Stratified medicine & translational science within large scale randomized trials

Kristel Van Steen, PhD[2] (*)

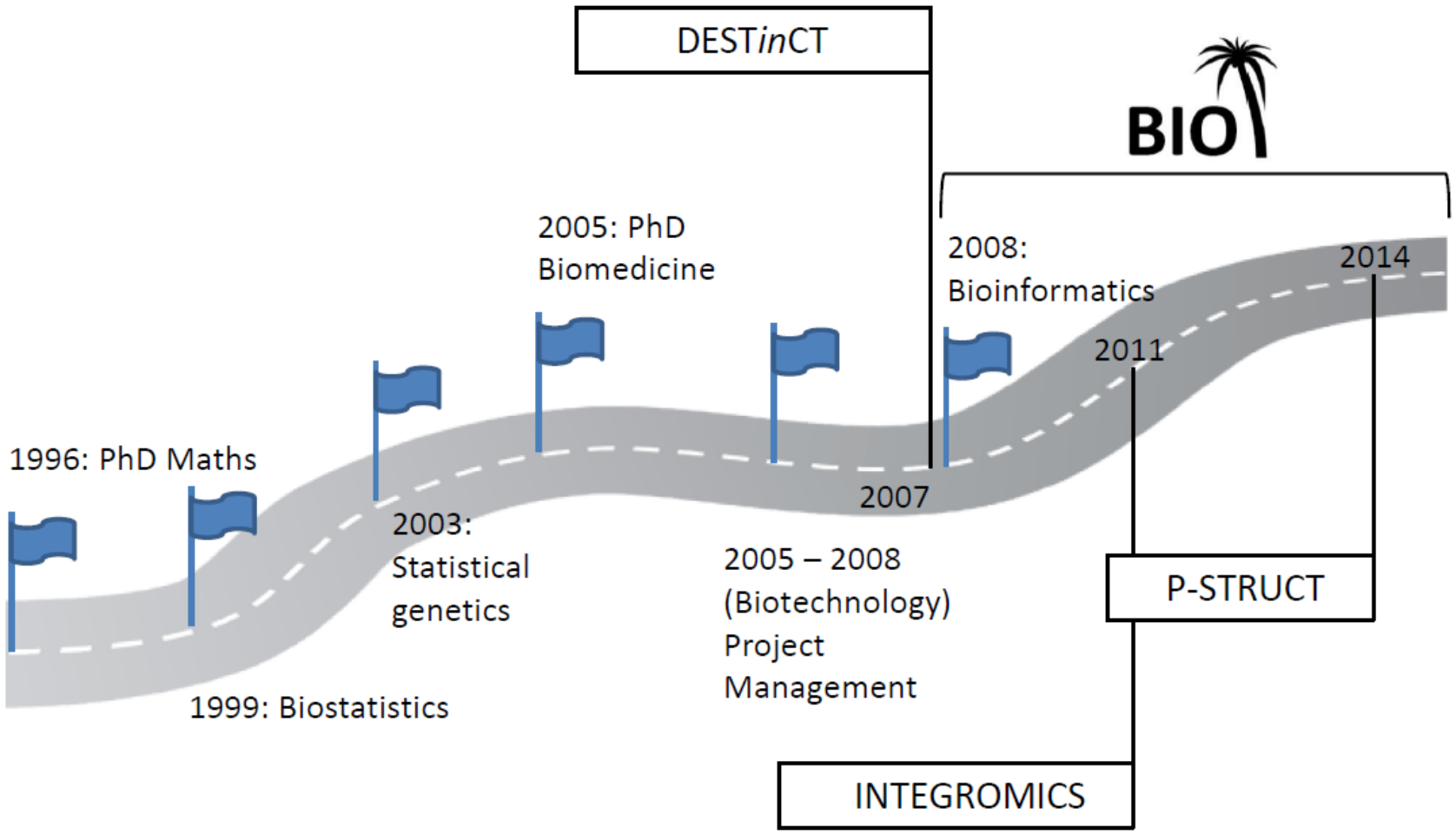
kristel.vansteen@ulg.ac.be

(*) WELBIO, GIGA-R, Medical Genomics, University of Liège, Belgium

Systems Medicine Lab, KU Leuven, Belgium

BIO

The road less traveled



OUTLINE

- Molecular biomarkers
- Basic science – How does it work?
- Translational science – Turning knowledge into sth useful?
- Clinical science – Is it really useful?
- Take-home messages

Molecular biomarkers

What are molecular biomarkers?

- A biological marker, or biomarker, is something that can be measured, which points to the presence of a disease, a physiological change, response to a treatment, or a psychological condition.
- A molecular biomarker is a molecule that can be used in this way.
- Biomarkers are used in different ways at different stages of medicines development, including in some cases as a surrogate endpoint to indicate and measure the effect of medicines in trials.

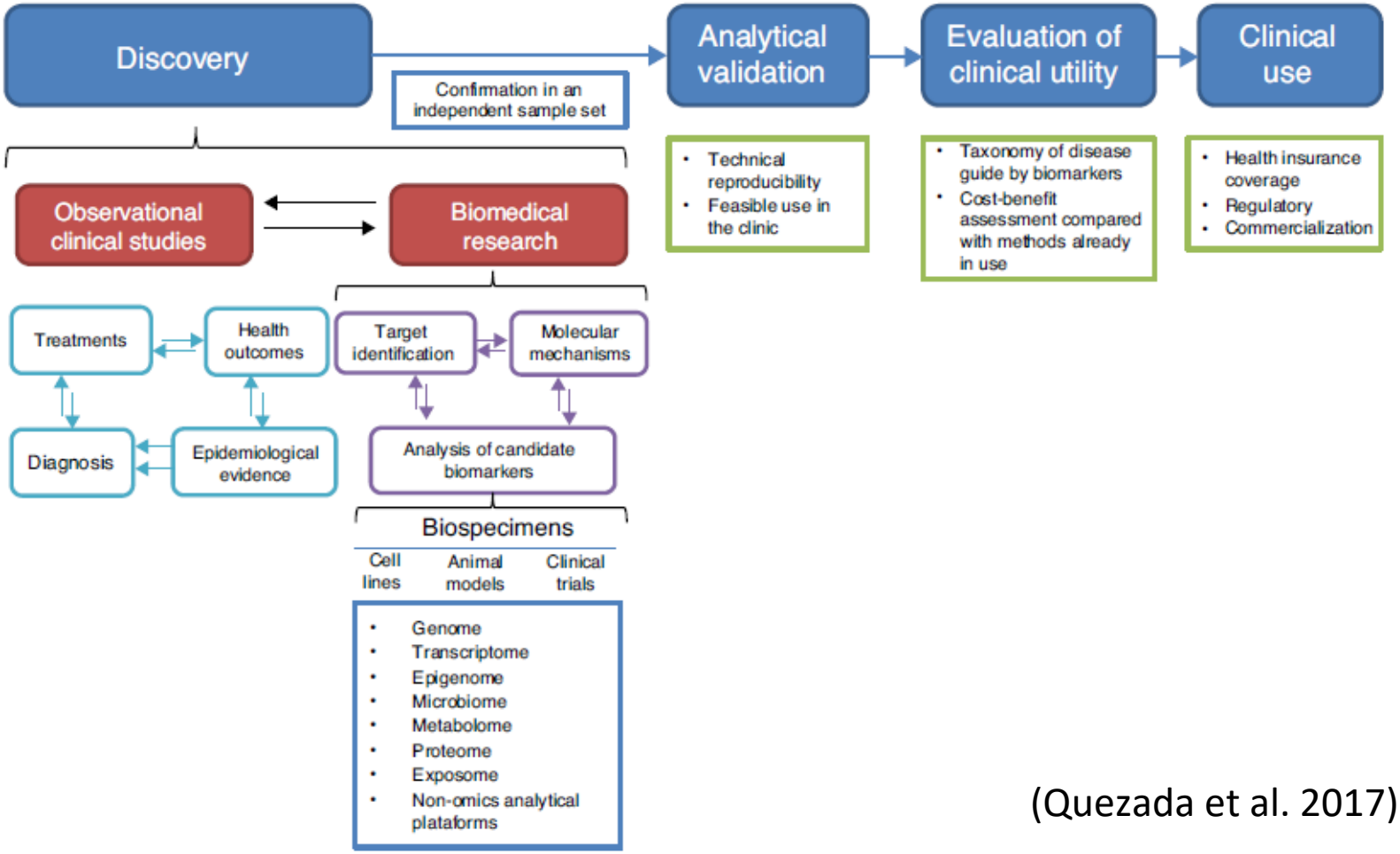
(www.eupati.eu)

Types of biomarkers

- **Diagnostic** biomarkers are used to determine the specific health disorder of the patient
- **Prognostic** biomarkers help to chart the likely course of the disease
- **Predictive** biomarkers indicate the probable response to a particular medicine
- **Predisposition** biomarkers indicate the risk of developing a disease

(Quezada et al. 2017)

The biomarker development process



(Quezada et al. 2017)

Instruments to discover molecular biomarkers?

Basic Science

"how things work"

- Understanding:
  - Comparisons
  - Profiling / Subtyping
- Predicting:
  - Future educated guesses

Translational Science

"how to create sth useful" (for whom…)

- Bioinformatics-driven pipeline based on molecular biomarkers to drive treatment management
- Stratified medicine: how much heterogeneity is allowed in strata to target?

Stratification in the frame of personalized medicine

- Stratification is the identification of a group of patients with shared "biological" characteristics by using molecular, biochemical and imaging diagnostic testing to select the optimal management for the patients and achieve the best possible outcome in terms of (based on the category and disease characteristics):

  - Risk assessment and prevention
  - Achievement of the optimal treatment outcome

(ec.europa.eu/research/health/)

Basic science

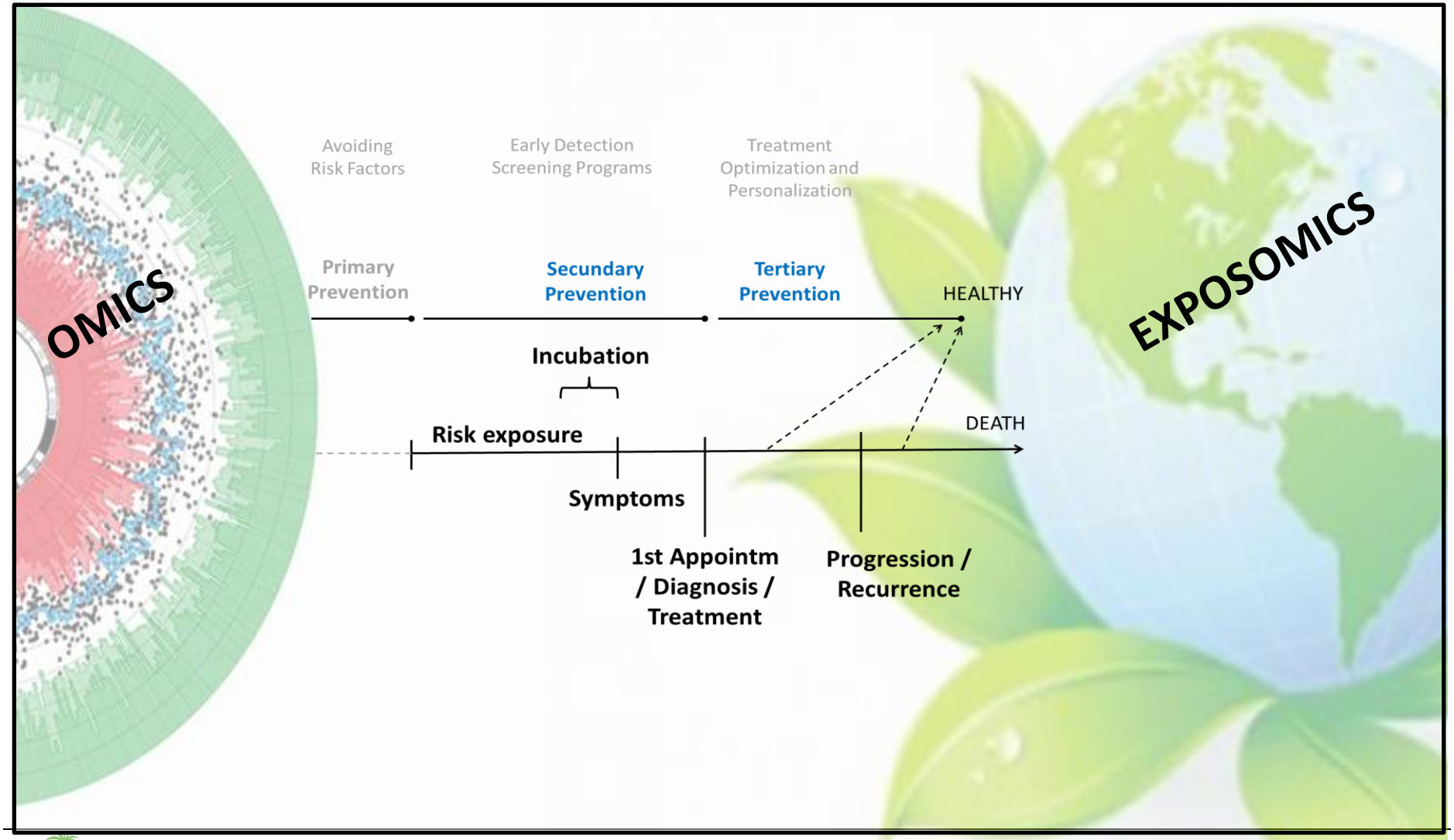
The context of precision medicine

Precision medicine is …

"a medical model using the 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)

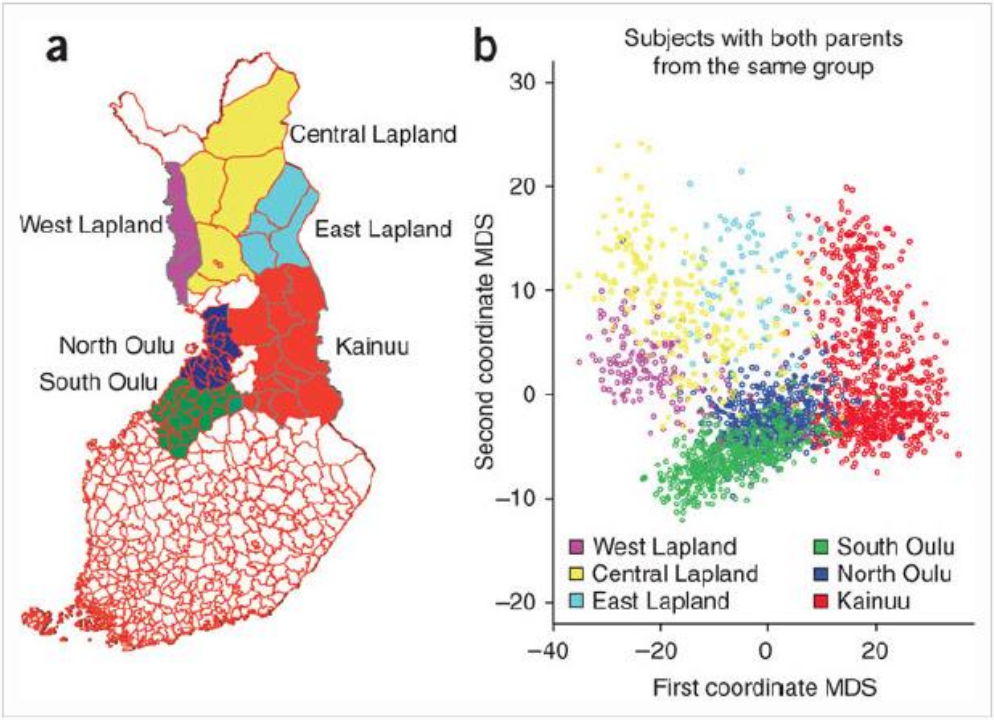
An individual's ecosystem comes with a data deluge



BIO

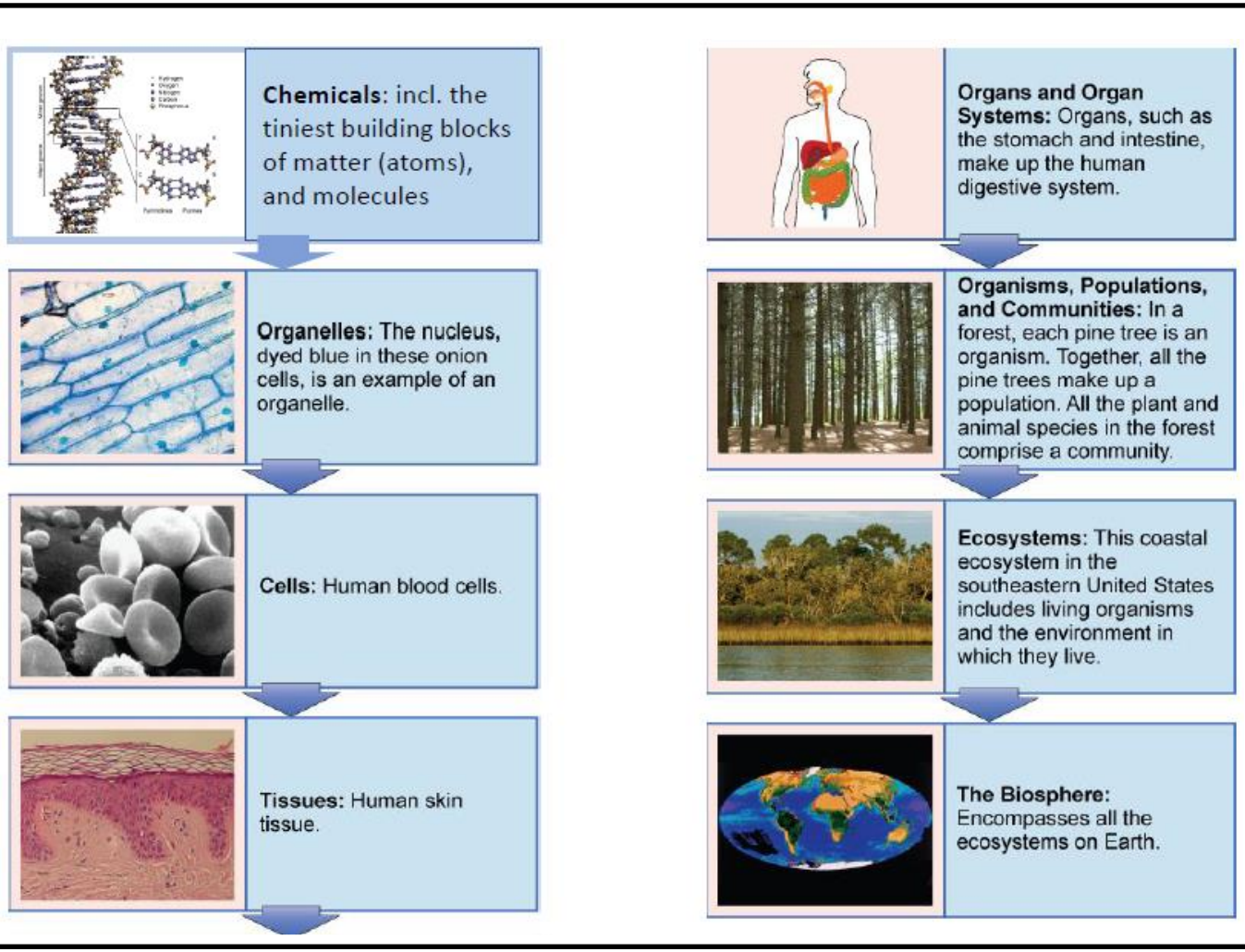
Population heterogeneity

- There can be population structure in all populations, even those that appear to be relatively “homogeneous”

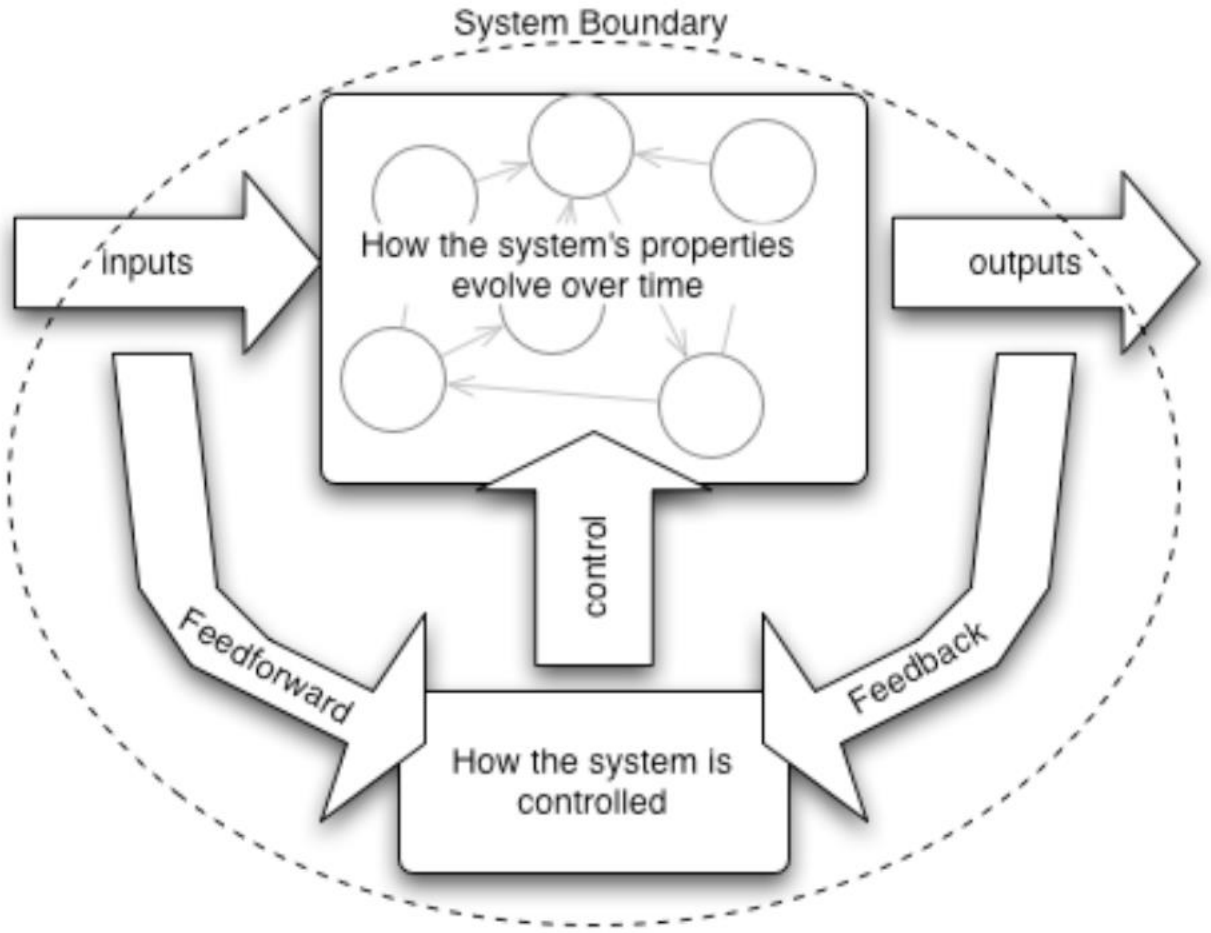


(Sabatti et al. 2009)

Patient heterogeneity

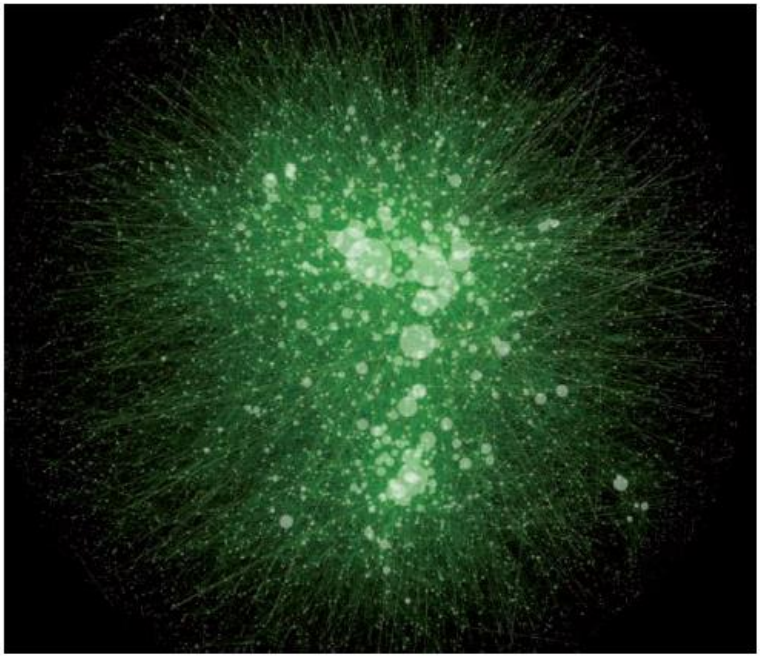


Systems and their eco-system - interactions



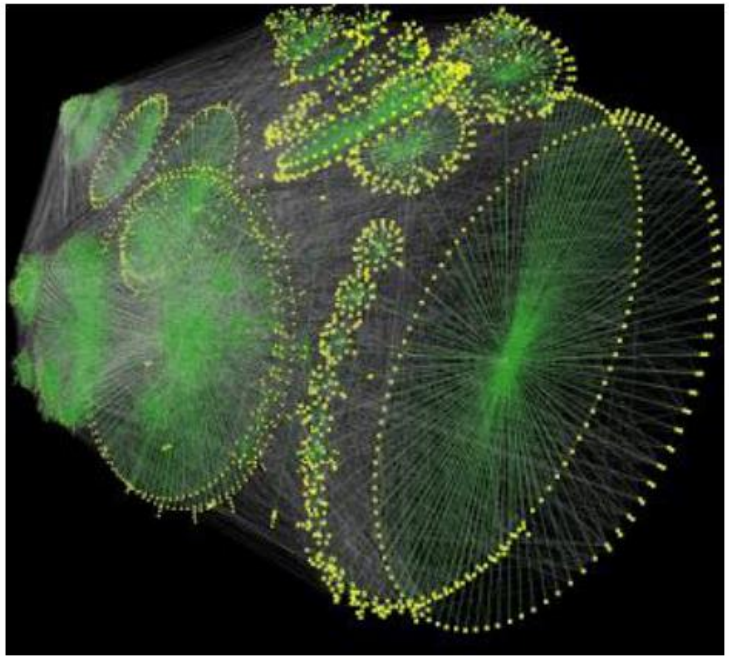
(@2004-5 Steve Easterbrook)

From interactions to the interactome



Human interactome (PPI)

(Bonetta 2010)



Fruit fly interactome

(www.molgen.mpg.de)

How ready are we to start integrating data?

- **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

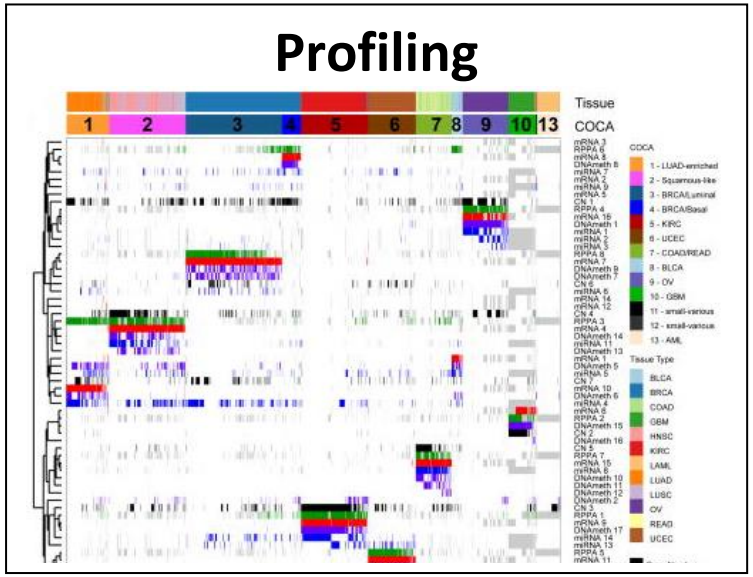
- **ANALYTIC LEVEL:**

  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)

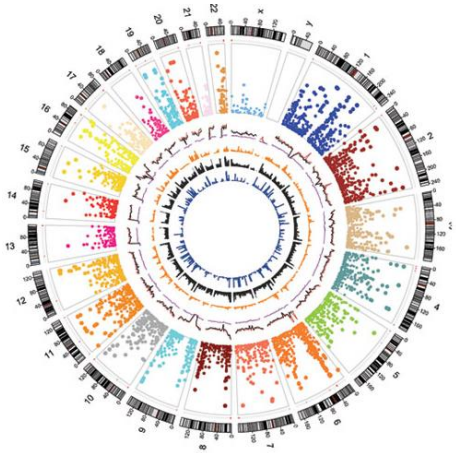
How ready are we for genuine data integration?

Profiling



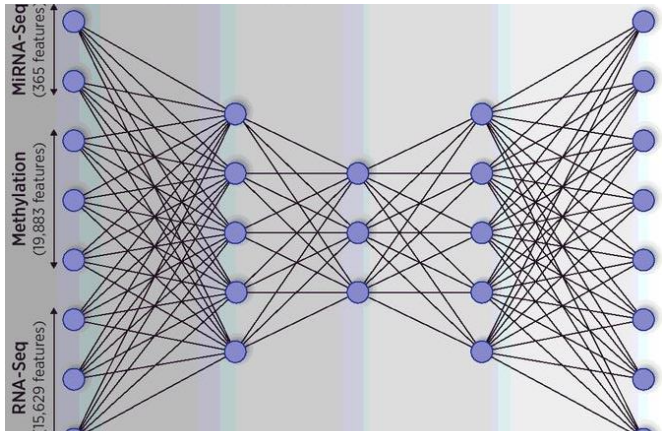
(Hoadley et al. 2014; Consensus Clust)

Comparing



(Jun Li et al. ’12; GWASrap)

Prediction

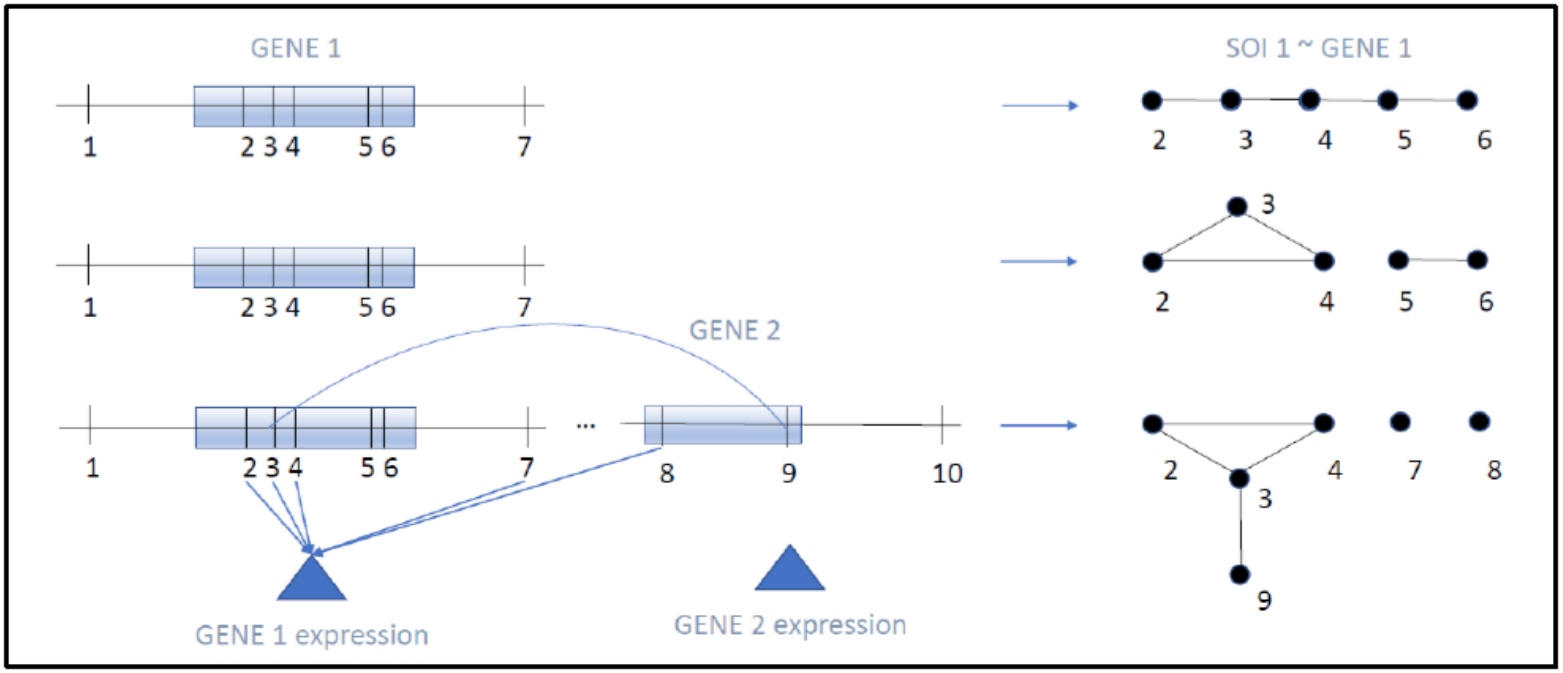


(Chaudhary et al. 2018; Deep Learning)

BIO

BIO3's approach: advanced integration in smaller systems

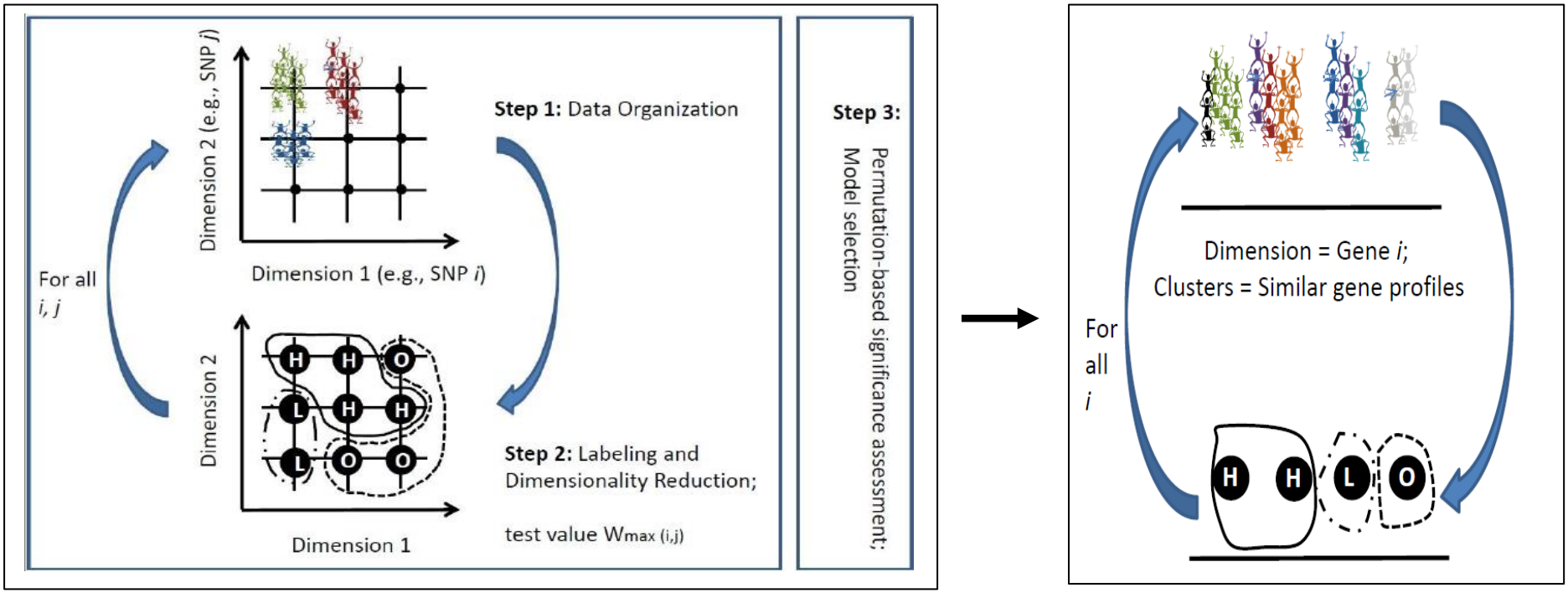
- Define sets of features that make sense …
- Example: consider a gene to be a system to be comprehensive about



BIO3's approach: advanced integration in smaller systems

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

Ex: MB-MDR + diffusion kernels on graphs

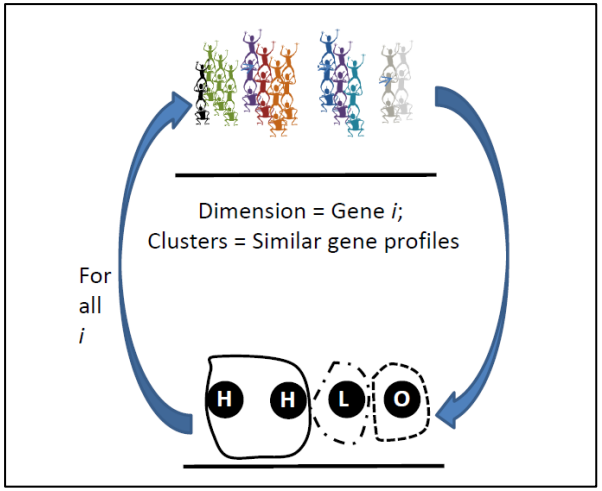


BIO3's approach

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

Ex: MB-MDR + diffusion kernels on graphs

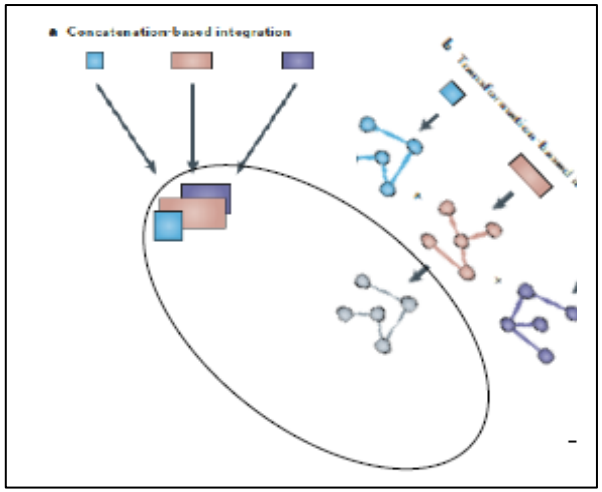
to perform omics-integrated gene-based sample clustering



(DEST*in*CT : MB-MDR)

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

(Fouladi et al. 2015-2018)



(Ritchie et al. 2015)

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

Translational science

Basic Science

"how things work" -

INTEGRATE for

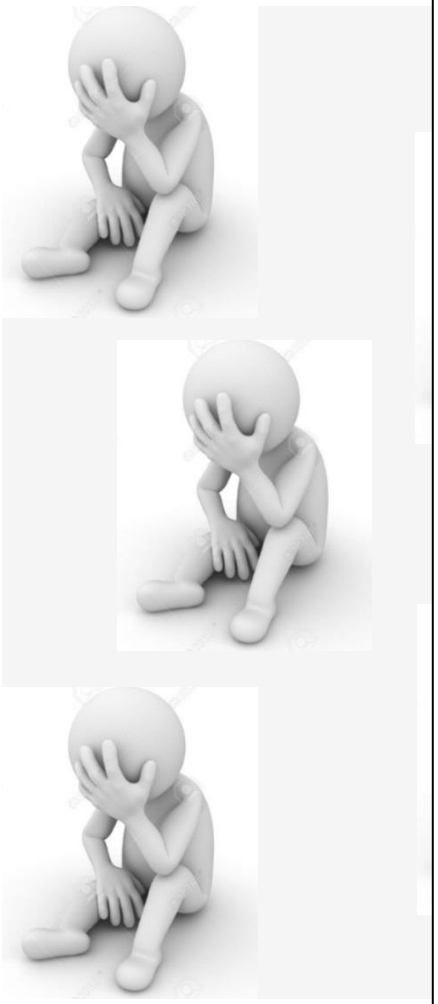
- Understanding:
  - Comparisons
  - Profiling / Subtyping
- Predicting:
  - Future educated guesses

Translational Science

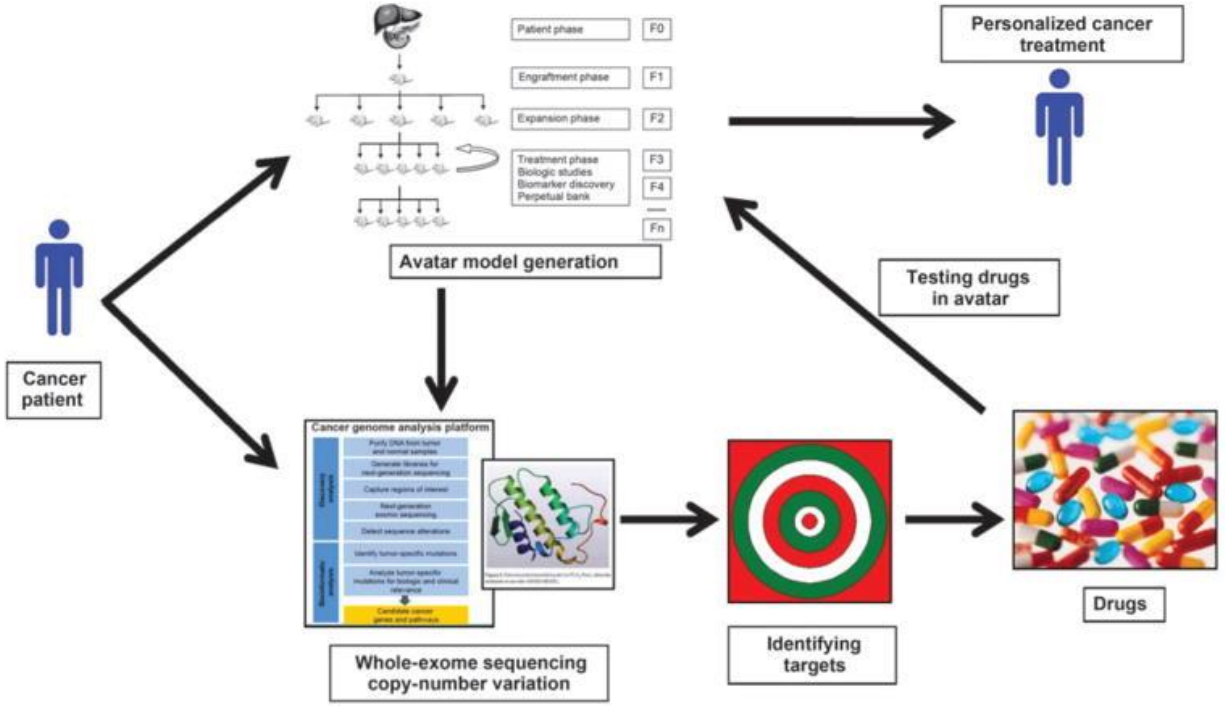
"how to create sth useful" (for whom…)

- Bioinformatics-driven pipeline based on molecular biomarkers to drive treatment management
- Stratified medicine: how much heterogeneity is allowed in strata to target?

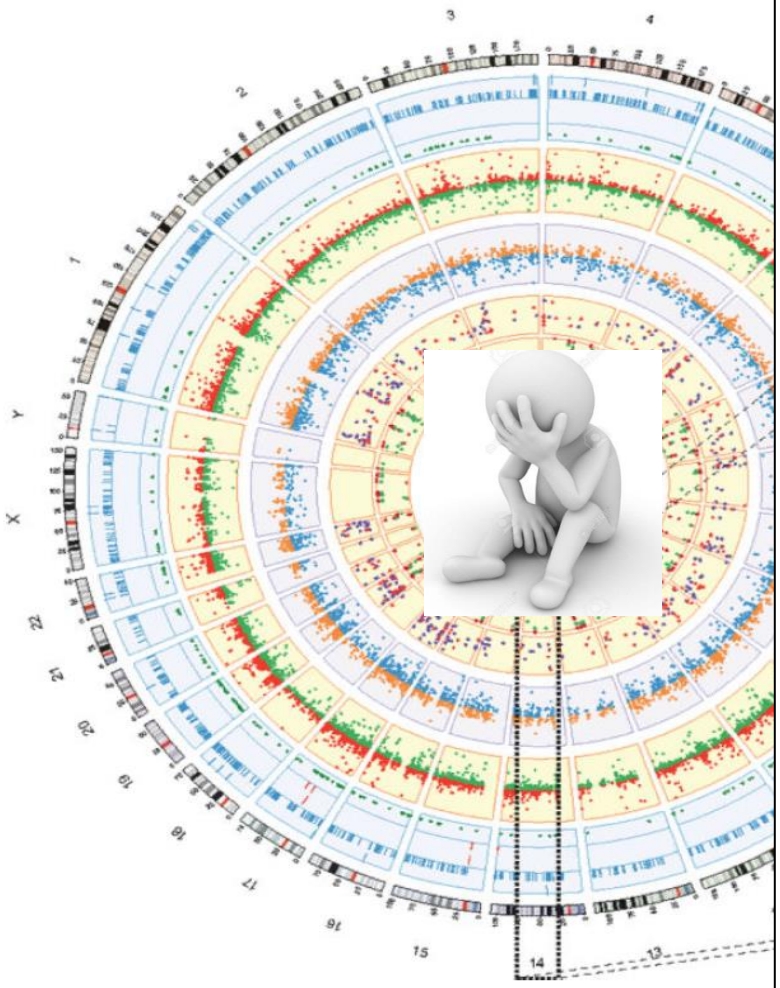
TIER 1: Bionformatics-driven treatment management



Integrating sequencing and avatar mouse models



(Garralda et al. 2014)



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."

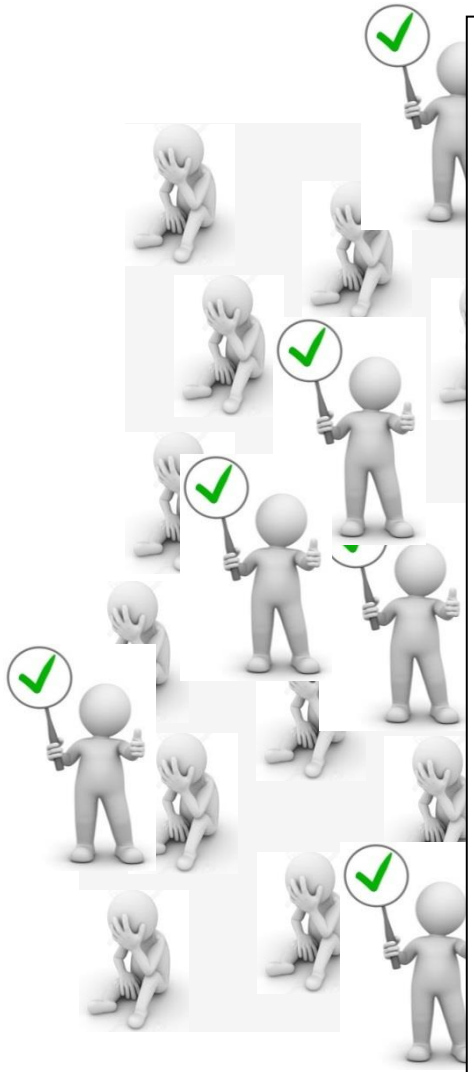
(Snyder 2014)

Redundancy – Informativity ?

Feature selection

- In machine learning contexts, our goal may be to reduce the number of dimensions without losing “predictive” power:
  - exhaustive search
  - random feature selection
  - minimum redundancy maximum relevance feature selection
  - simultaneous learning feature representation and cluster assignment using deep learning network (Xie et al. 2016)
- Machine learning *can* benefit from redundancy
  - leading to increased performance by adding robustness (Lorenzen 1999)

TIER 2: Stratified medicine



Molecular profiling

OPEN ACCESS Freely available online

PLoS one

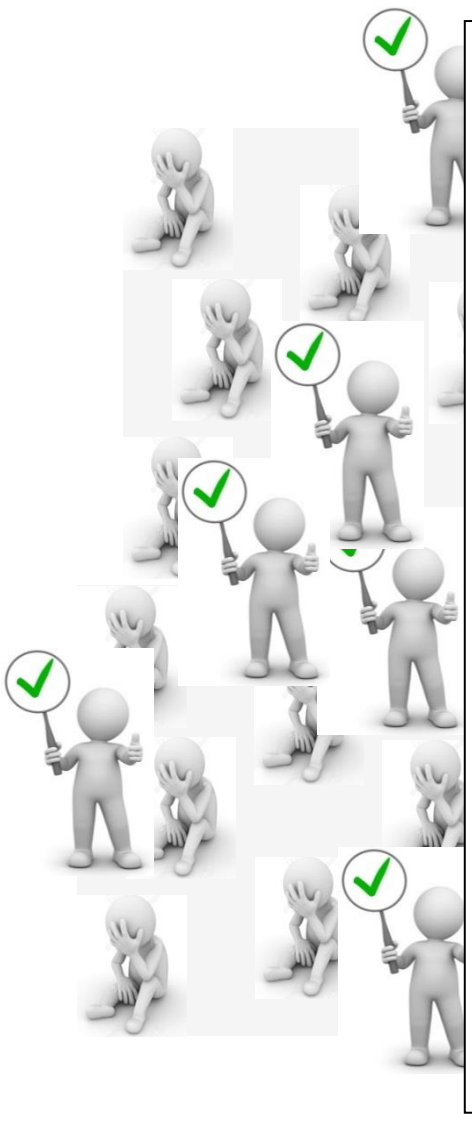
Molecular Reclassification of Crohn's Disease by Cluster Analysis of Genetic Variants

Isabelle Cleynen[1]*, Jestinah M. Mahachie John[2,3], Liesbet Henckaerts[4], Wouter Van Moerkercke[1], Paul Rutgeerts[1], Kristel Van Steen[2,3], Severine Vermeire[1]

1 Department of Gastroenterology, KU Leuven, Leuven, Belgium, **2** Systems and Modeling Unit, Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium, **3** Bioinformatics and Modeling, GIGA-R, University of Liège, Liège, Belgium, **4** Department of Medicine, UZ Leuven, Leuven, Belgium

(Cleynen et al. 2012)

Heterogeneity as a target



Molecular profiling

OPEN ACCESS Freely available online

PLOS ONE

Molecular Reclassification of Crohn's Disease: A Cautionary Note on Population Stratification

Bärbel Maus[1,2*], Camille Jung[3,4,5], Jestinah M. Mahachie John[1,2], Jean-Pierre Hugot[3,4,6], Emmanuelle Génin[7,8], Kristel Van Steen[1,2]

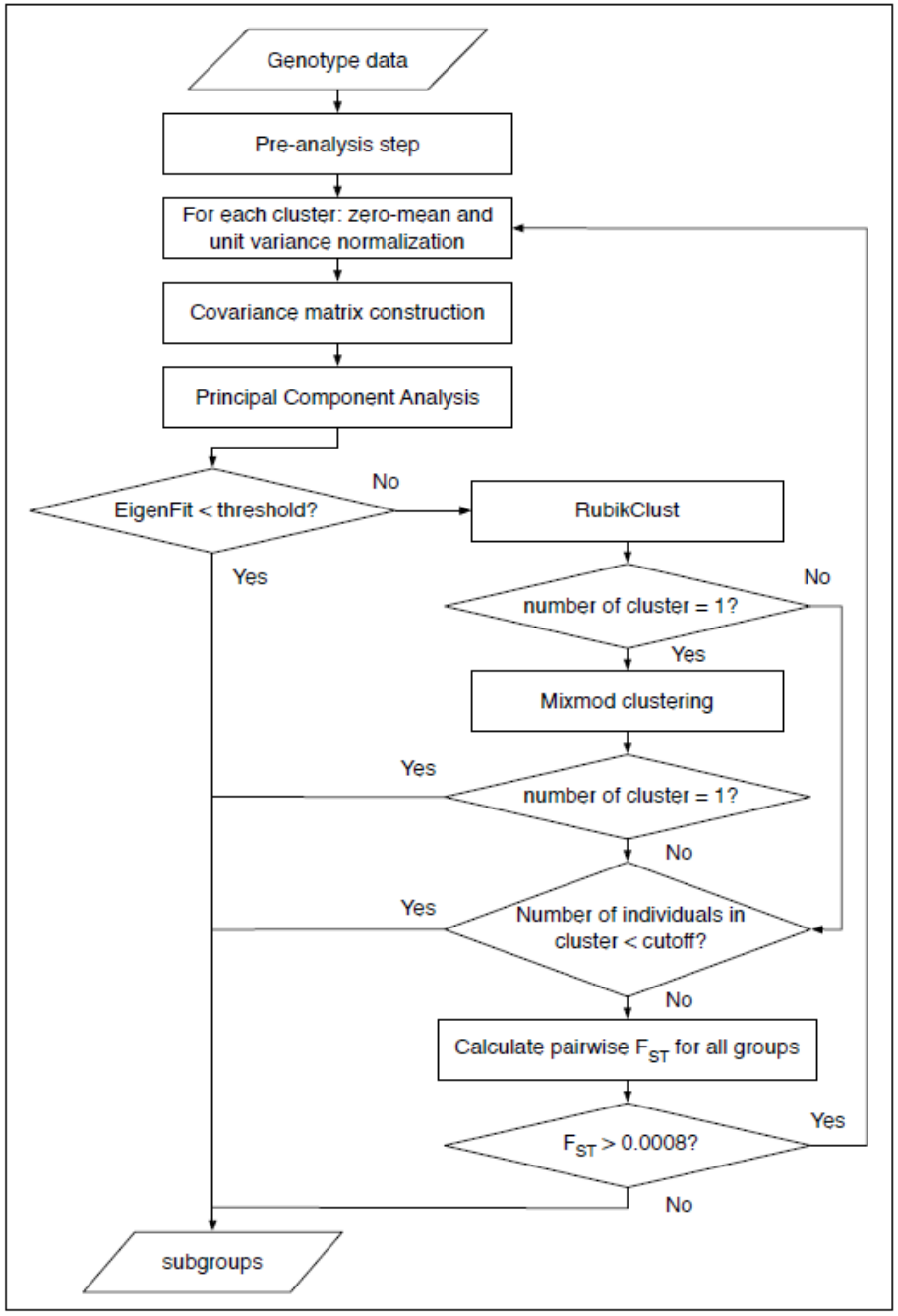
1 UMR843, INSERM, Paris, France, **2** Bioinformatics and Modeling, GIGA-R, University of Liège, Liège, Belgium, **3** UMR843, Institut National de la Sante et de la recherche Medicale, Paris, France, **4** Service de Gastroentérologie Pédiatrique, Hôpital Robert Debré, APHP, Paris, France, **5** CRC-CRB, CHI Creteil, Creteil, France, **6** Labex Inflamex, Université Paris Diderot, Paris, France, **7** UMR1078, Génétique, Génomique fonctionnelle et Biotechnologies, INSERM, Brest, France, **8** Centre Hospitalier Régional Universitaire de Brest, Brest, France

(Maus et al. 2013)

Heterogeneity <u>as a target and a nuisance</u>

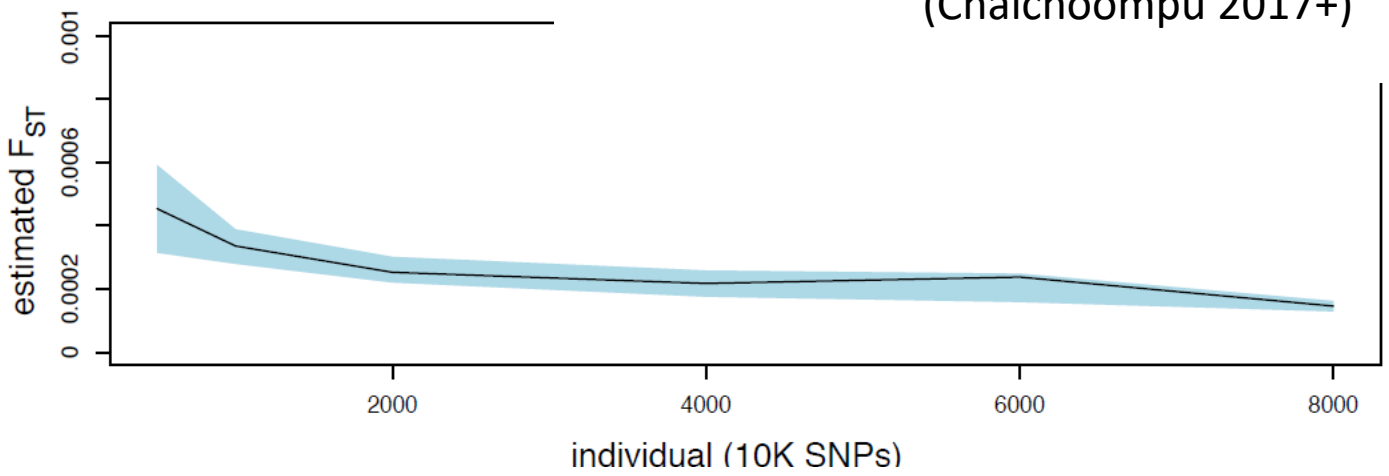
IPCAPS workflow

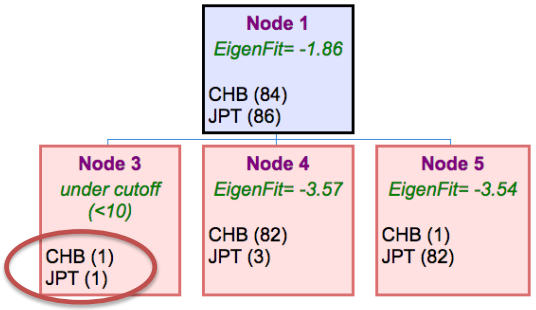
(Chaichoompu 2017+)

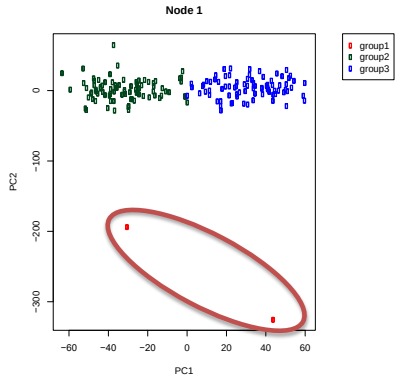


Fine-scale structure detection

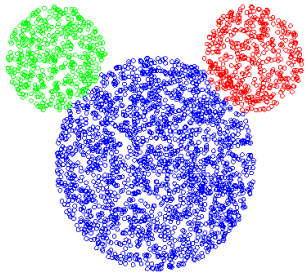
(Chaichoompu 2017+)



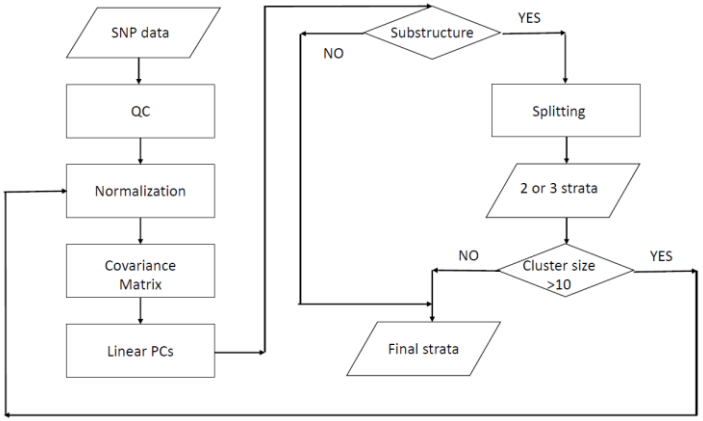




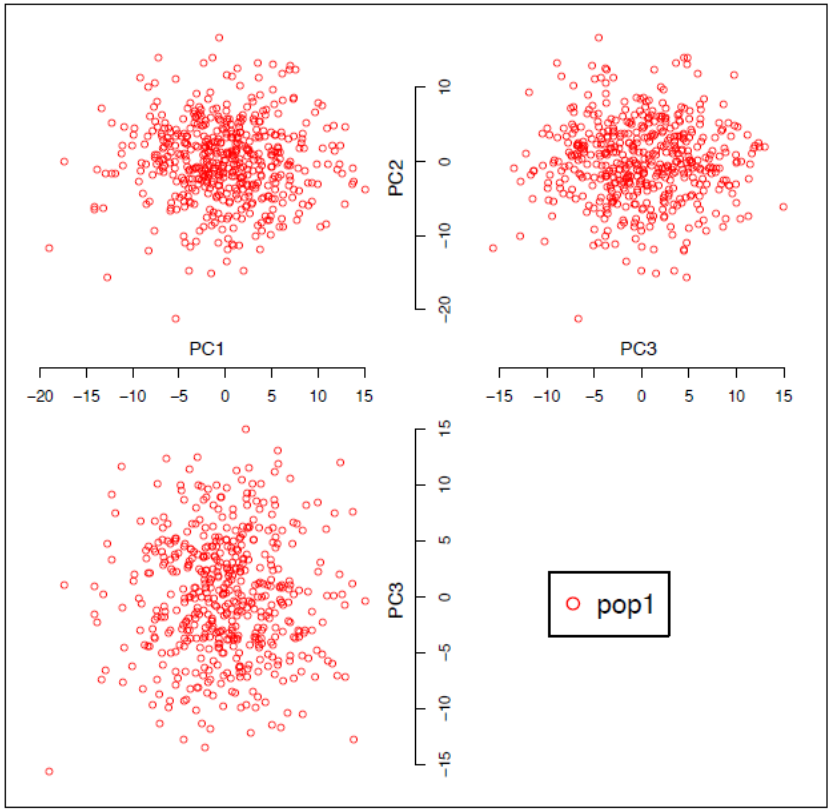
Combine with EM clustering



IPCAPS



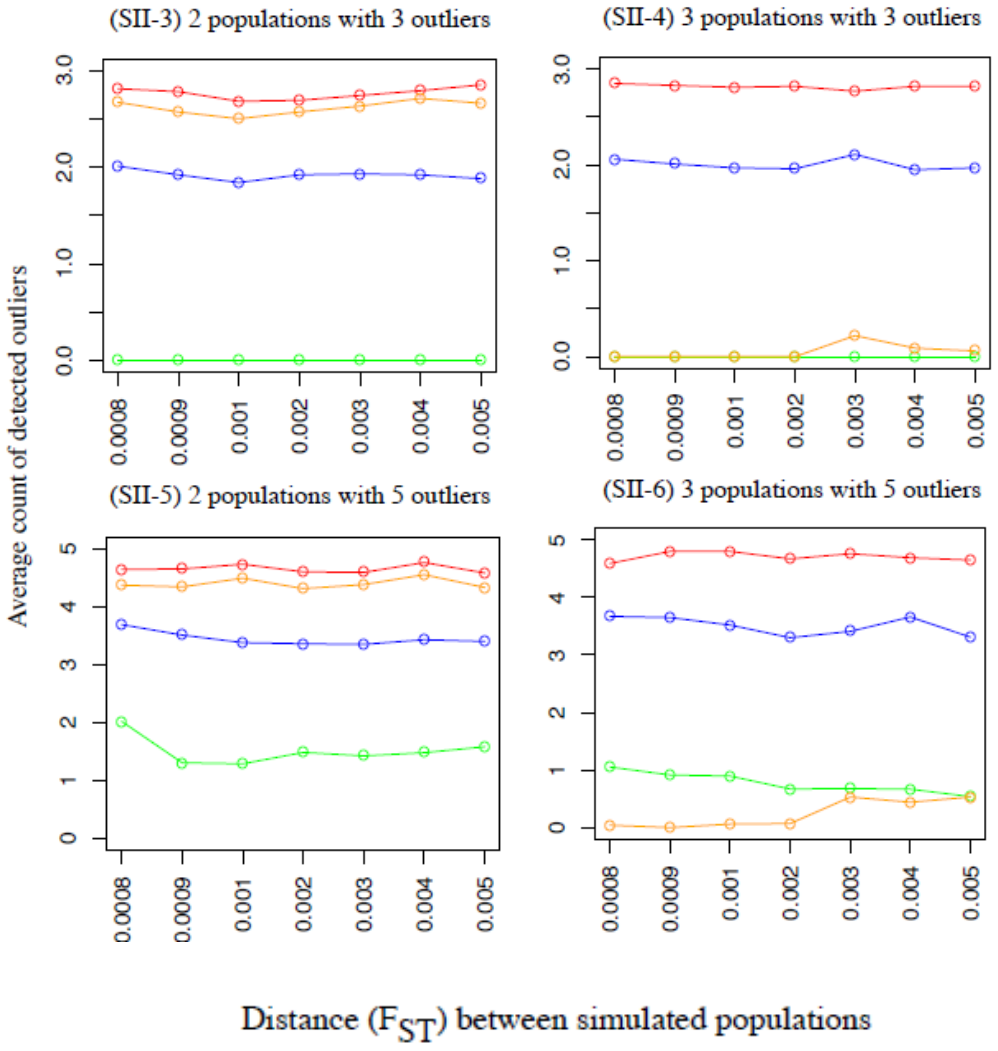
Type I error of IPCAPS



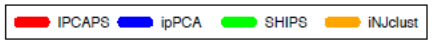
Method	Av. # clusters
IPCAPS	1
ipPCA	2
SHIPS	1
iNJclust	>150

(Kridsadakorn Chaichoompu 2017,
PhD thesis – Chapter 2)

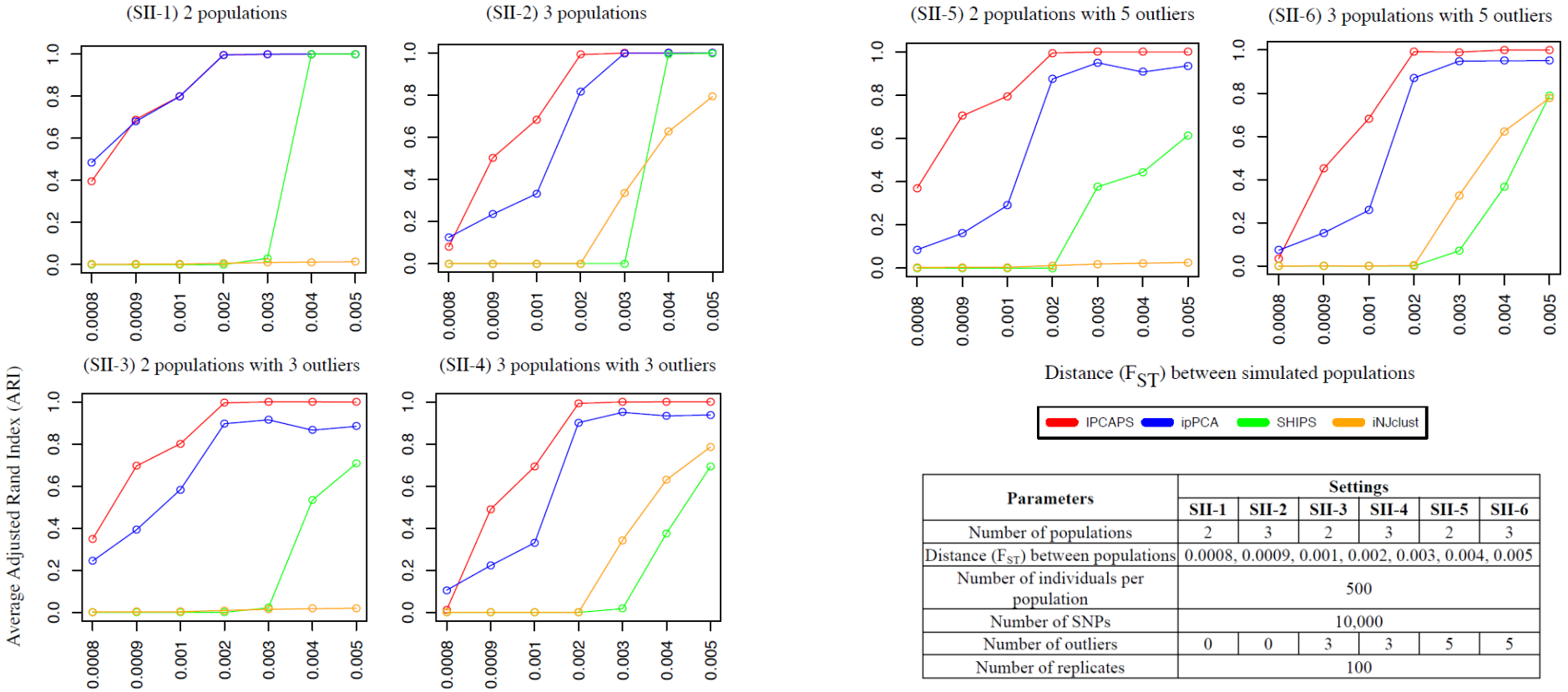
Performance of IPCAPS as outlier detection tool



Parameters	Settings					
	SII-1	SII-2	SII-3	SII-4	SII-5	SII-6
Number of populations	2	3	2	3	2	3
Distance (F_{ST}) between populations	0.0008, 0.0009, 0.001, 0.002, 0.003, 0.004, 0.005					
Number of individuals per population	500					
Number of SNPs	10,000					
Number of outliers	0	0	3	3	5	5
Number of replicates	100					



Accuracy of IPCAPS as a clustering technique



Parameters	Settings					
	SII-1	SII-2	SII-3	SII-4	SII-5	SII-6
Number of populations	2	3	2	3	2	3
Distance (F_{ST}) between populations	0.0008, 0.0009, 0.001, 0.002, 0.003, 0.004, 0.005					
Number of individuals per population	500					
Number of SNPs	10,000					
Number of outliers	0	0	3	3	5	5
Number of replicates	100					

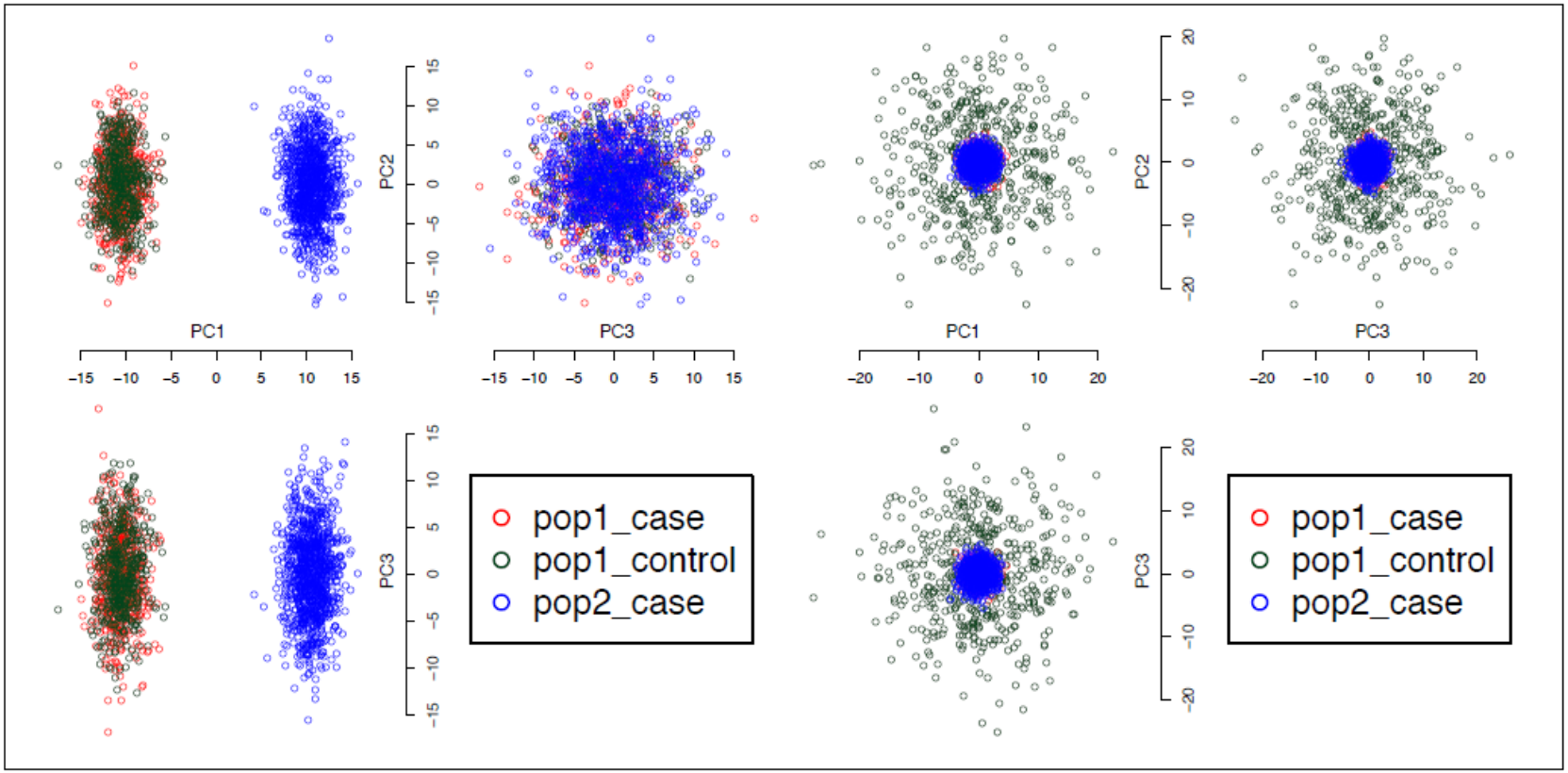
(Chaichoompu 2017+)

F_{ST} among some populations

	Sp	Fr	Be	UK	Sw	No	Ge	Ro	Cz	Sl	Hu	Po	Ru	CEU	CHB	JPT
Fr	0.0008															
Be	0.0015	0.0002														
UK	0.0024	0.0006	0.0005													
Sw	0.0047	0.0023	0.0018	0.0013												
No	0.0047	0.0024	0.0019	0.0014	0.0010											
Ge	0.0025	0.0008	0.0005	0.0006	0.0011	0.0016										
Ro	0.0023	0.0017	0.0018	0.0028	0.0041	0.0044	0.0016									
Cz	0.0033	0.0016	0.0013	0.0014	0.0016	0.0024	0.0003	0.0016								
Sl	0.0034	0.0017	0.0015	0.0017	0.0019	0.0026	0.0005	0.0014	0.0001							
Hu	0.0030	0.0015	0.0013	0.0016	0.0020	0.0026	0.0004	0.0011	0.0001	0.0001						
Po	0.0053	0.0032	0.0028	0.0027	0.0023	0.0034	0.0012	0.0028	0.0004	0.0004	0.0006					
Ru	0.0059	0.0037	0.0034	0.0032	0.0025	0.0036	0.0016	0.0030	0.0008	0.0007	0.0009	0.0003				
CEU	0.0026	0.0008	0.0005	0.0002	0.0011	0.0012	0.0006	0.0028	0.0014	0.0016	0.0016	0.0026	0.0031			
CHB	0.1096	0.1094	0.1093	0.1096	0.1073	0.1081	0.1085	0.1047	0.1080	0.1069	0.1058	0.1086	0.1036	0.1095		
JPT	0.1118	0.1116	0.1114	0.1117	0.1095	0.1103	0.1107	0.1068	0.1102	0.1091	0.1079	0.1108	0.1057	0.1117	0.0069	
YRI	0.1460	0.1493	0.1496	0.1513	0.1524	0.1531	0.1502	0.1463	0.1503	0.1498	0.1490	0.1520	0.1504	0.1510	0.1901	0.1918

(Heath et al. 2008)

Linear population structure correction (Chaichoompu 2017+)



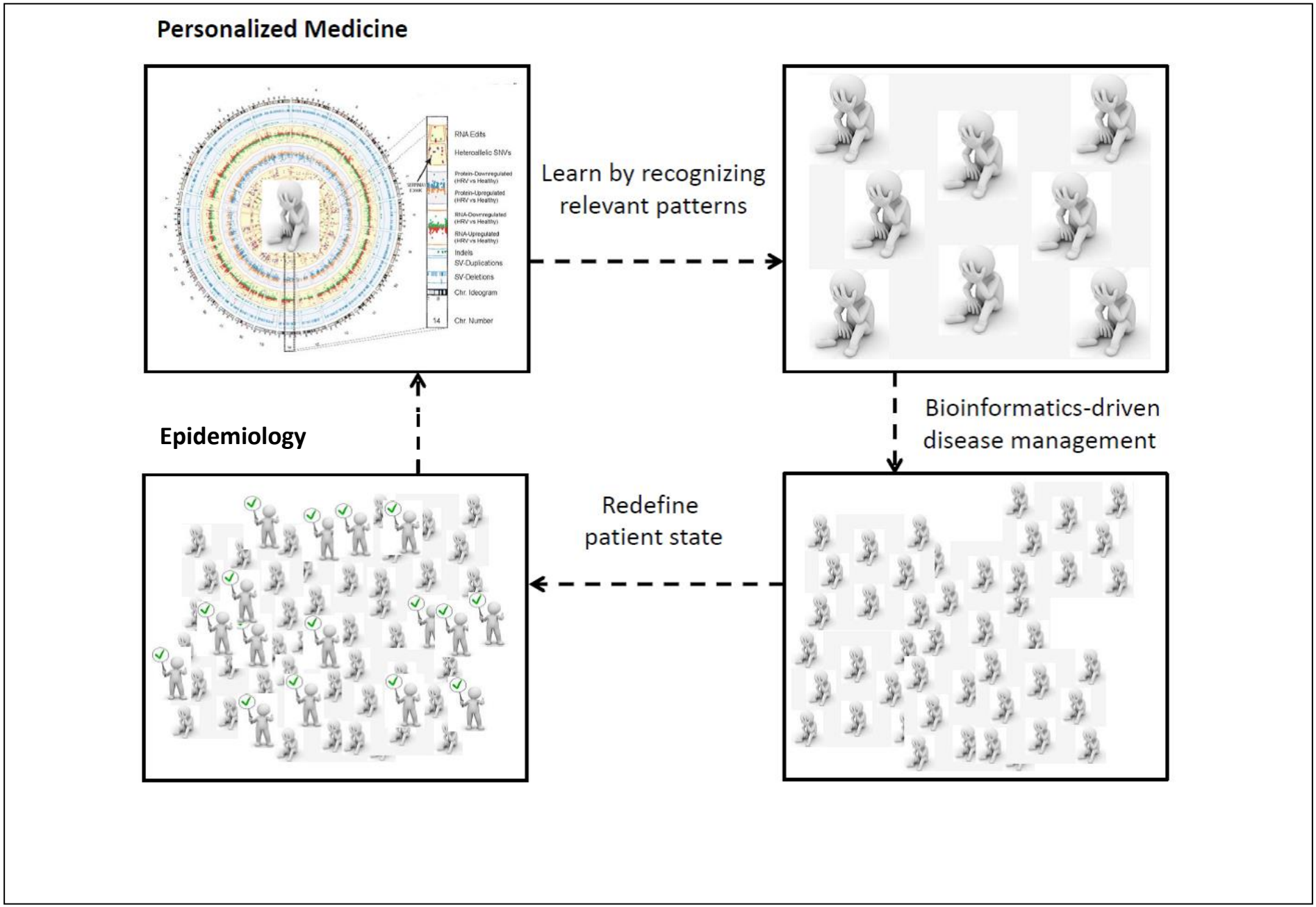
Pooled case/control PCs (left) vs Case-Projected PCs (right)

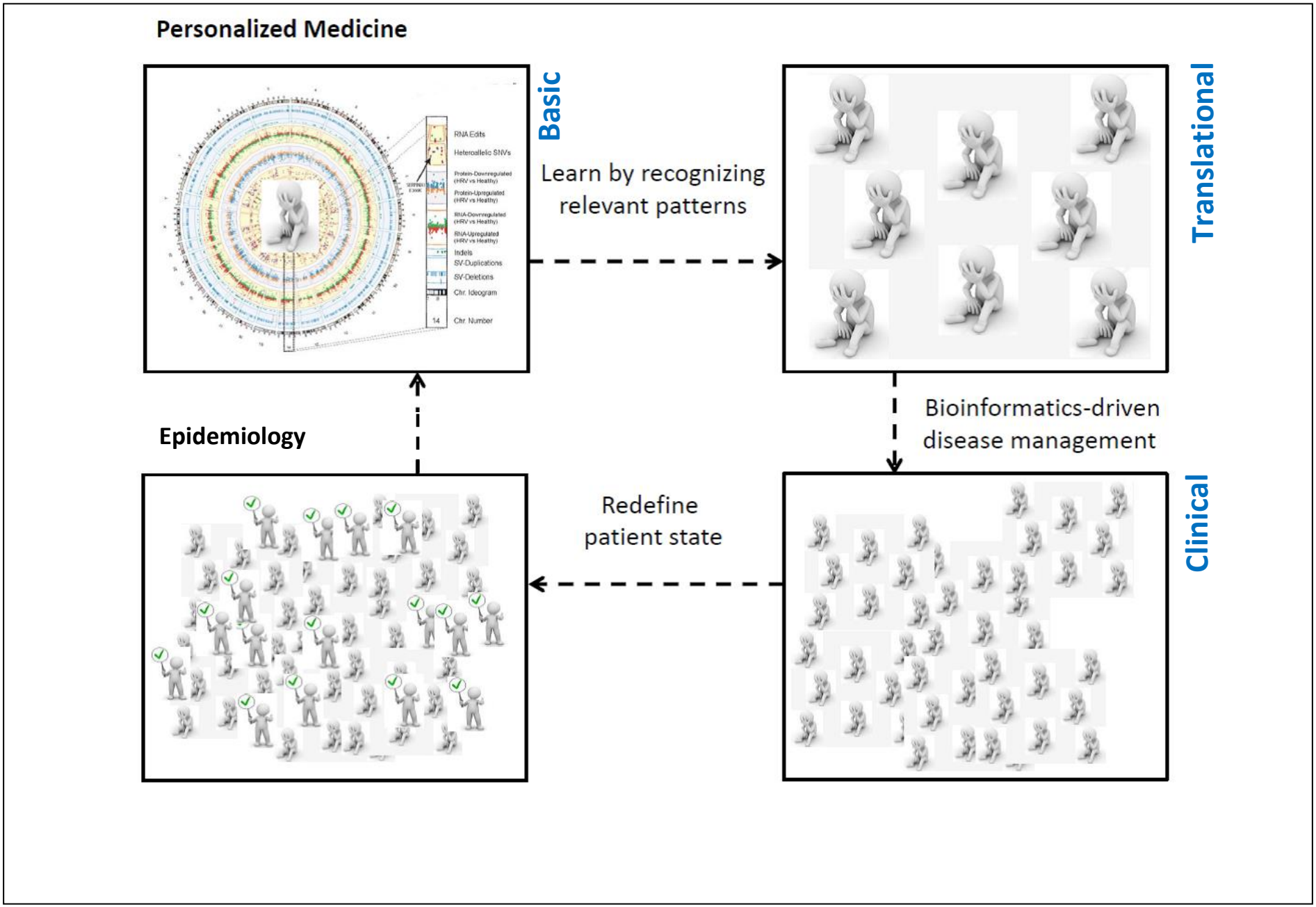
Pooled PCs but on which SNPs? (Chaichoompu 2017+)

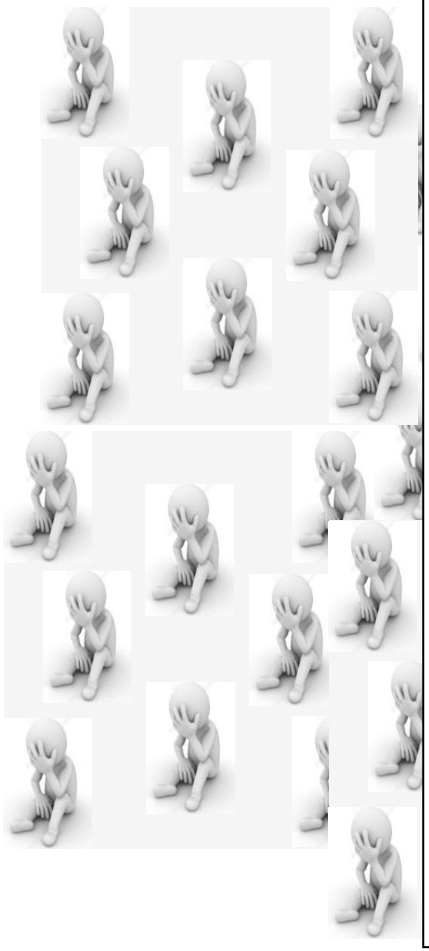
Set	Uncorrected CON		CON		CD		UC	
	Dis.	Rep.	Dis.	Rep.	Dis.	Rep.	Dis.	Rep.
1	5	4	1	1	3	8	3	3
2	3	5	1	1	3	5	3	3
3	5	5	1	1	3	3	3	5
4	5	5	1	1	3	3	3	3
5	5	5	1	1	3	5	3	3
6	5	4	1	1	3	3	3	3
7	6	5	1	1	3	3	3	3
8	6	4	1	1	6	3	3	3
9	4	4	1	1	3	8	3	5
10	4	5	1	1	6	5	3	3
Average	4.8	4.6	1.0	1.0	3.6	4.6	3.0	3.4

(cluster sizes less than 20 are considered to be outlying and are removed)

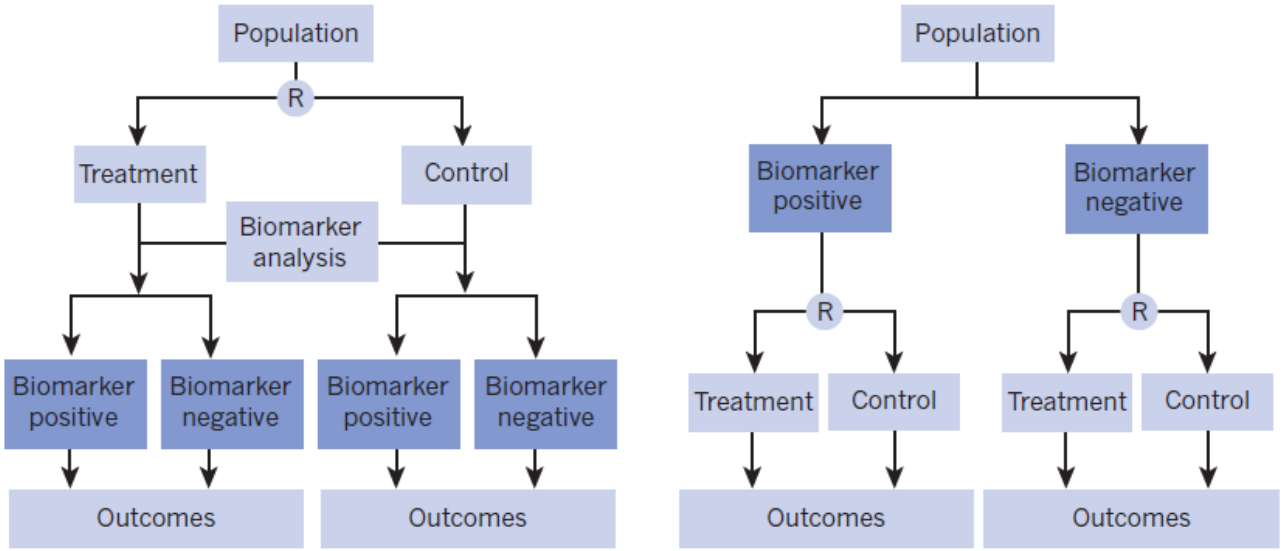
Clinical science


Personalized Medicine
RNA Edits
Heteroallelic SNVs
Protein-Downregulated (HRV vs Healthy)
Protein-Upregulated (HRV vs Healthy)
RNA-Downregulated (HRV vs Healthy)
RNA-Upregulated (HRV vs Healthy)
Indels
SV-Duplications
SV-Deletions
Chr. Ideogram
14
Chr. Number
Learn by recognizing relevant patterns
Bioinformatics-driven disease management
Redefine patient state
Epidemiology

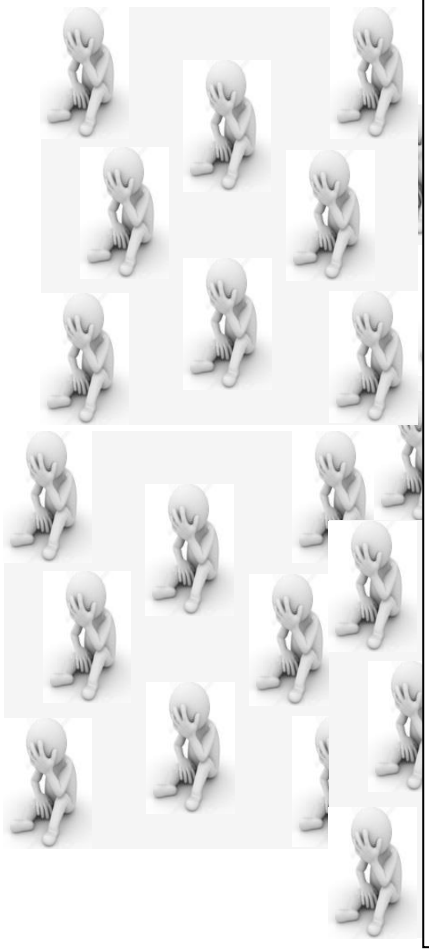

Personalized Medicine
Basic
RNA Edits
Heteroallelic SNVs
Protein-Downregulated (HRV vs Healthy)
Protein-Upregulated (HRV vs Healthy)
RNA-Downregulated (HRV vs Healthy)
RNA-Upregulated (HRV vs Healthy)
Indels
SV-Duplications
SV-Deletions
Chr. Ideogram
14
Chr. Number
Learn by recognizing relevant patterns
Translational
Bioinformatics-driven disease management
Clinical
Redefine patient state
Epidemiology



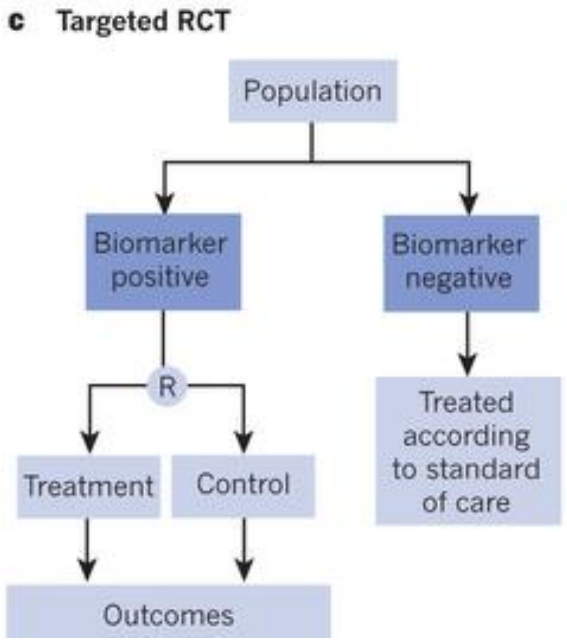
Testing precision-medicine strategies

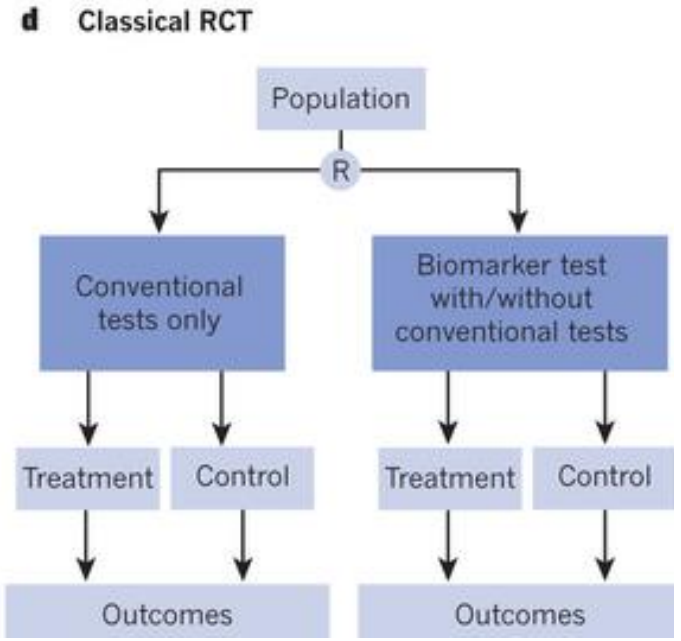


(Biankin et al. 2015)



Testing precision-medicine strategies





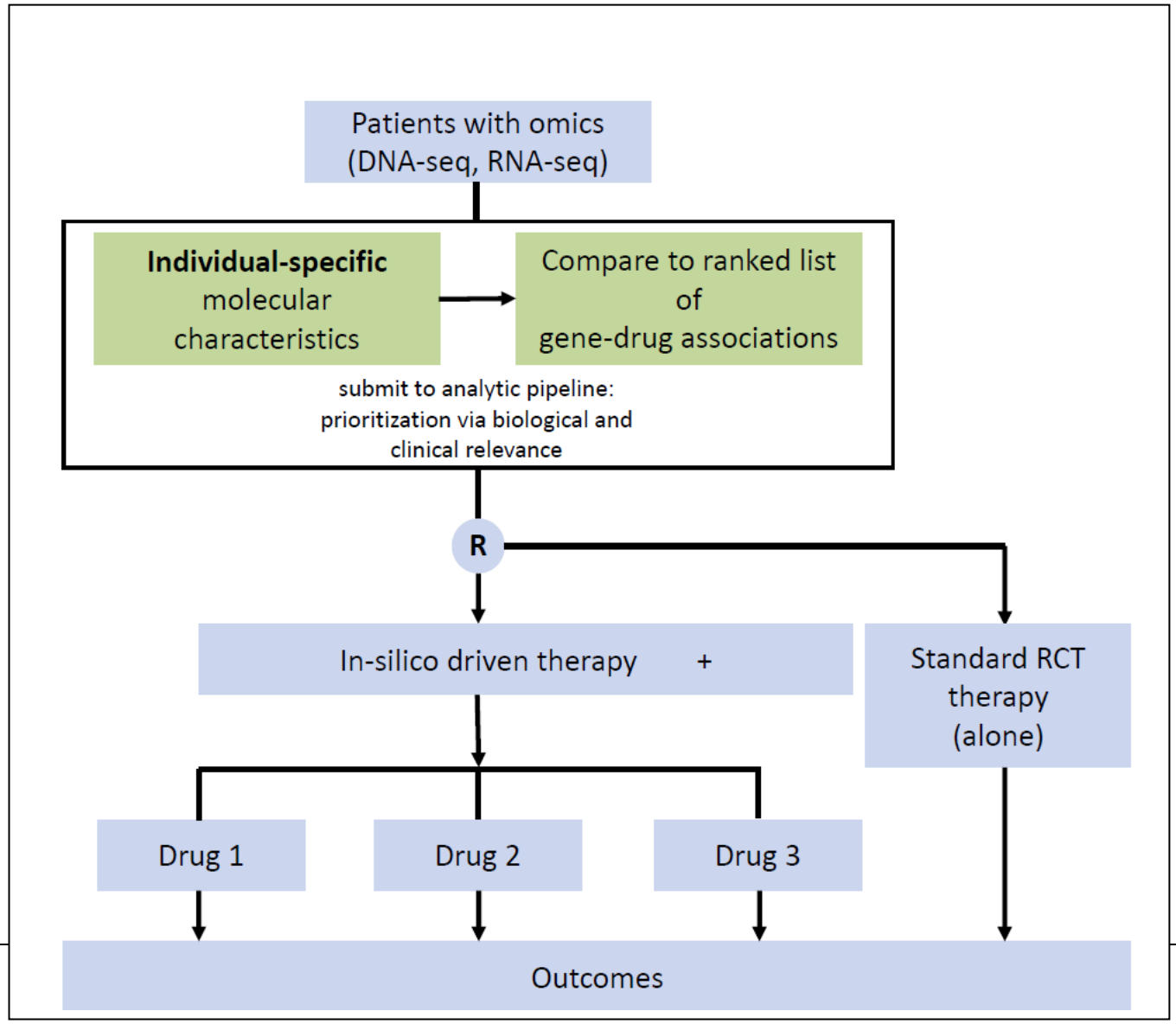
(Biankin et al. 2015)

Patient selection – where are we?

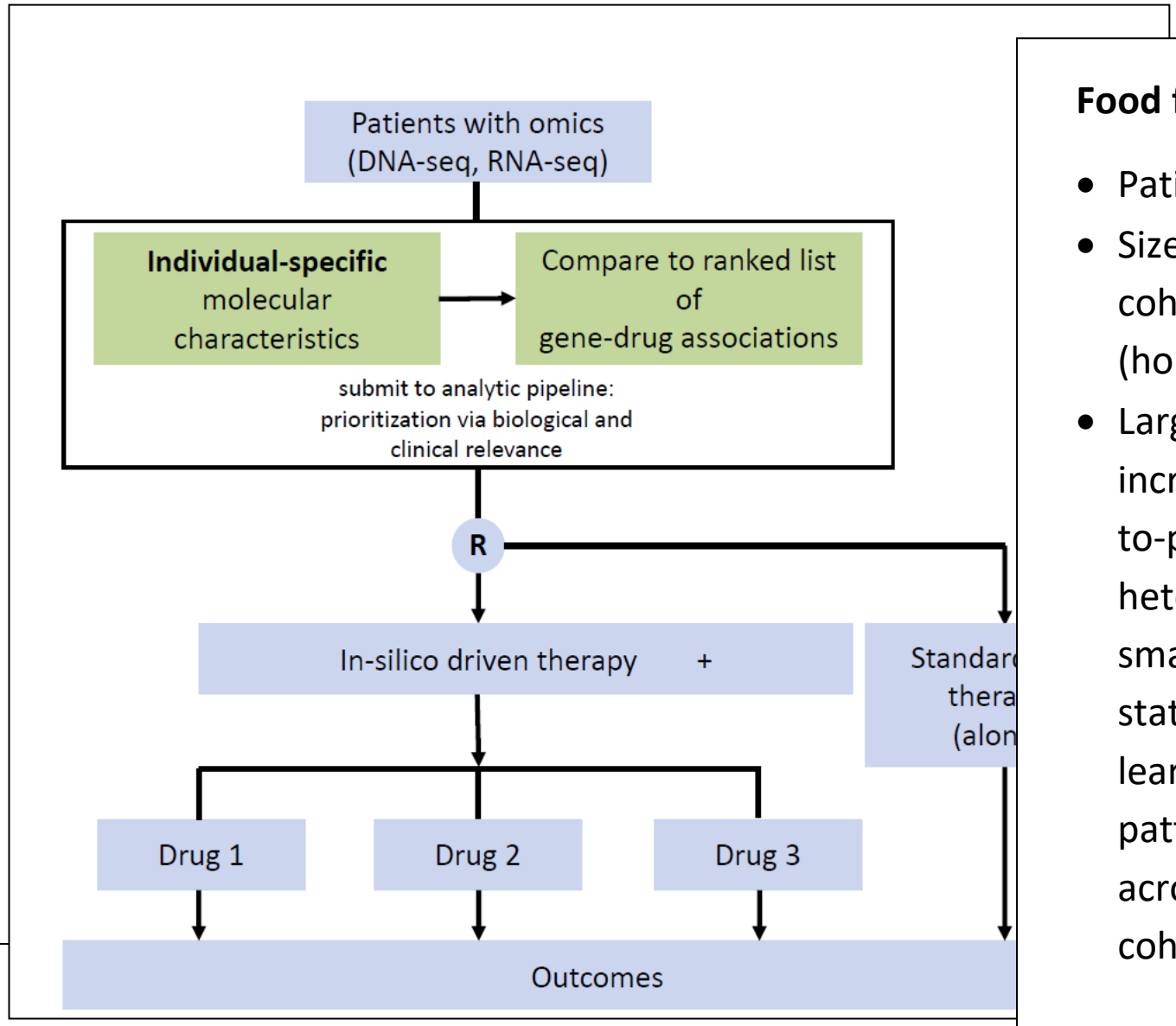
- Basket CTs: multiple diseases with the same genetic mutation, randomized treatment allocation
- Umbrella CTs: 1 “disease”, different genetic mutations which define sub-cohorts, each receiving randomized treatment regimen
- Adding complexity:
  - cellular heterogeneity - assign based on the mutation detected in the higher percentage of cancer cells?
  - highly multi-dimensional profiles are expected to lead to very small cohorts

(Sumitrhra Mandrekar,
INSERM atelier 248, Bordeaux, 2017)

CTs in view of personalized medicine – where are we going?



CTs in view of personalized medicine – where are we going?



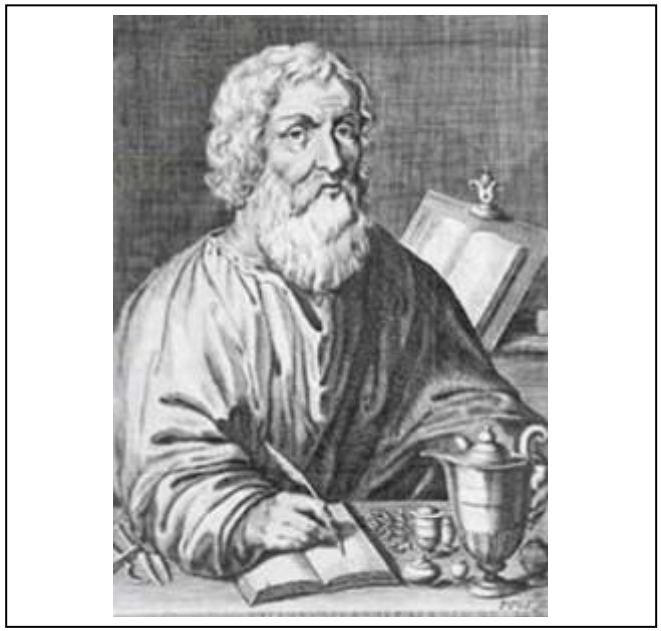
Food for thought

- Patient selection
- Size of sub-cohorts (homogeneity)
- Large CTs may increase patient-to-patient heterogeneity: small sample statistics, machine learning for pattern detection across sub-cohorts

Take-home messages

Hippocrates (460-370 BC)

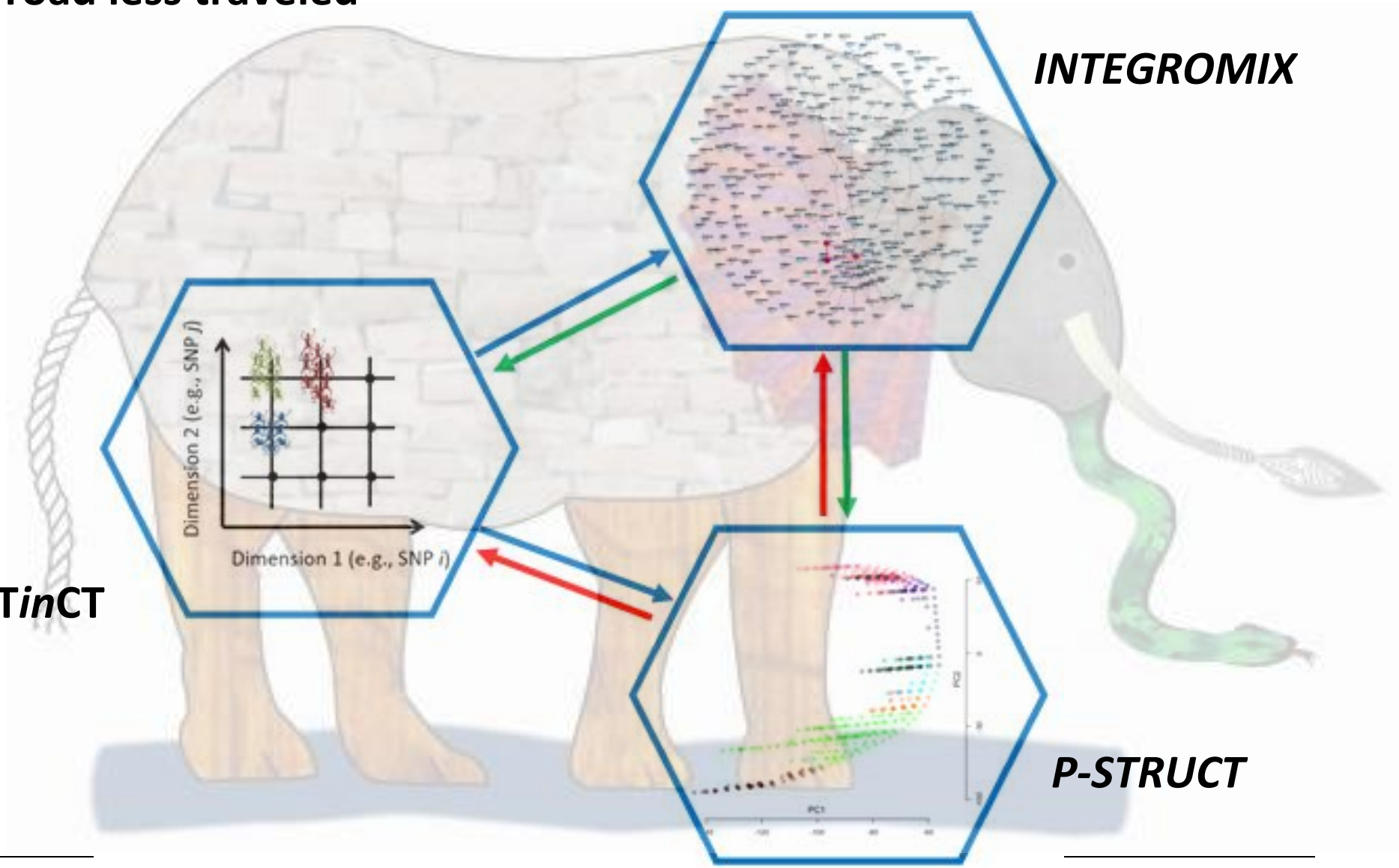
"It's far more important to know what person the disease has than what disease the person has."



Imagine a world …

- in which missing data handling strategies hold, despite data heterogeneity
- in which multi-omics summaries can be (deep-) learned from data
- in which machine learning taxonomy addresses an interdisciplinary community
- in which confounding information is adequately described or accounted for
- in which disease prediction can be extended to accommodate a latent spectrum of diseases or a continuum of disease presentations
- in which neural network parameters aid in deriving meaningful/relevant relationships

The road less traveled



Acknowledgements

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

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