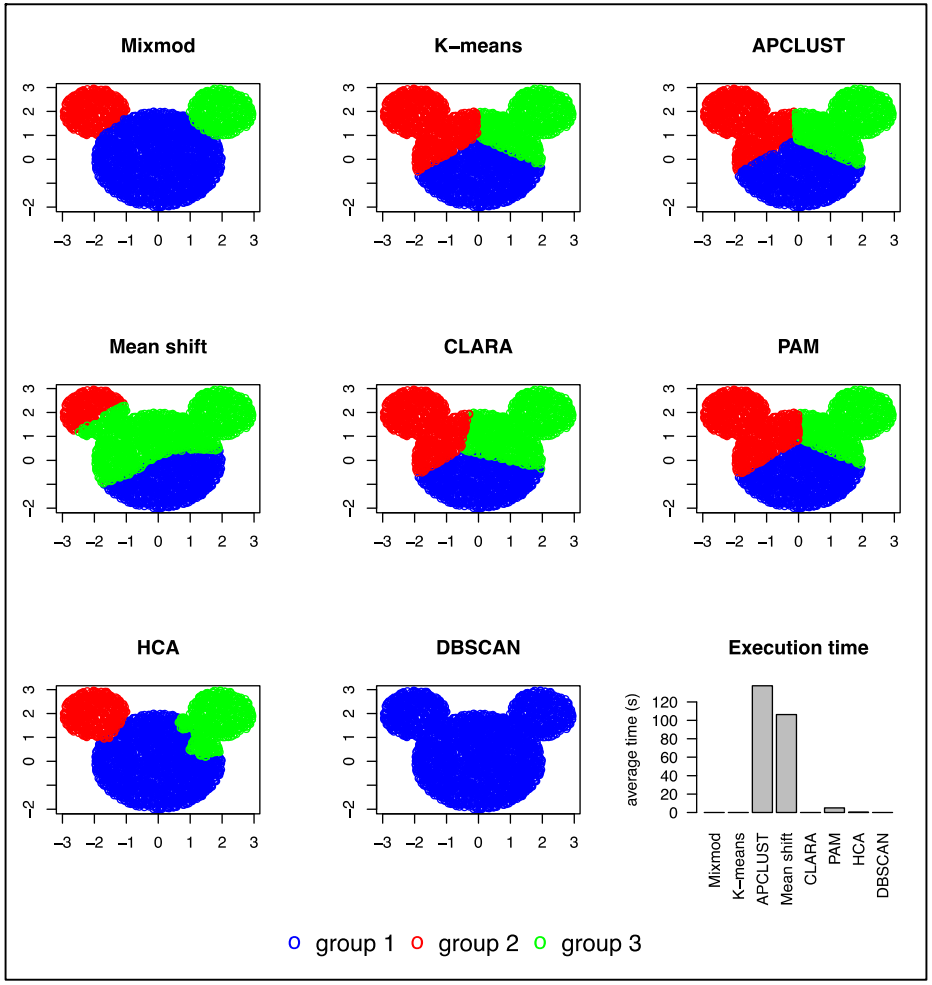


RubikClust for outlier separation

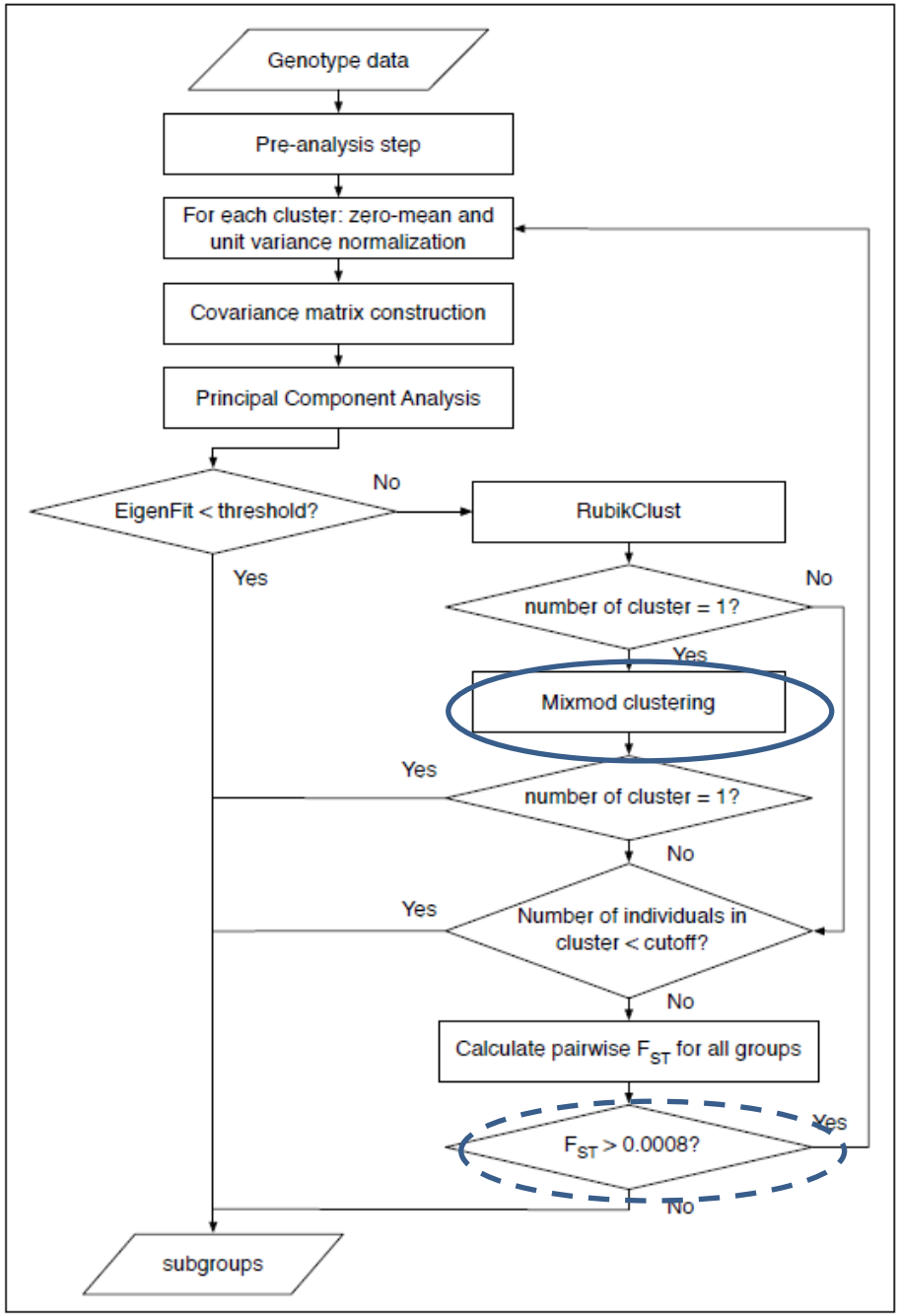
(Chaichoompu et al. 2019)



IPCAPS



(Chaichoompu et al. 2019)



IPCAPS Software:

Chaichoompu *et al. Source Code for Biology and Medicine* (2019) 14:2
https://doi.org/10.1186/s13029-019-0072-6

Source Code for Biology and Medicine

SOFTWARE Open Access

IPCAPS: an R package for iterative pruning to capture population structure

Kridsadakorn Chaichoompu[1*], Fentaw Abegaz[1], Sissades Tongsima[2], Philip James Shaw[3], Anavaj Sakuntabhai[4,5], Luísa Pereira[6,7] and Kristel Van Steen[1,8*]

IGES October 2019 highlighted paper:

Human Genetics (2020) 139:45–59
https://doi.org/10.1007/s00439-019-02069-7

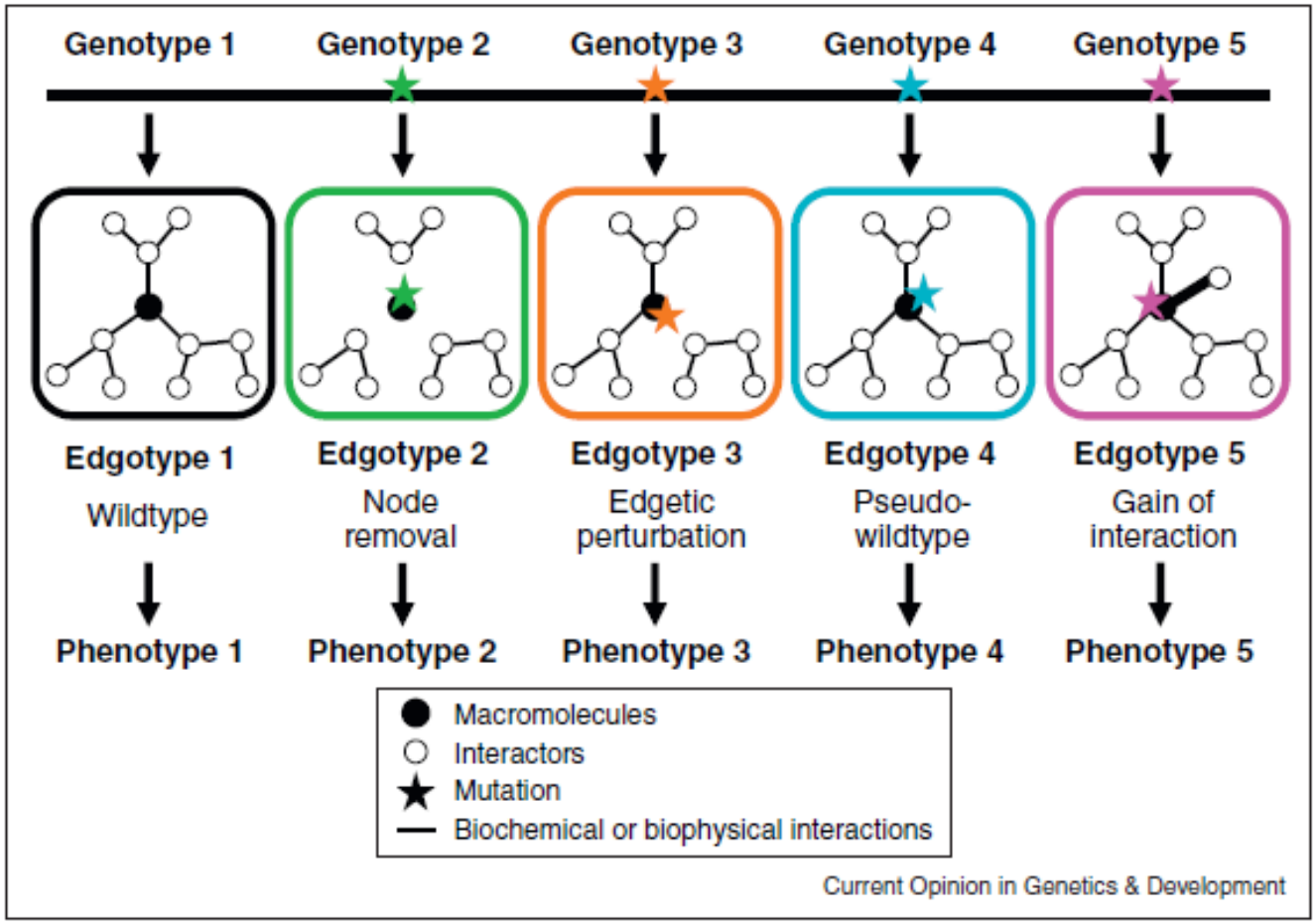
ORIGINAL INVESTIGATION

A different view on fine-scale population structure in Western African populations

Kridsadakorn Chaichoompu[1,2] · Fentaw Abegaz[1] · Bruno Cavadas[3,4] · Verónica Fernandes[3,4] · Bertram Müller-Myhsok[2] · Luísa Pereira[3,4] · Kristel Van Steen[1,5]

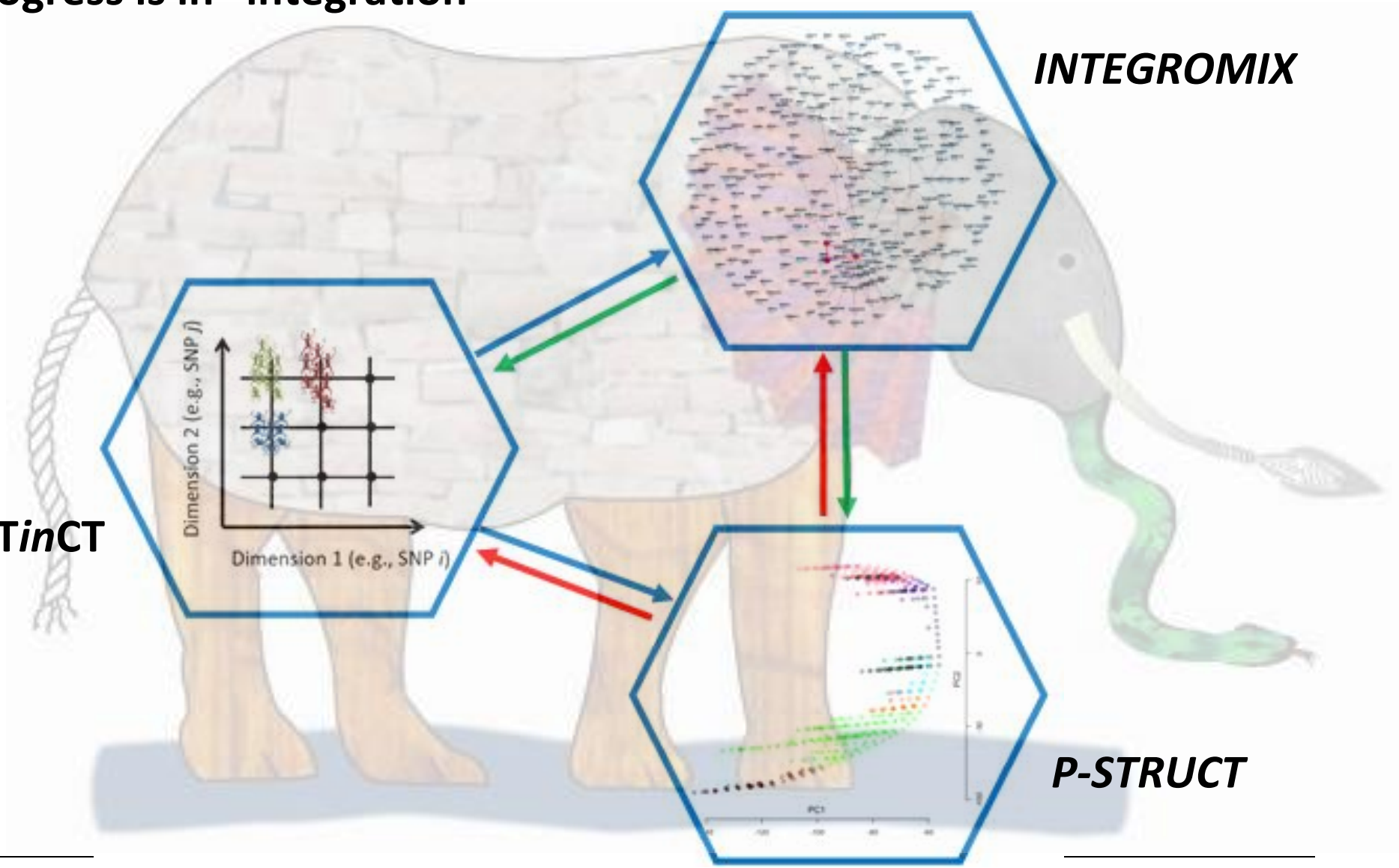
Take-home messages

1: Information is in “the edges” & learn from mini-systems

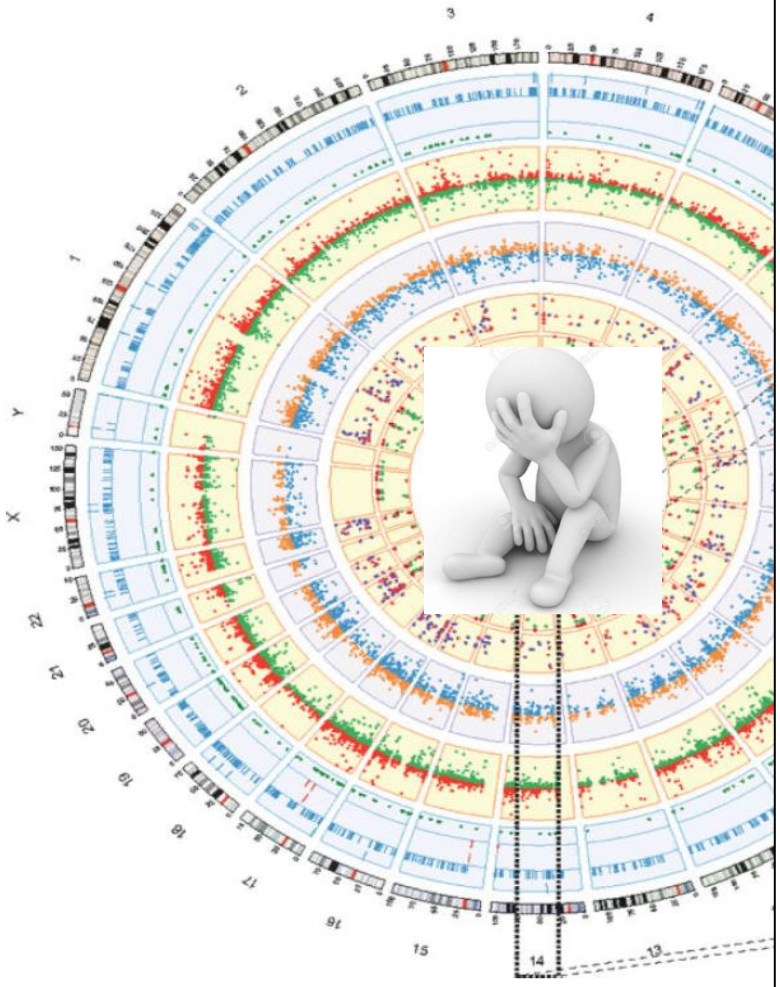


(moving towards individual networks)

2: Progress is in “integration”



3: The need for more data is relative



Do you think that omics profiling will be routinely used in the clinic in future?

"Not in the form we are doing it. At the moment we have a very incomplete picture of what's going on, whereas if we were able to make thousands of measurements we would have a much better feeling. We just don't know, for the clinical tests, which thousand measurements are going to be most useful. We'll need certain measurements for diabetes, others for cancer, and specific tests will probably reveal themselves useful for different diseases."

(Snyder 2014)

Redundancy – Informativity - Structure

Acknowledgements

GIGA-R, Medical Genomics Thematic Research Unit, Liège, Belgium

http://bio3.giga.ulg.ac.be/

BIO

	22/11/2020 Sunday	23/11/2020 Monday	24/11/2020 Tuesday	25/11/2020 Wednesday	26/11/2020 Thursday	27/11/2020 Friday
8.30			**Pharmacogenomics in clinical practice** (Vita Dolžan, University of Ljubljana, Slovenia)			
8.45						
9.00	Dynamic presenting & Personal Coaching (UK Bodytalk)	Moonlanding exercise powered by NASA (Expert academy)	**Addressing the heterogeneity in liver diseases using biological networks** (Adil Mardinglu, KTH Stockholm, Sweden)	**Study Design 1: post-GWAS cohort studies** (L Franke, UMCG, the Netherlands)	**Mining and digesting tons of information** (M. Butz-Ostendorf, BIOMAX, Germany)	**Technologies 2: Imaging in the era of Big Data** (P Claes, KU Leuven, Belgium)
9.15						
9.30			**Challenges in the introduction of next-generation sequencing in hematology** (Nataša Debeljak, Irena Preložnik Zupan, Uni of			
9.45						
10.00			BREAK	BREAK	BREAK	BREAK
10.15						
10.30			**Systems thinking as the framework of systems technicality** (Christian Pristipino, University of Rome, Italy)	**Keynote lecture** L Milani, Institute of Genomics, University of Tartu	**Study Design 3: Experimental designs** (P. Real, CNIO, Spain)	**Technologies 3: iPSC and single cell RNAseq** (A Skupin, LCSB, Luxembourg)
10.45						
11.00			**Keynote lecture:** Artificial intelligence in systems biology (Andrej Zinvoyev, Institute Curie Paris, France)	SELECTED TALKS		
11.15						
11.30				BREAK	Supervisory Board meeting	Valorisation Board Meeting
11.45						
12.00			INTERACTIVE POSTER SESSION	ESR elections to boards		
12.15						
12.30	BREAK	BREAK				
12.45						
13.00						
13.15						
13.30	Dynamic presenting & Personal Coaching (UK Bodytalk)	Formal welcome: D. Rozman and guests	**Keynote lecture** Eran Segal, Weizmann Institute of Science	**Study Design 2: Cohorts for Aging Research** (C Van Duijn, NDPH Oxford, UK)	**Identify your blogging voice** (A Winfield, UWE Bristol, UK)	**TranSYS ESR flash presentations**
13.45		Introduction to TranSYS: K.V.Steen				
14.00		**Reproducibility in Systems Biology Modelling — Sometimes** (H. Hermjakob, EBI-EMBL, UK)				
14.15						
14.30			SELECTED TALKS			SELECTED TALKS
14.45		**Disease-oriented data mining** (Vitor Martins dos Santos, LIFEGLIMMER Gmbh, Germany)				
15.00			BREAK	BREAK	BREAK	BREAK
15.15		BREAK				
15.30						
15.45		**From experimental to computational models** (T. Režen, M. Moškon, D. Rozman, University of Ljubljana, Slovenia)	**Workshop: from sample to insight** (Leif Schauser, Qiagen, Denmark)	**Agile Development** (Lab900)	**Technologies 1: Emerging concepts in liquid biopsies** (D Ge, Apostle/Duke University Medical Center, USA)	**Open & Closed Science & Data** (R Head, CERATIUM, UK/the Netherlands)
16.00						
16.15						**Gender bias/impact in the context of health sciences** (B Greenwood, George Mason University)
16.30		**Keynote lecture:** Toward clinical decision support in oncology: Identifying driver alterations and therapeutic options (Nikolaus Schultz, Memorial Sloane Kettering Ctr, USA)				
16.45						
17.00				WG Leaders / Lab900 Meeting	TranSYS PhD Committee meetings	**Closure** (K Van Steen & D Dozman)
17.15						
17.30		Discussion				Organising Committee Meeting
17.45						
18.00				TranSYS helpdesk (C.Olsen, CERATIUM)		
	SESSIONS:	TranSYS Workshops	TranSYS affiliated speakers	Selected from abstract submission	Invited speakers	TranSYS only activities

https://h2020transys.eu/dissemination-exploitation/dissemination/conferences/