Towards Clinical Trials in PM

- a systems genomics viewpoint -

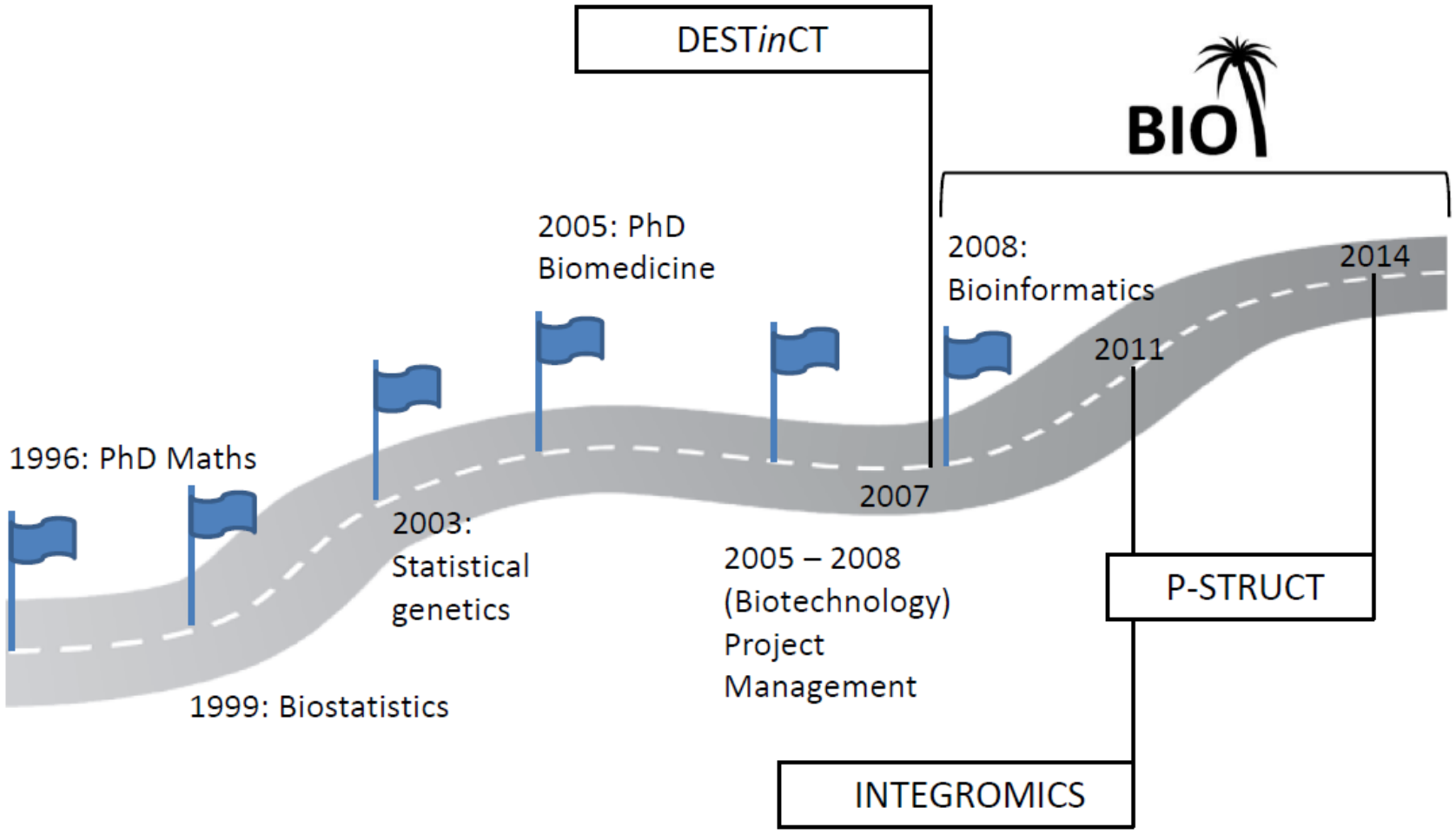
Kristel Van Steen, PhD[2] (*)

kristel.vansteen@uliege.be

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

Systems Medicine Lab, KU Leuven, Belgium

The road less traveled



OUTLINE

- **Personalized Medicine**
- **Systems**
- **Biomarker Sciences**
  - Basic Science – How do things work?
  - Translational Science – Turning knowledge into sth useful?
  - Clinical Science – Is it really useful?
- **Take-home messages**

Personalized Medicine

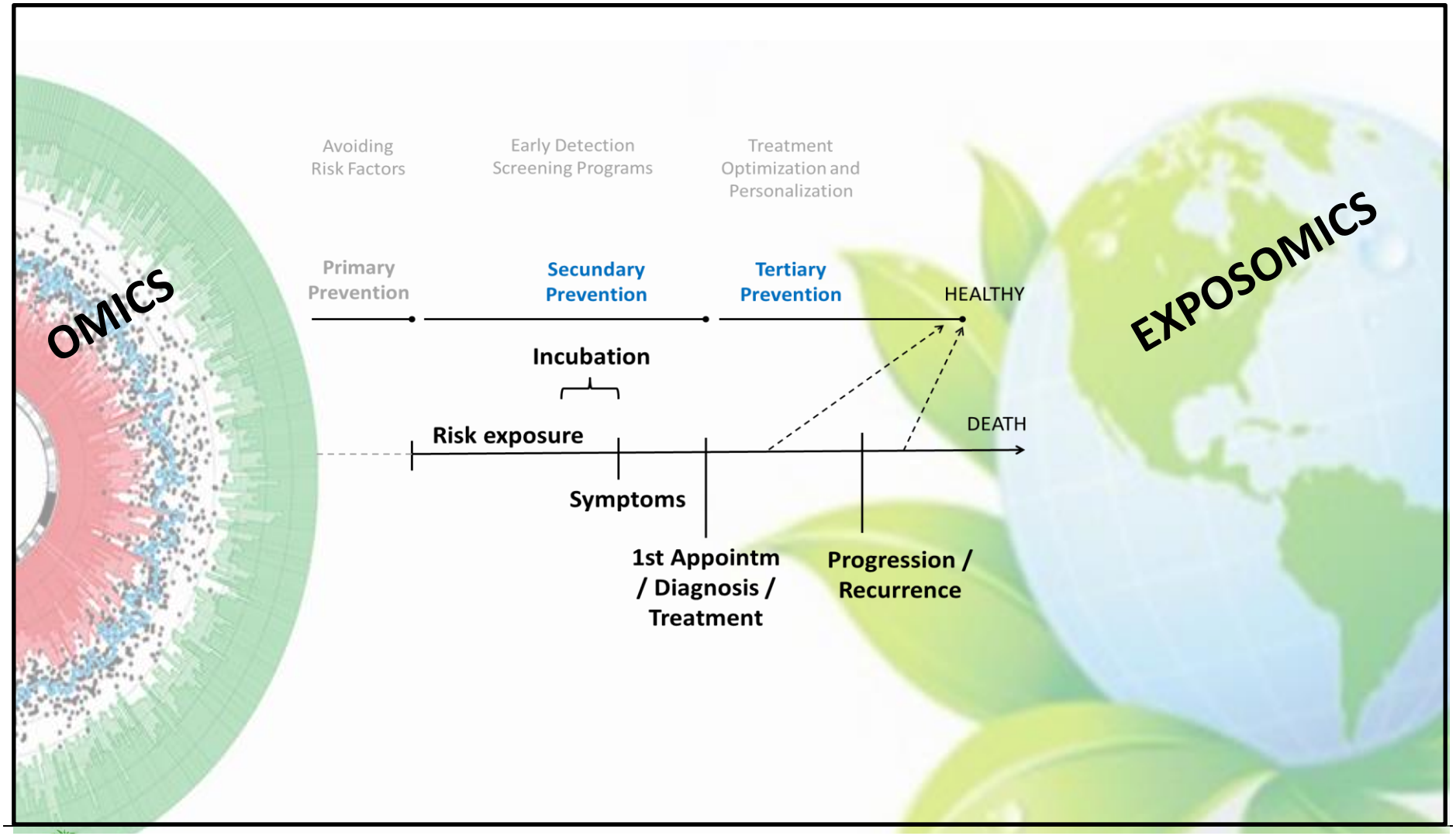
The context of precision medicine

Precision medicine is …

"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."

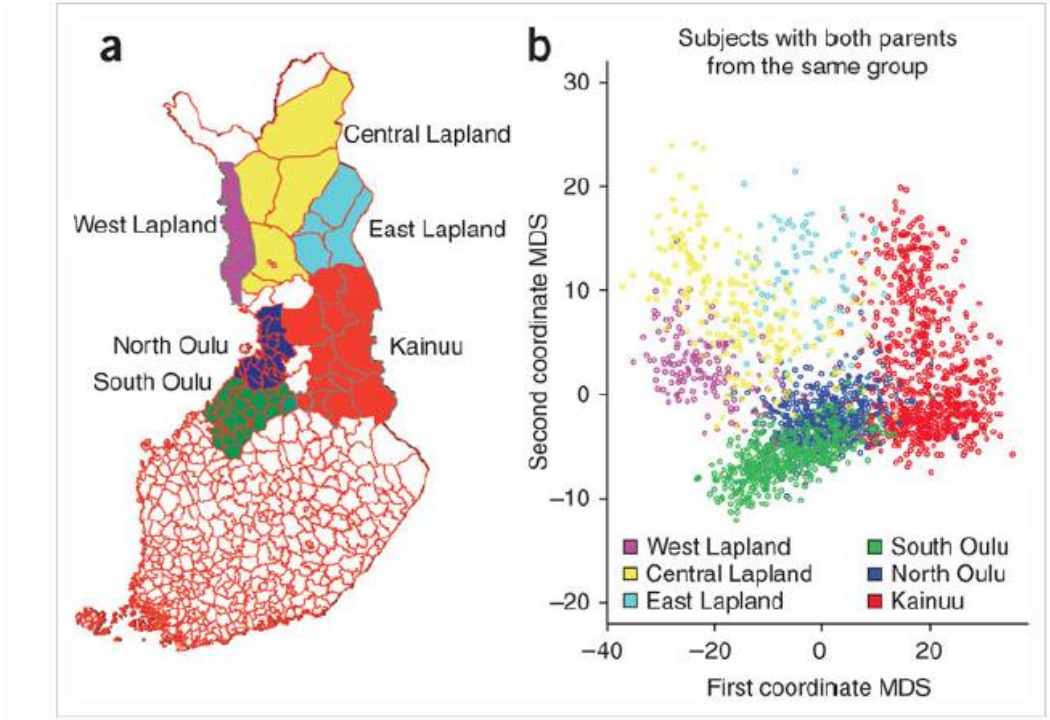
(HORIZON2020 Advisory Group;
EU Health Ministers – December 2015)

The more data, the better to characterize a patient *(subtyping)*



The more SNPs, the better to characterize a population

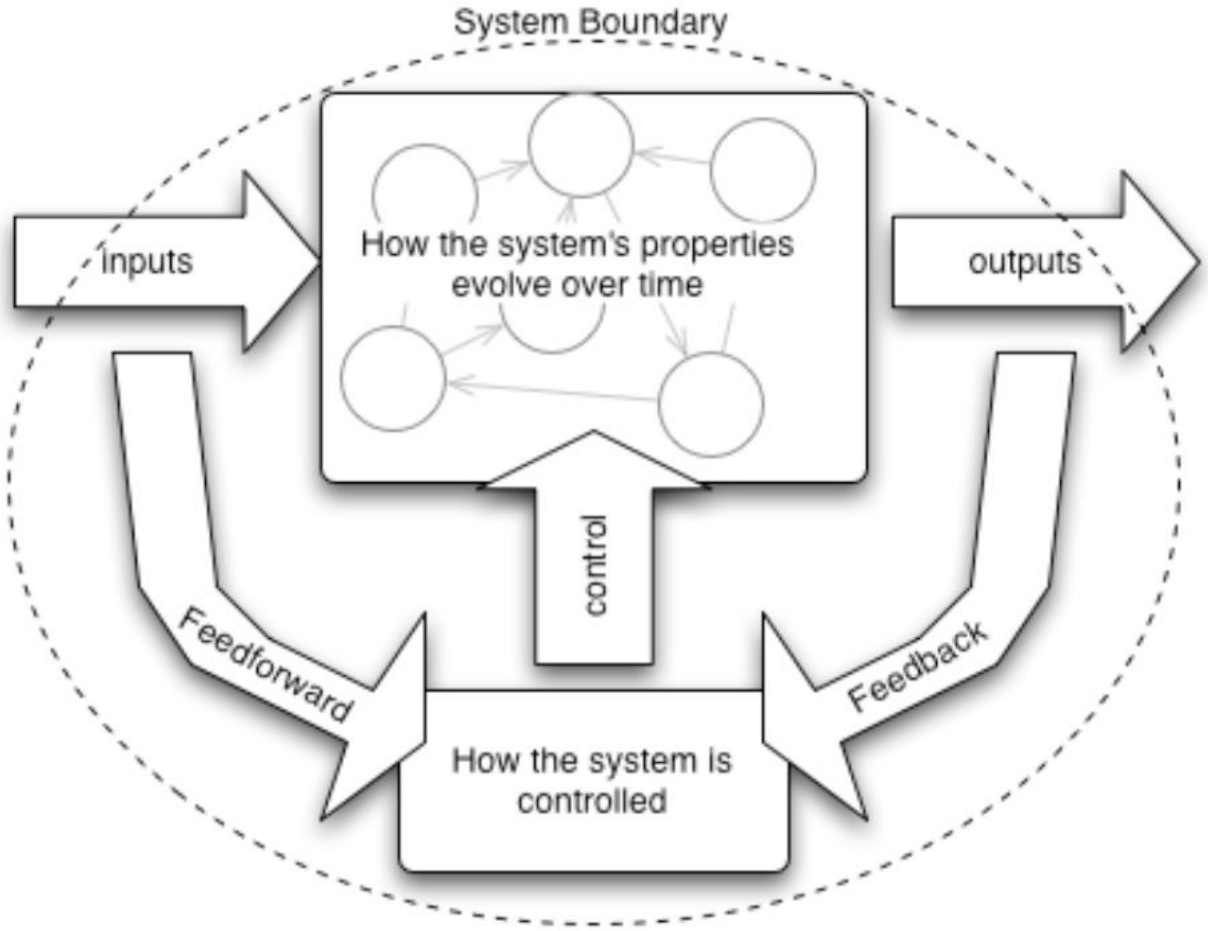
- There can be population structure in all populations, even those that appear to be relatively “homogeneous”



(Sabatti et al. 2009)

Systems

Systems and their eco-system



(@2004-5 Steve Easterbrook)

Interactome

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

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!

Studying 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

Studying interactions with human SNPs

- 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 epistasis 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.

Studying epistasis at the DNA level – lessons learned

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]

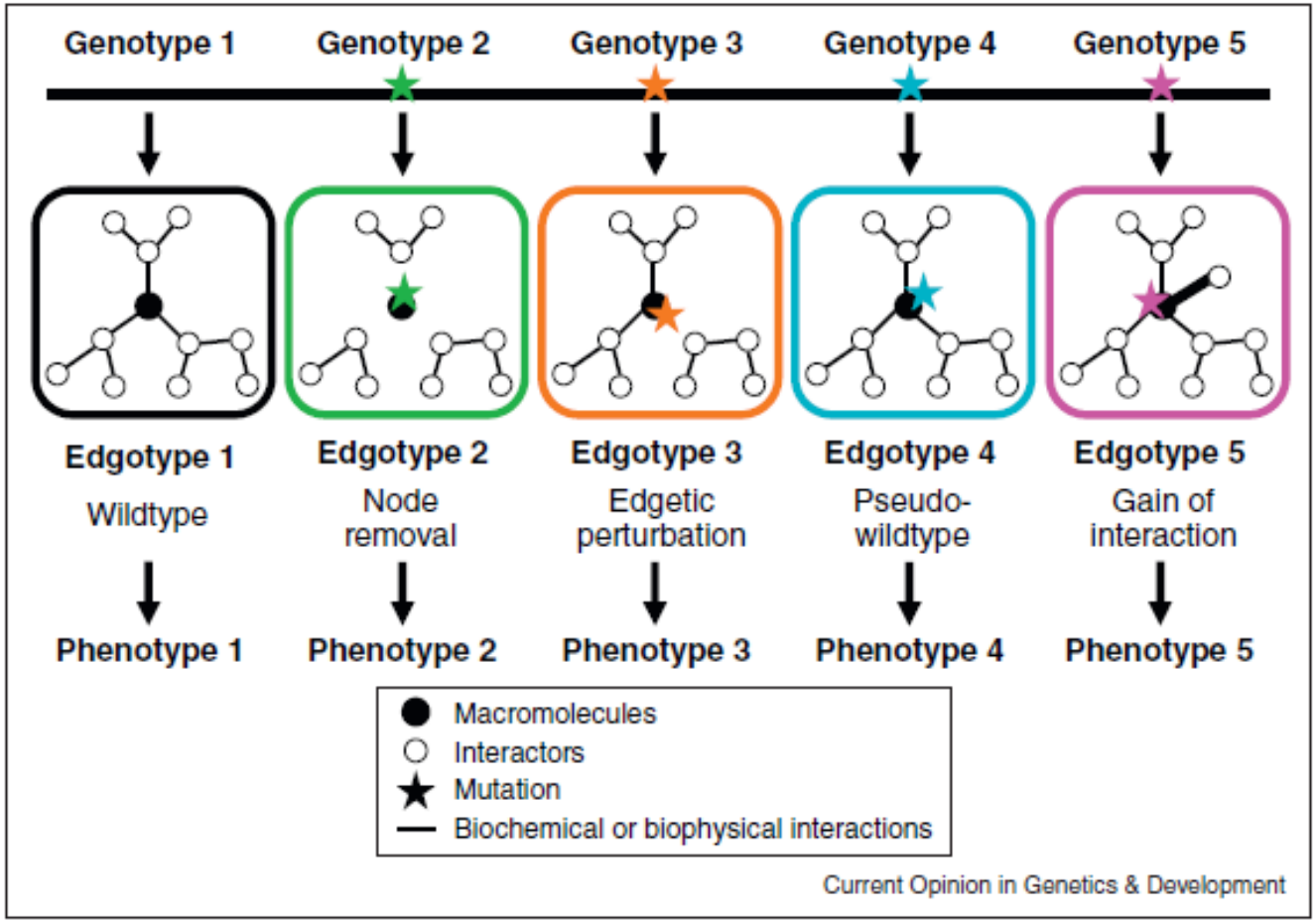
Received: 26 July 2018 / Accepted: 20 February 2019 / Published online: 6 March 2019

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.

Confirmed information in "the edges" (Filteau et al. 2015)

- Zhong et al. 2009: Interactomes of 29 mutant alleles of genes implicated in five human Mendelian disorders. Each allele was cloned and its interactions tested with ~ 8100 other proteins (*edgetics*)
  - ≠ mutant alleles of the same protein caused ≠ perturbations on their PPI profiles
- Sahni et al. 2015: ~ 2500 mutant alleles of proteins implicated in human diseases and their ~1000 corresponding wild-type proteins; Y2H to test the interactions of these proteins with a set of ~ 7200 human ORFs
  - ~ roughly 1/3rd of the mutations, the cause of the disease may result from perturbations of PPIs

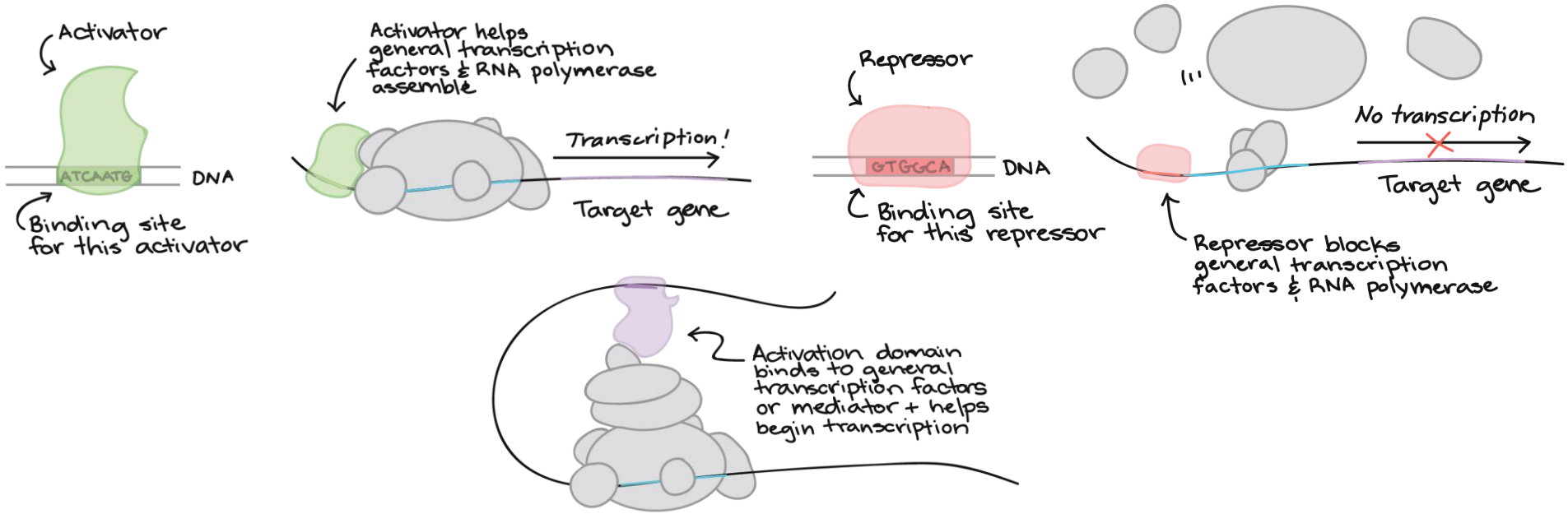
Information is in “the edges” - knowledge boosting via PPIs



(Sahni et al. 2013: *edgotypes*)

"Message passing" between layers of information

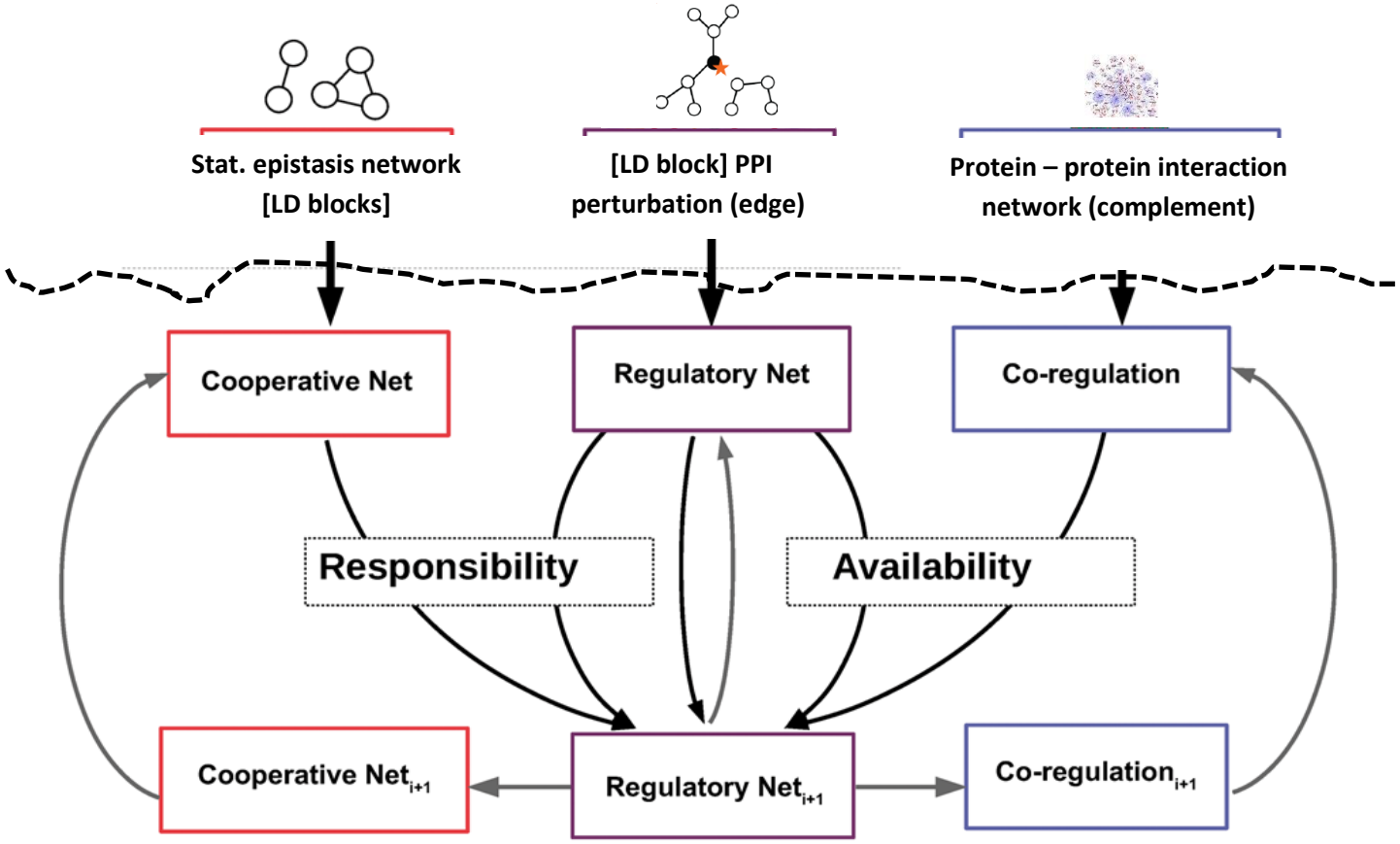
- Example: Transcription Factors (TFs)



TF interaction network (TranSignalTM TF ProteinArray) ⟷ Gene co-expression network

TRRUSTv2 (Han et al. 2018)

Knowledge boosting via PPI perturbations [→ Vidal, Glass]



Glass et al. 2013: "PANDA searches for agreement between different data-types by using the information from each to iteratively refine predictions in the others"

Integration

- **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

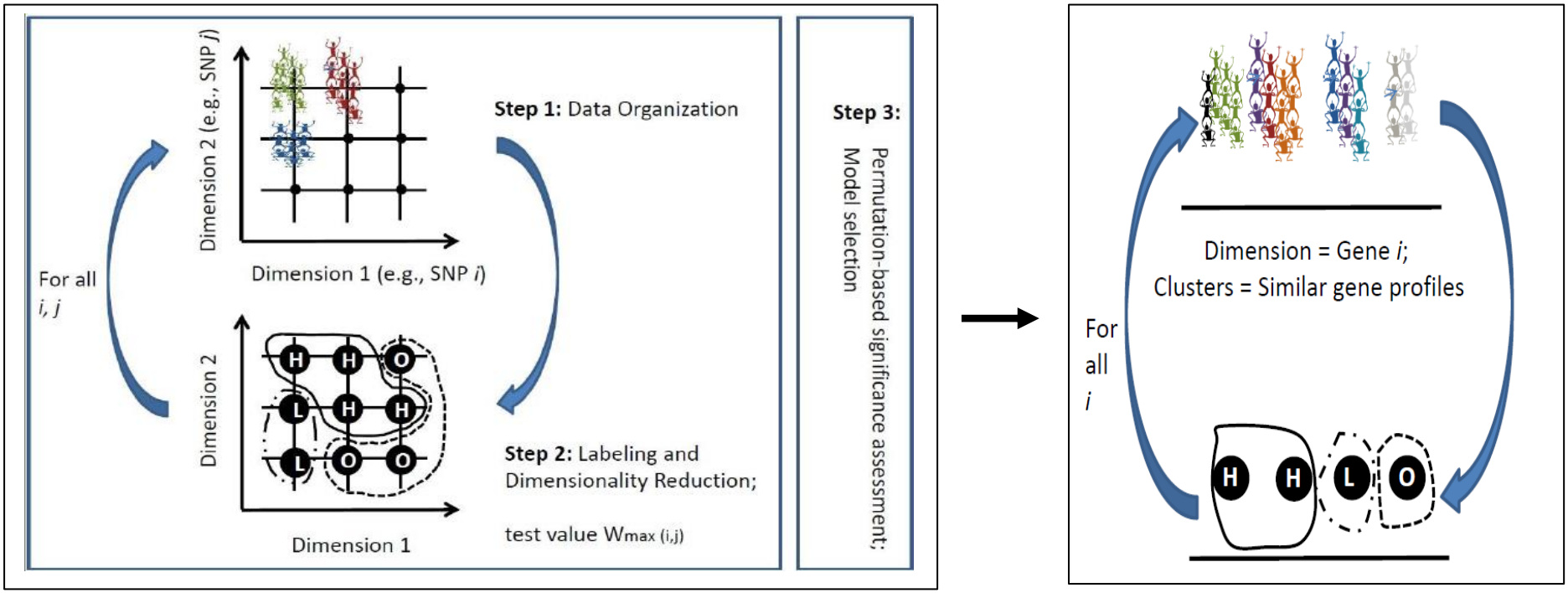
  (Van Steen, Cluj, 2015)

BIO3's approach: advanced integration in smaller systems

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

Ex: MB-MDR + new units of analyses

(e.g., SNPs allocated to genes; gene scores via kernel PCA)



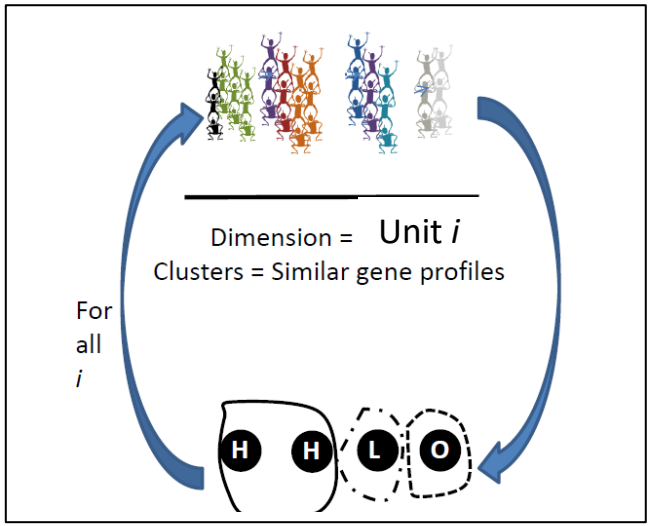
(Fouladi et al. 2015)

BIO3's approach (DEST*in*CT grant)

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

Ex: MB-MDR + revised units of analysis

to allow omics-integration for improved *disease mapping*



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

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

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

Biomarker Sciences

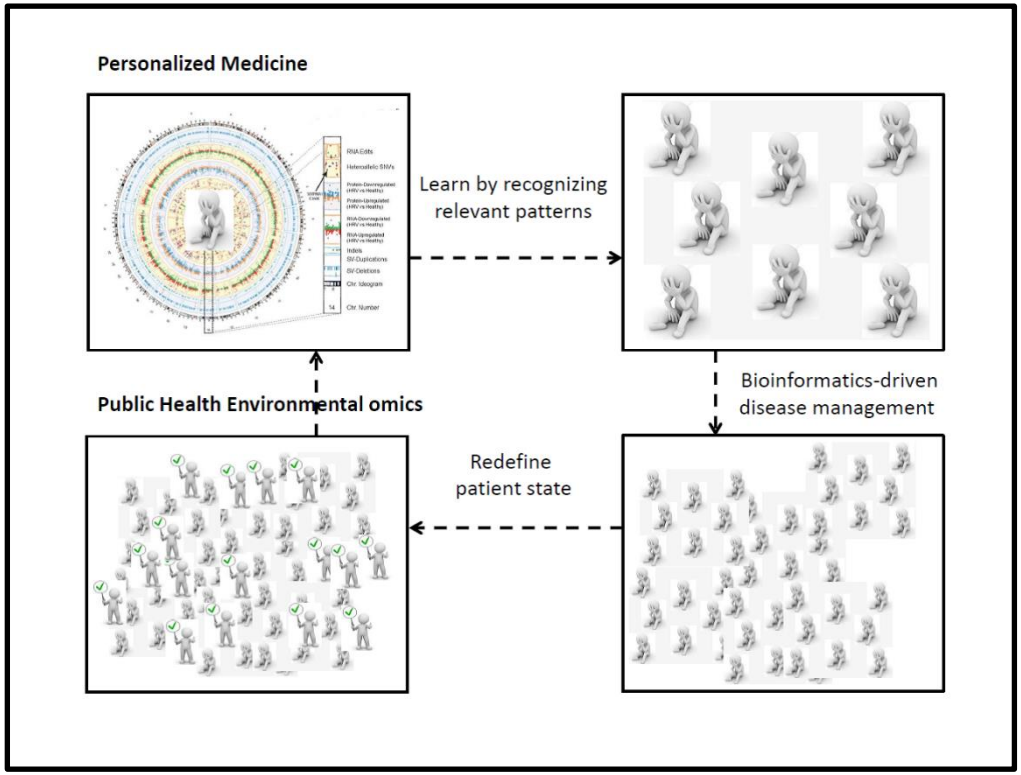
Towards CTs in PM

Molecular biomarkers to target a patient

- Integrated/Interacting molecules that can be measured and point to the presence of a disease, a physiological change, response to a treatment, or a psychological condition.

- Biomarkers are used in different ways at different stages of medicines development

(www.eupati.eu)



Biomarker-related "Sciences"

Basic Science

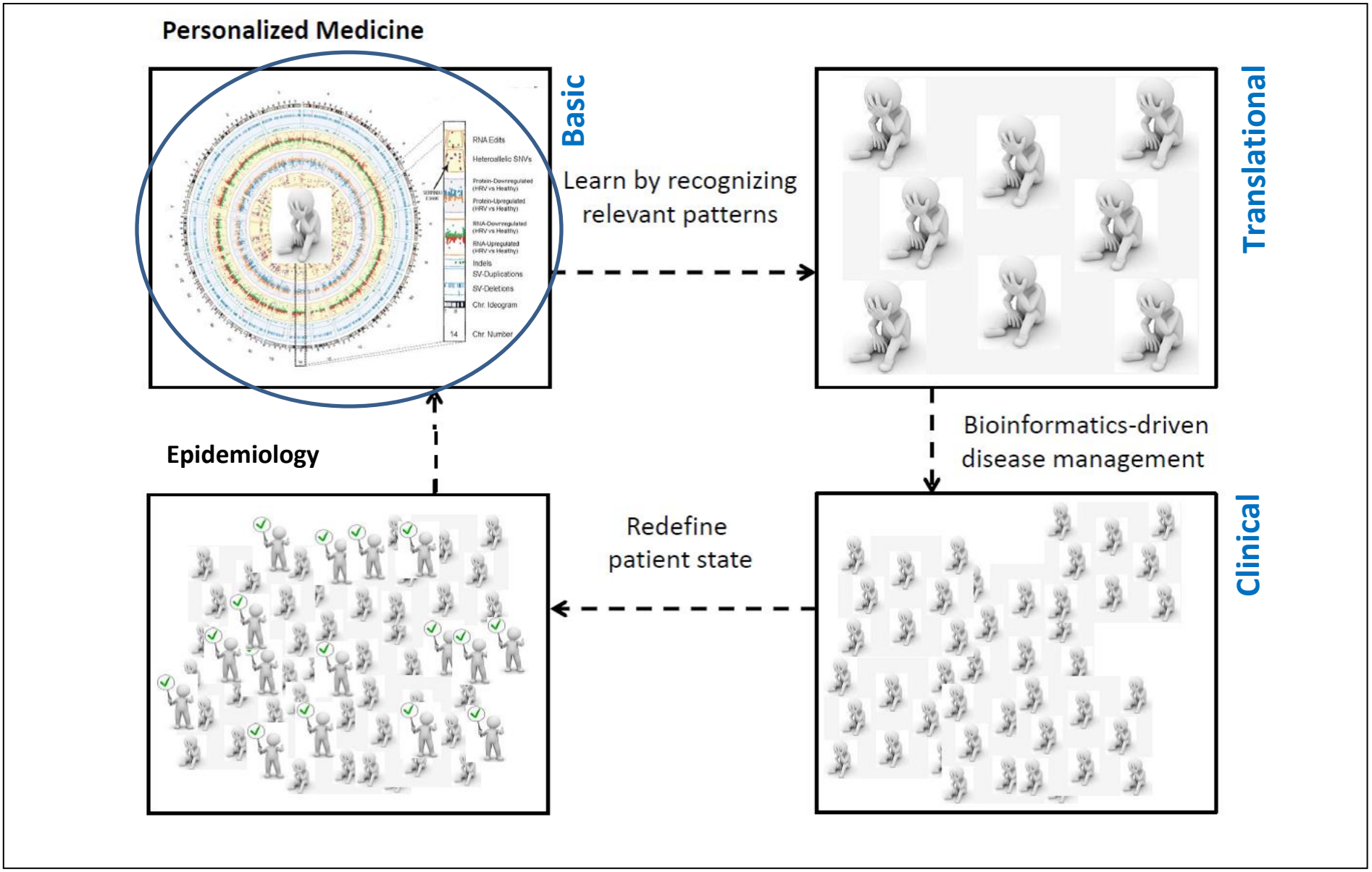
"how things work"

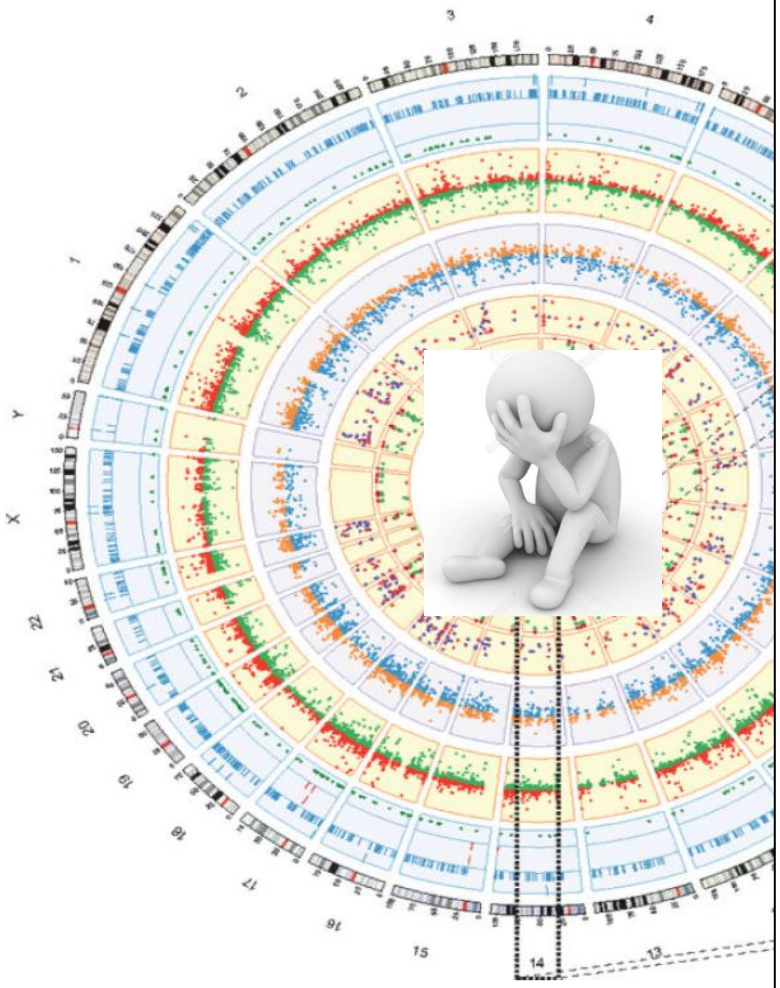
- *Understanding:*
  - *Comparisons*
  - Profiling / *Subtyping*
- Prediction:
  - Future educated guesses

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?


Personalized Medicine
Basic
Learn by recognizing relevant patterns
Translational
Bioinformatics-driven disease management
Clinical
Redefine patient state
Epidemiology
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
Chr. Number



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
Missingness

Biomarker-related "Sciences"

Basic Science

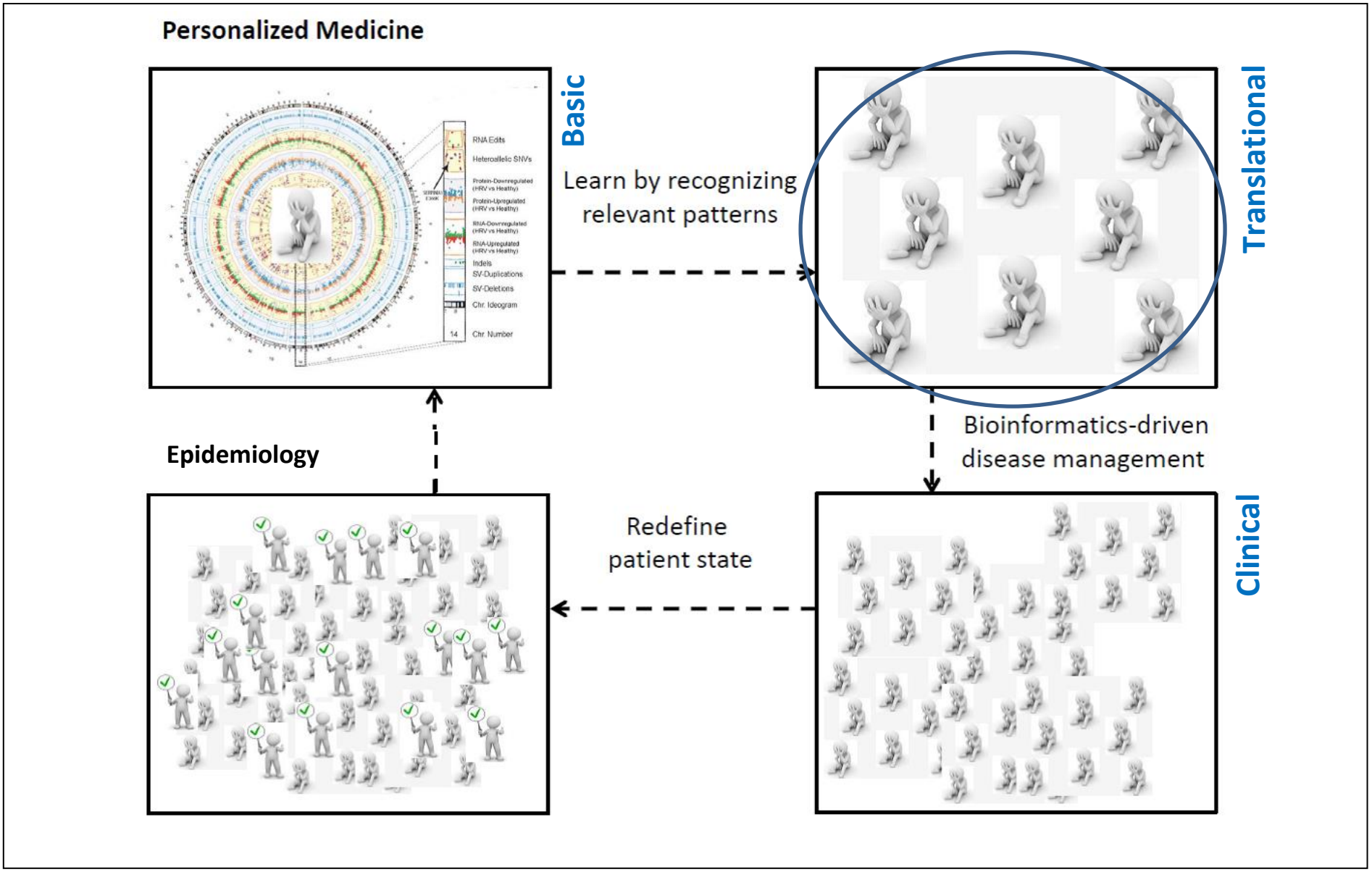
"how things work"

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

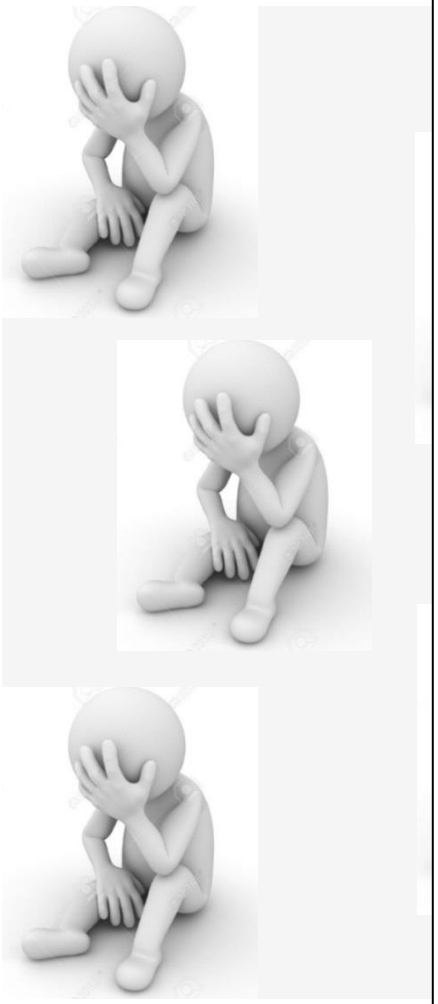
Translational Science

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

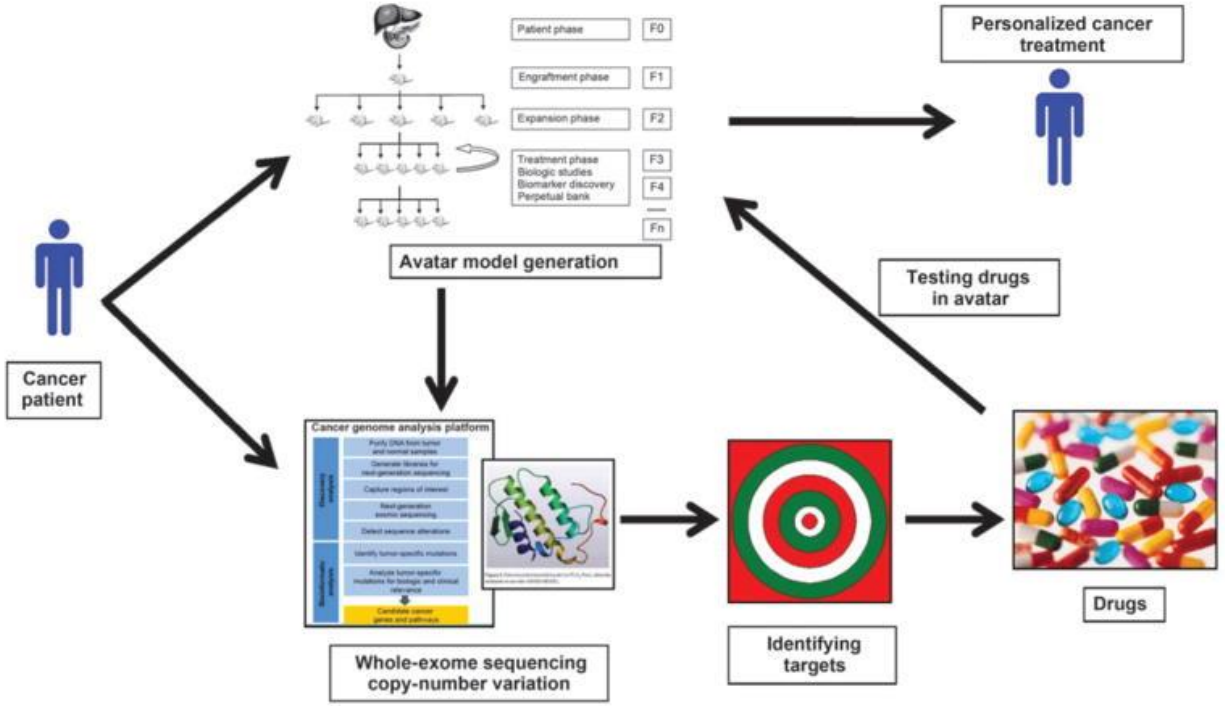
- Bioinformatics-driven pipeline based on molecular biomarkers to drive treatment management
- Stratified medicine: how much heterogeneity is allowed in strata to target?


Personalized Medicine
Basic
Learn by recognizing relevant patterns
Translational
Bioinformatics-driven disease management
Clinical
Redefine patient state
Epidemiology
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
Chr. Number

Bionformatics-driven treatment assignment

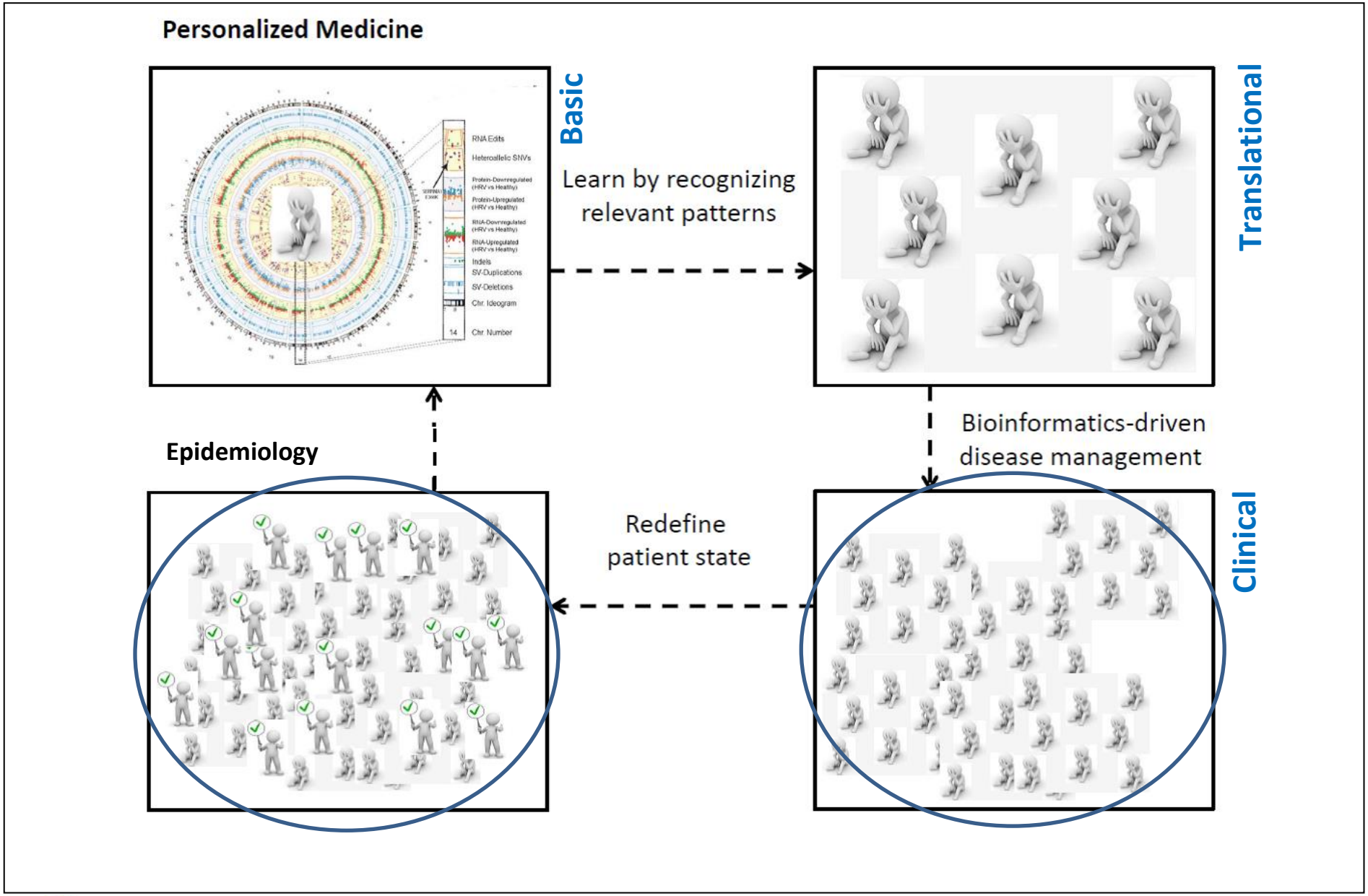


Integrating sequencing and avatar mouse models



Practical problems

(Garralda et al. 2014)


Personalized Medicine
Basic
Learn by recognizing relevant patterns
Translational
Bioinformatics-driven disease management
Clinical
Redefine patient state
Epidemiology
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
Chr. Number

Homogeneity vs heterogeneity

Molecular profiling; What does it mean to be „Diseased“?

OPEN ACCESS Freely available online

PLoS one

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

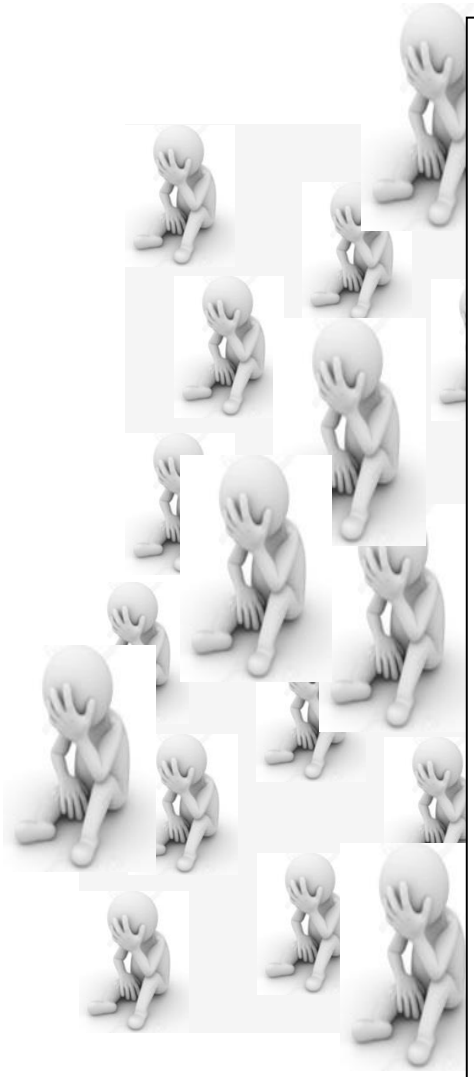
Isabelle Cleynen[1]*, Jestinah M. Mahachie John[2,3], Liesbet Henckaerts[4], Wouter Van Moerkercke[1], Paul Rutgeerts[1], Kristel Van Steen[2,3], Severine Vermeire[1]

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

Homogeneity vs heterogeneity



Molecular profiling; What does it mean to be „Diseased“?

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]

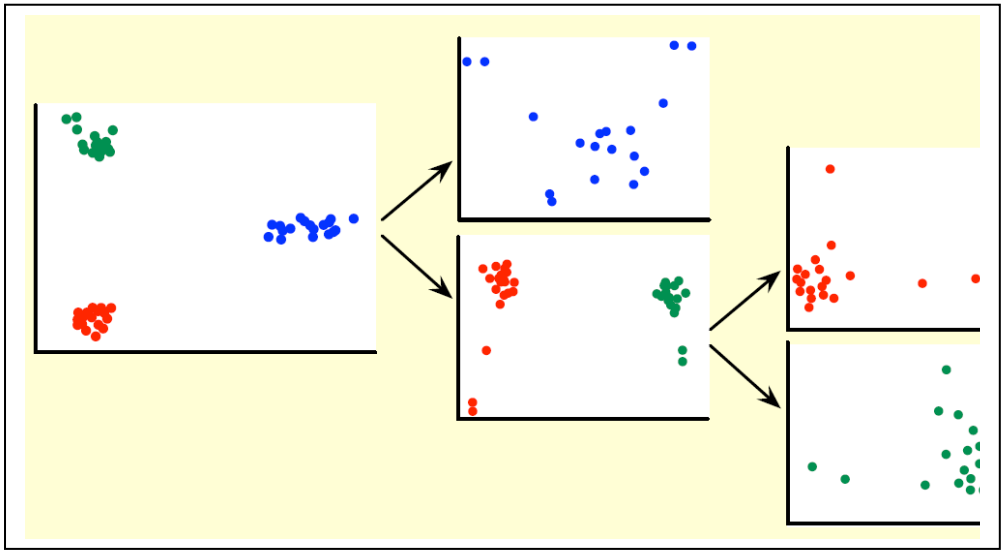
1 UMR843, INSERM, Paris, France, **2** Bioinformatics and Modeling, GIGA-R, University of Liège, Liège, Belgium, **3** UMR843, Institut National de la Sante et de la recherche Medicale, Paris, France, **4** Service de Gastroentérologie Pédiatrique, Hôpital Robert Debré, APHP, Paris, France, **5** CRC-CRB, 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>

BIO3's approach: create a fine-scale clustering tool

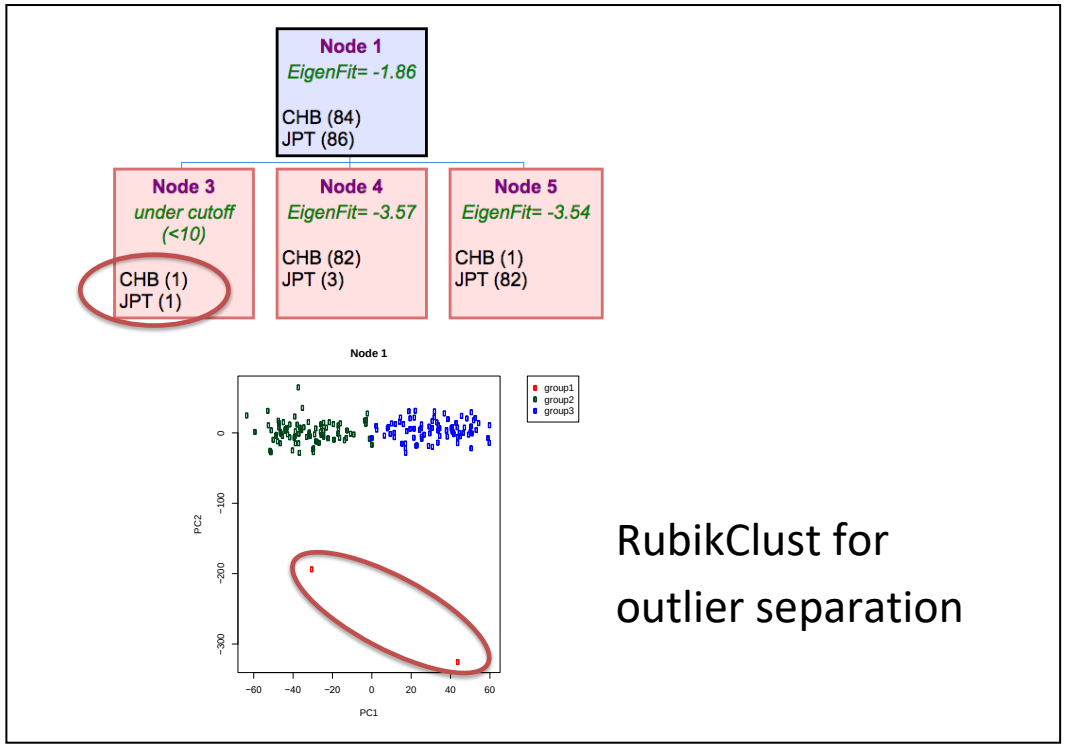
- Template: ipPCA (Intarapanich et al. 2009)
  - Performs PCA with genotype data (similar to EIGENSTRAT)
  - If substructure exists in PC space individuals are assigned to one of two clusters (2-means algorithm / fuzzy c-means)
  - Iteratively performs test for substructure and clustering on nested datasets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



BIO3's approach: create a fine-scale clustering tool

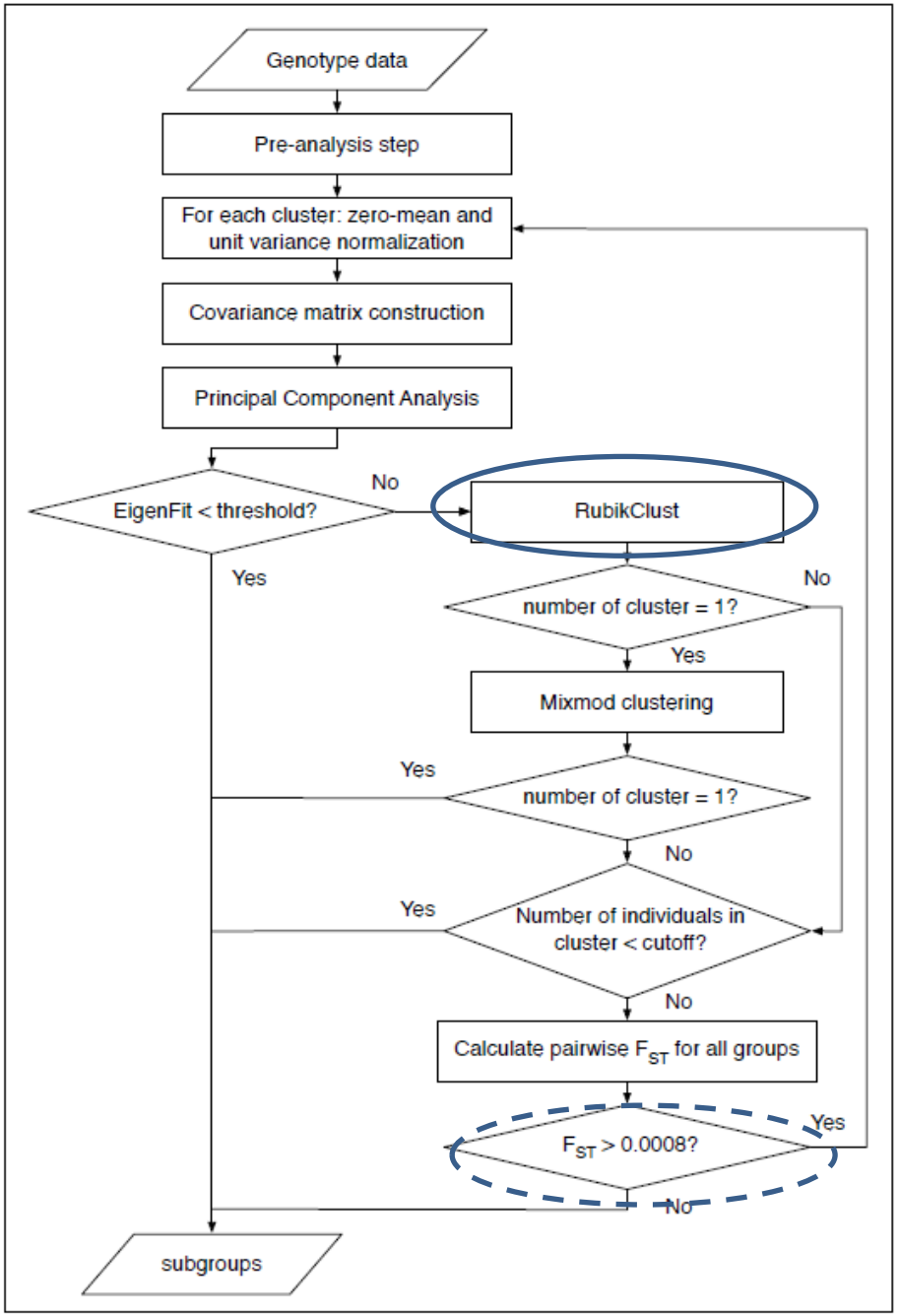
- ipPCA
  - Pros: outperformed others (STRUCTURE – 2000) in achieving higher accuracy for highly structure populations
  - Cons: binary splitting; outlier sensitive; difficult to integrate mixed data types
- Competitors:
  - SHIPS (2012) – divisive fine-scale structure detection; computational efficiency; together with STRUCTURE best accuracy (individual assignment and number of clusters)
  - iNJClust (2014) – allele sharing distance; graph based partitioning on neighbor joining tree; fixation index F_{ST}

BIO3's approach: IPCAPS

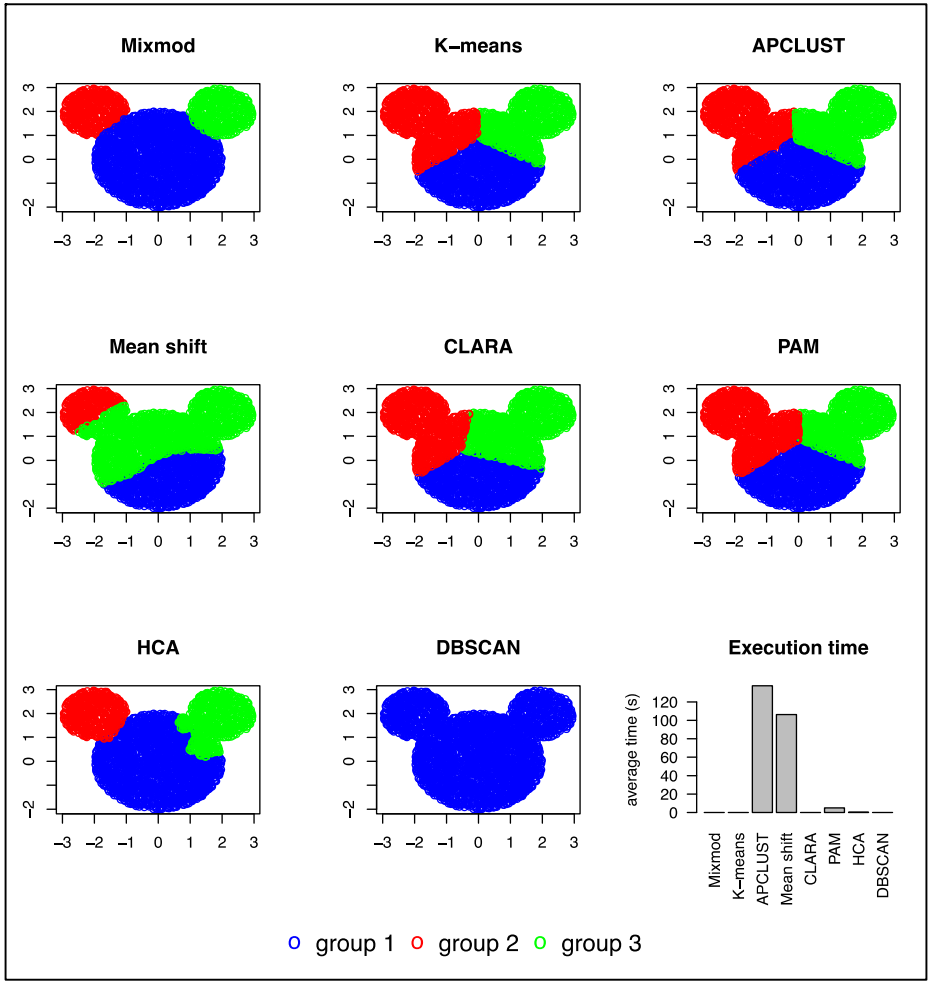


RubikClust for outlier separation

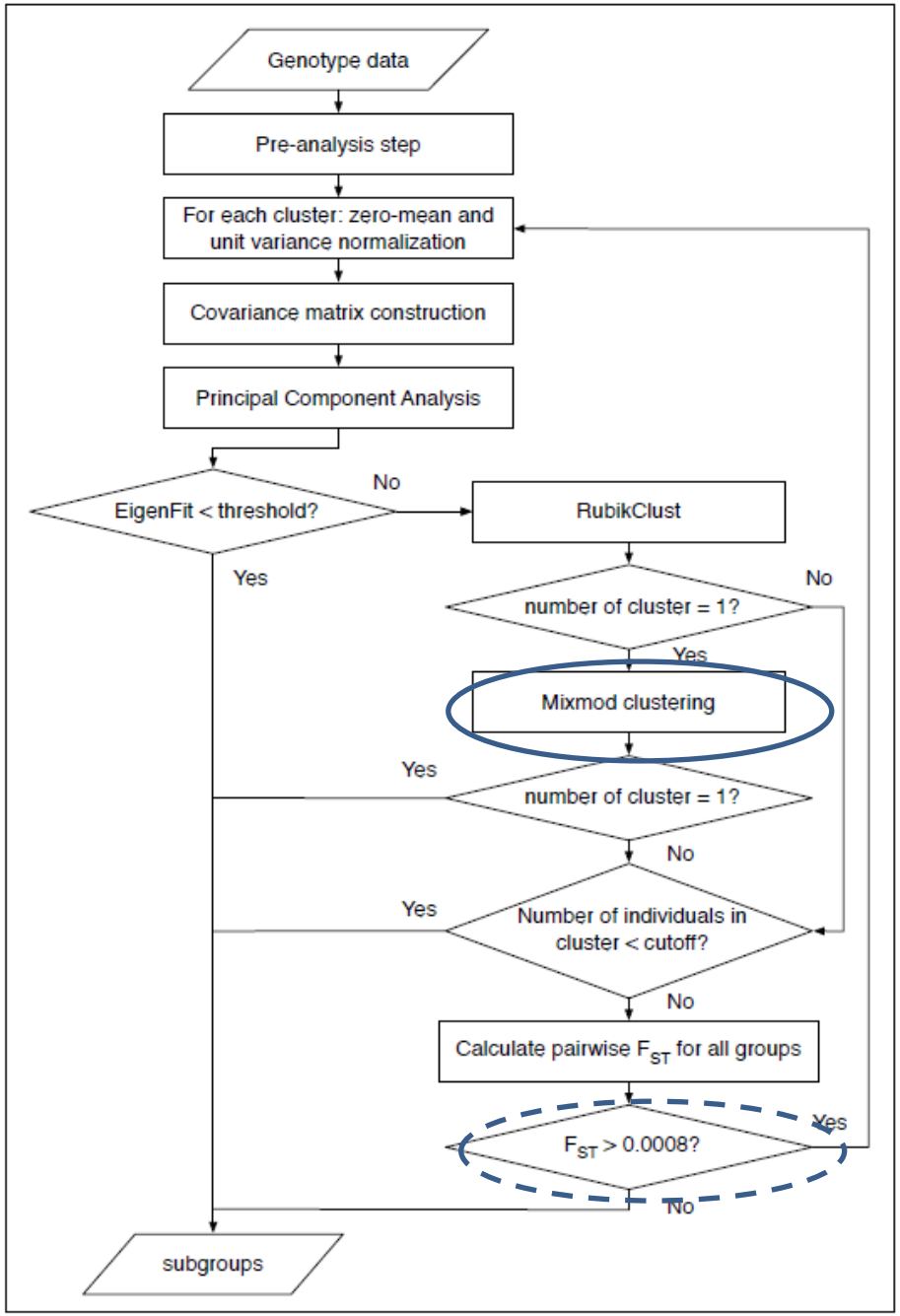
(Chaichoompu – thesis defense Oct 2017)



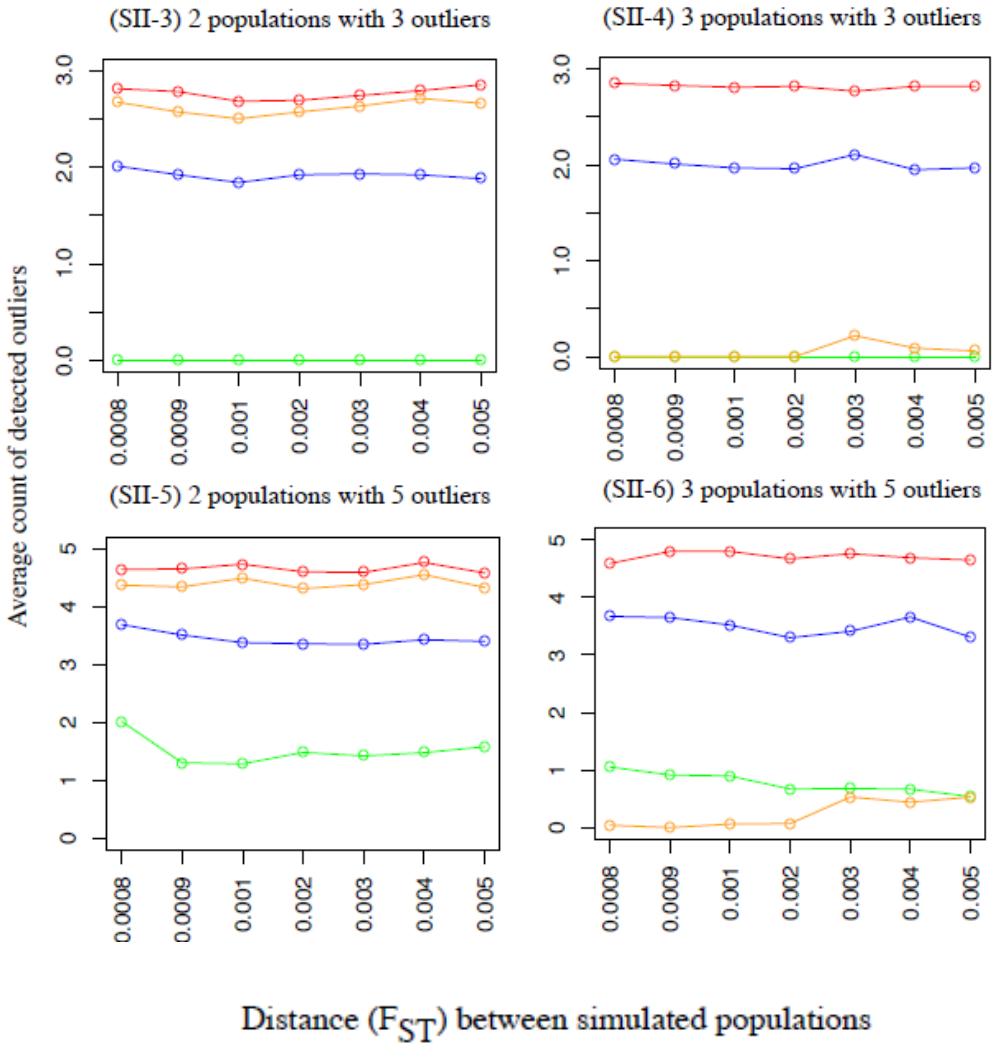
BIO3's approach: IPCAPS



(Chaichoompu – thesis defense Oct 2017)



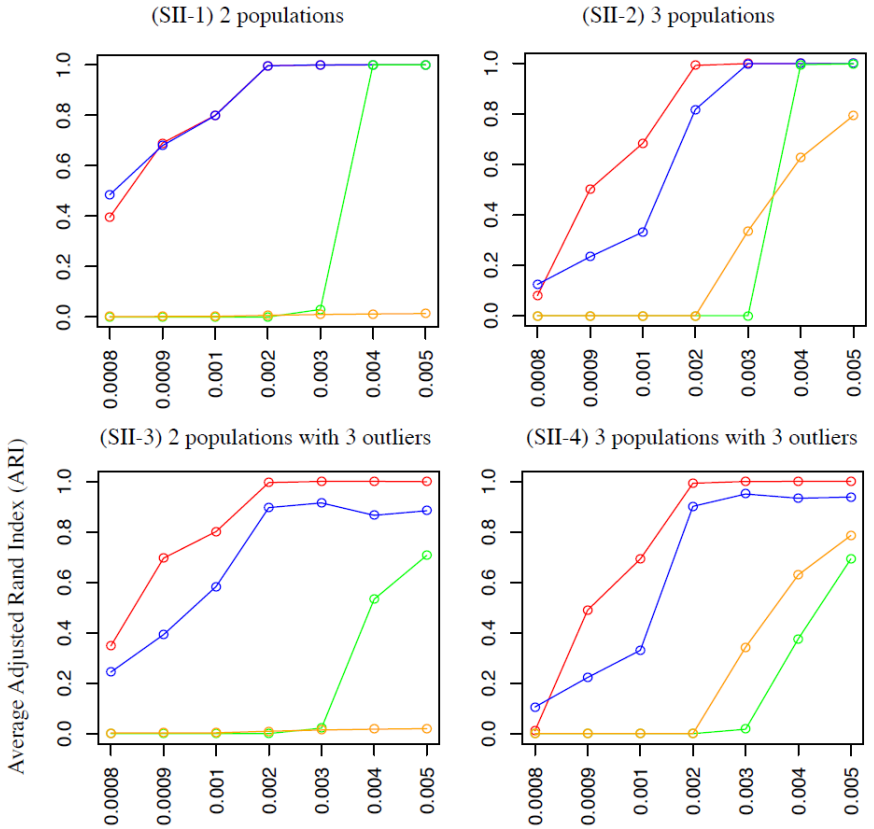
Performance of IPCAPS as outlier detection tool

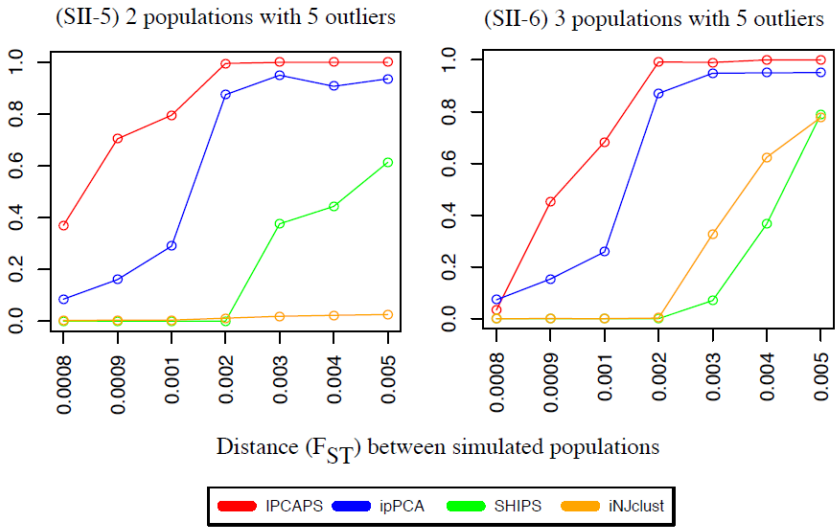


Parameters	Settings					
	SII-1	SII-2	SII-3	SII-4	SII-5	SII-6
Number of populations	2	3	2	3	2	3
Distance (F_{ST}) between populations	0.0008, 0.0009, 0.001, 0.002, 0.003, 0.004, 0.005					
Number of individuals per population	500					
Number of SNPs	10,000					
Number of outliers	0	0	3	3	5	5
Number of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

IPCAPS ipPCA SHIPS iNJclust

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					

(Rand Index = prob that two clusterings agree on a randomly chosen pair)

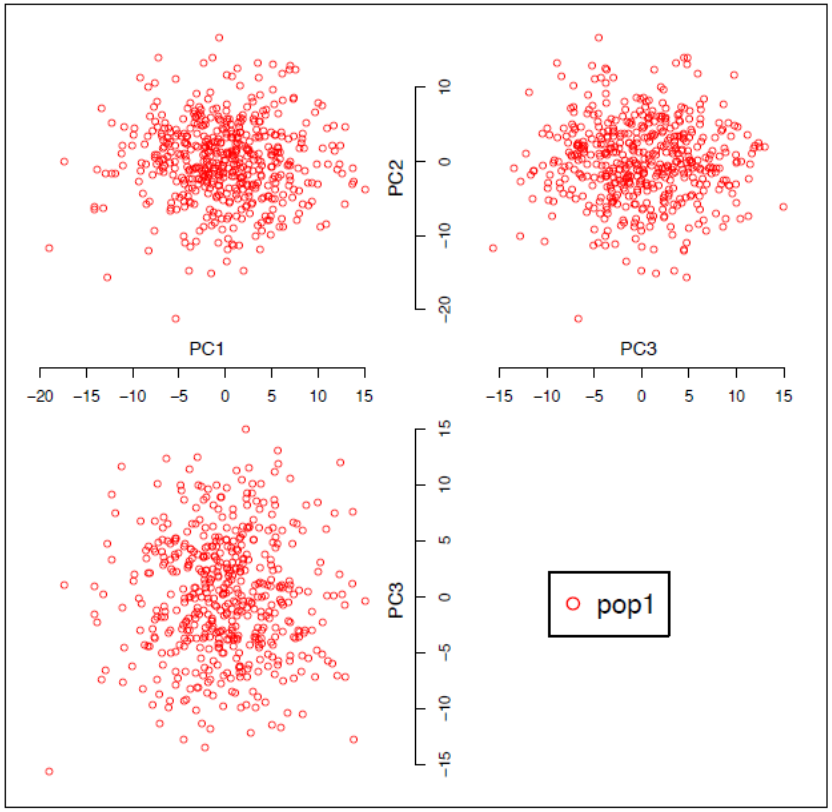
(Chaichoompu – thesis defense Oct 2017)

F_{ST} among populations – examples

	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

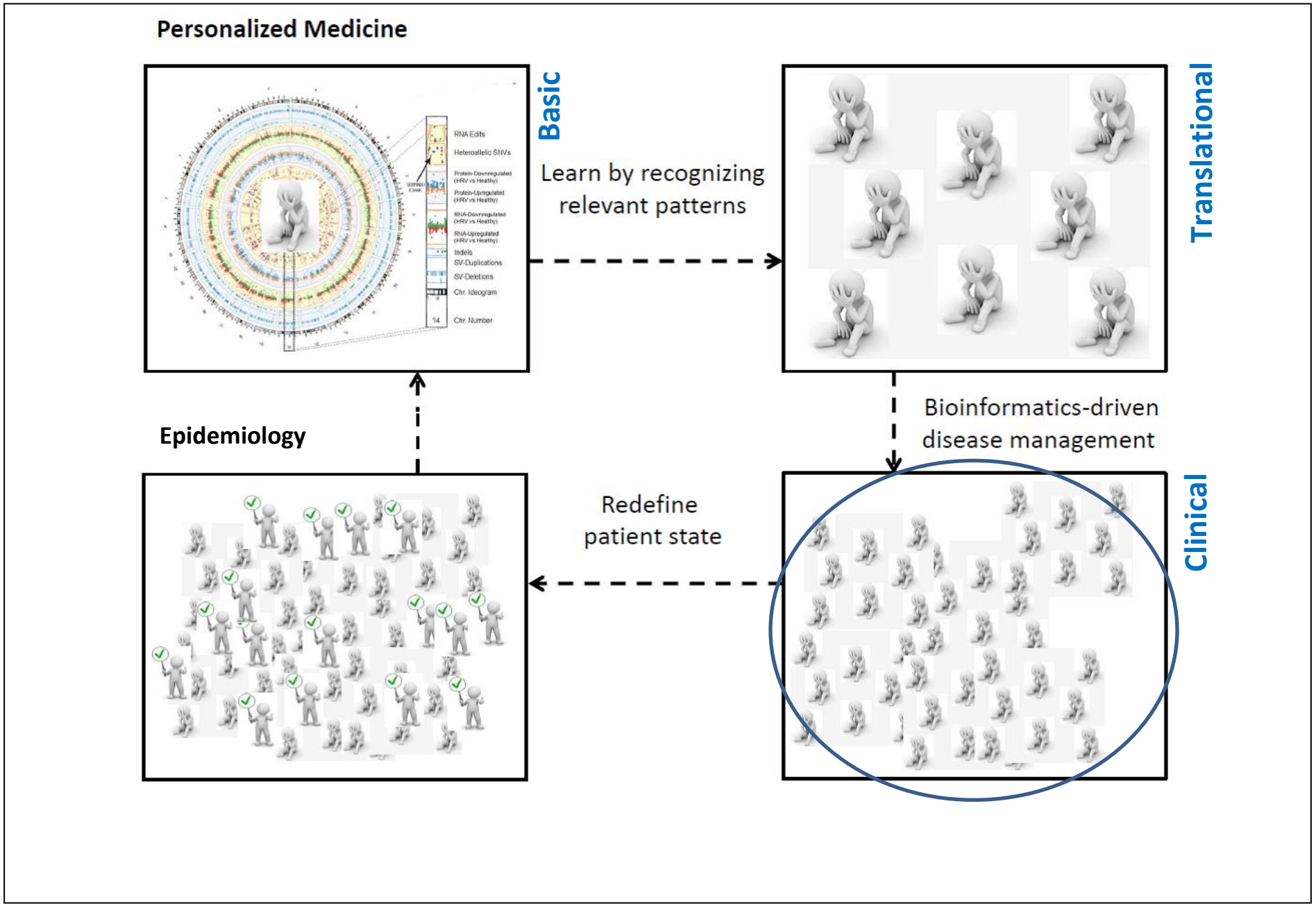
(Heath et al. 2008)

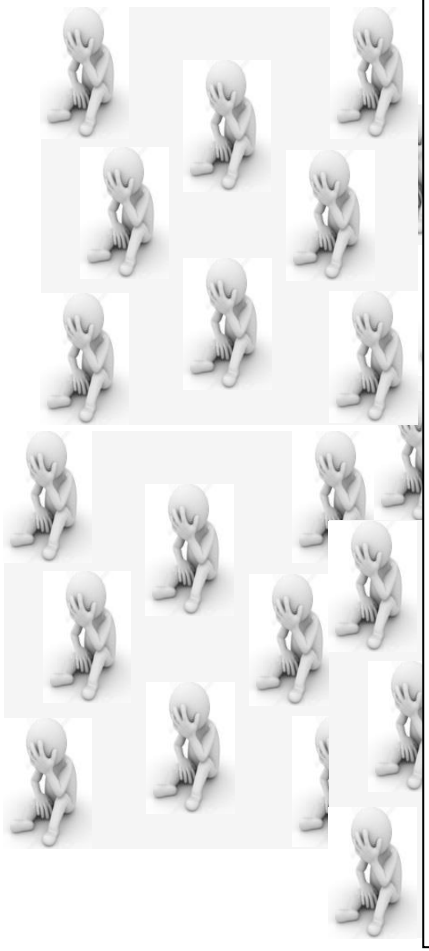
Type I error of IPCAPS



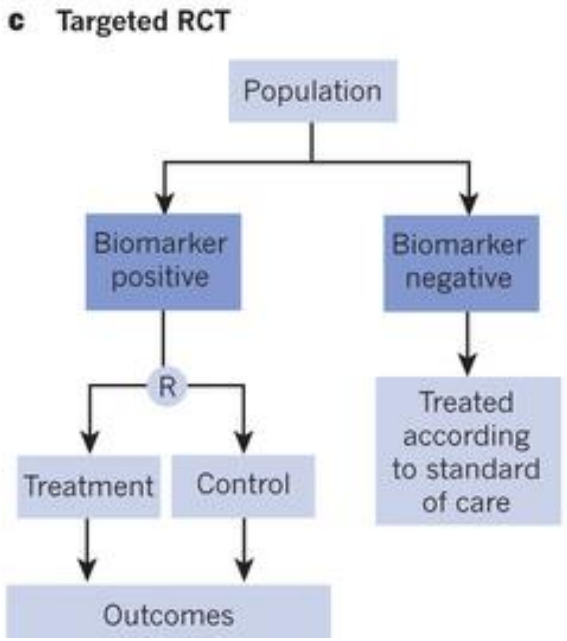
Method	Av. # clusters
IPCAPS	1
ipPCA	2
SHIPS	1
iNJclust	>150

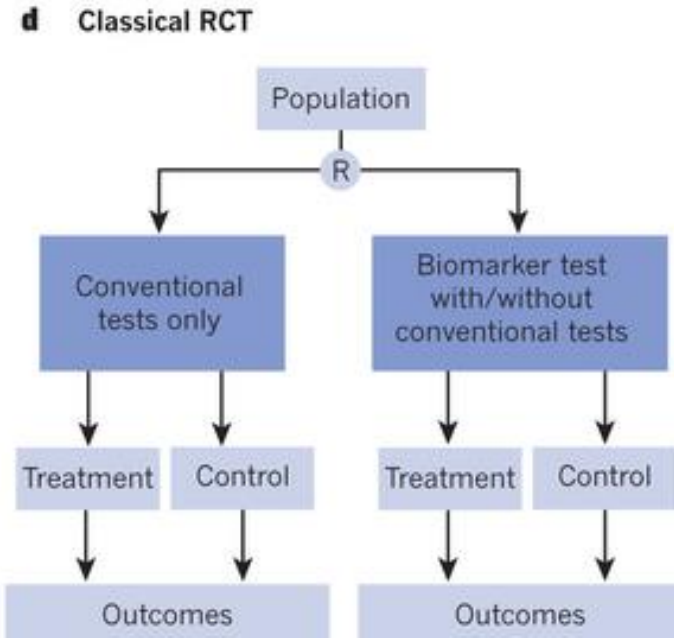
(Kridsadakorn Chaichoompu 2017,
PhD thesis – Chapter 2)


Personalized Medicine
Basic
Learn by recognizing relevant patterns
Translational
Bioinformatics-driven disease management
Clinical
Redefine patient state
Epidemiology
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
Chr. Number



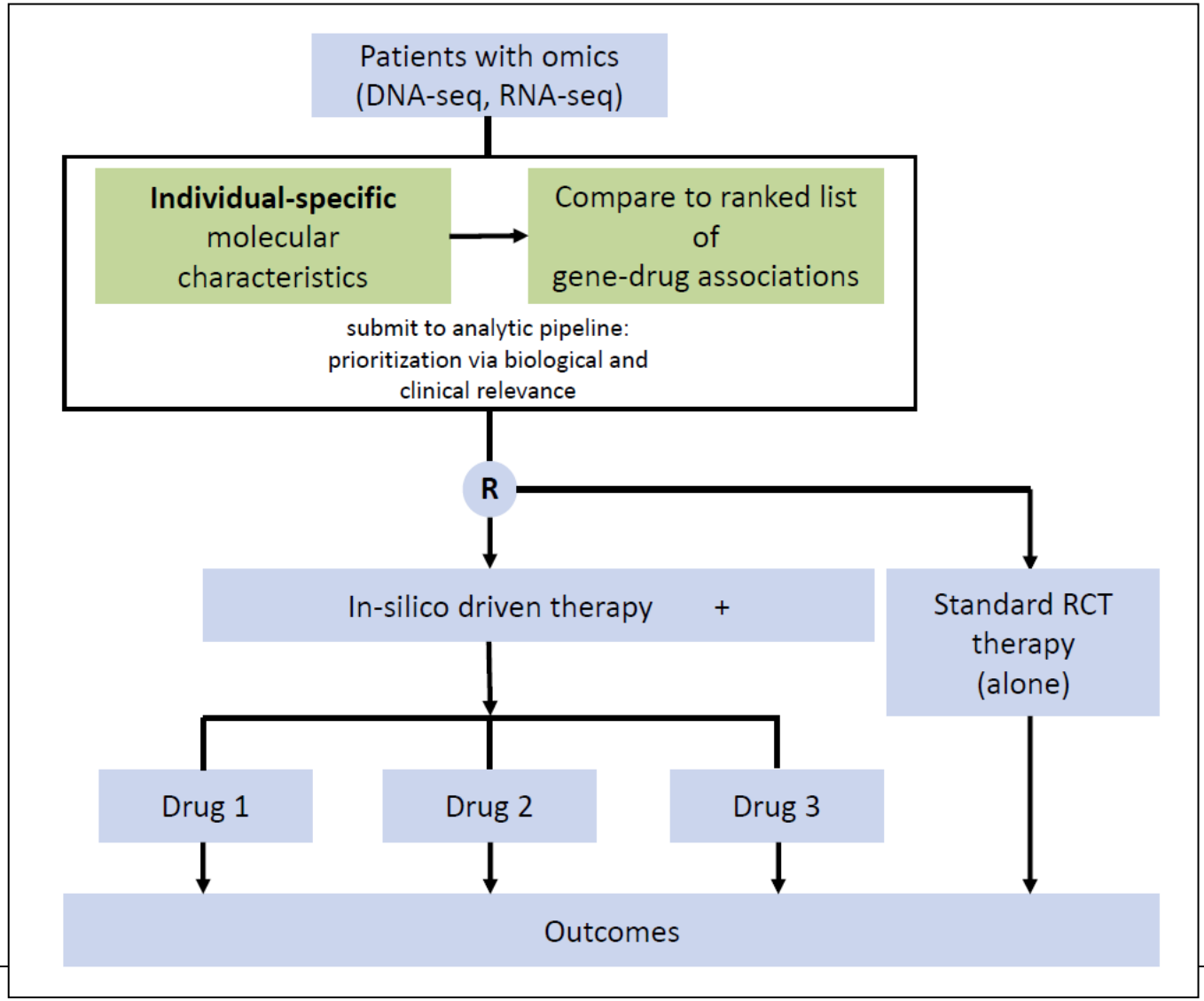
Testing precision-medicine strategies





(Biankin et al. 2015)

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



CTs in view of personalized medicine – where are we?

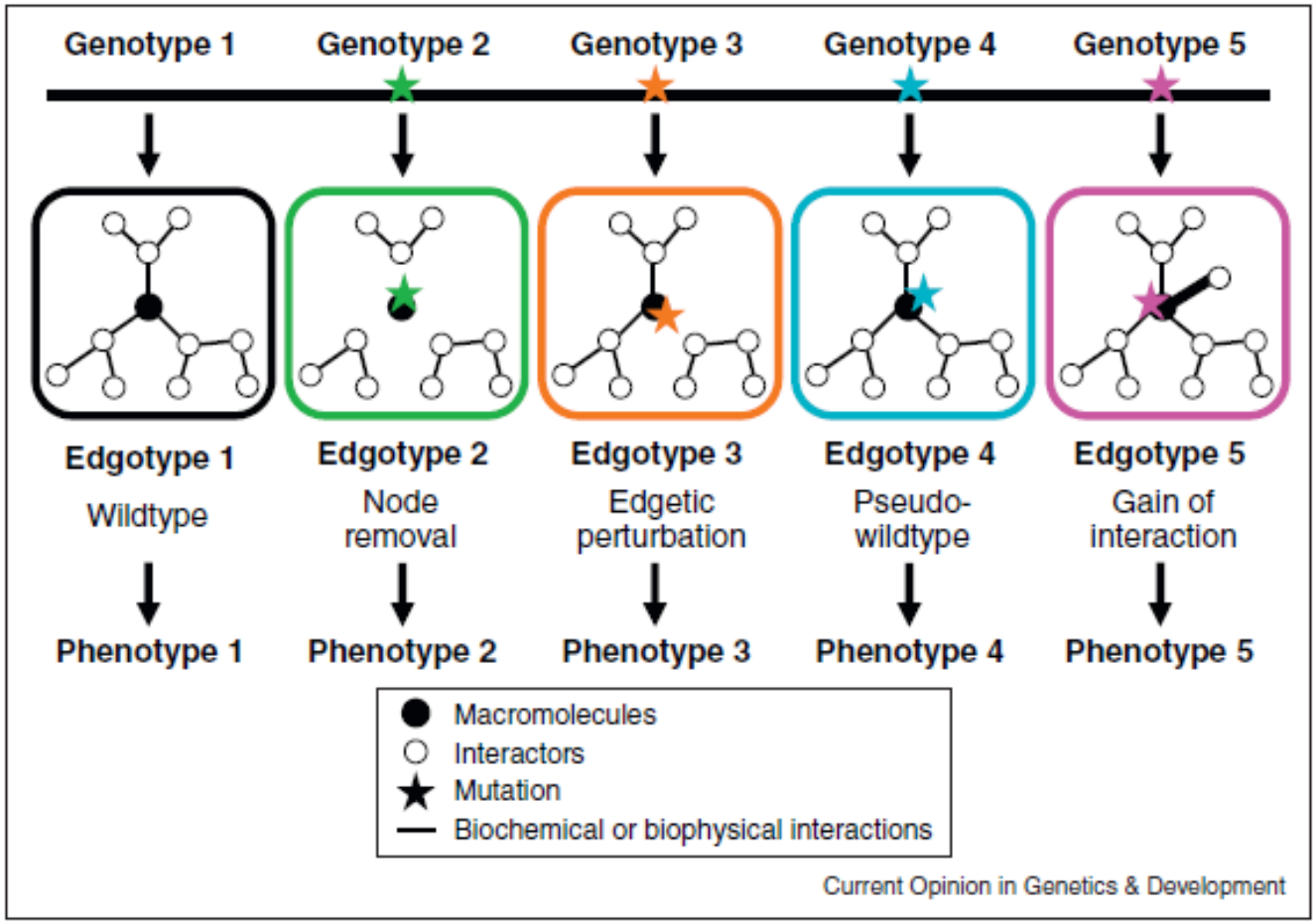
- **Basket CTs**: multiple diseases with the same genetic mutation, randomized treatment allocation
- **Umbrella CTs**: 1 "disease", different genetic mutations which define sub-cohorts, each receiving randomized treatment regimen
- Added complexities:
  - highly multi-dimensional profiles are expected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

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

Take-home messages

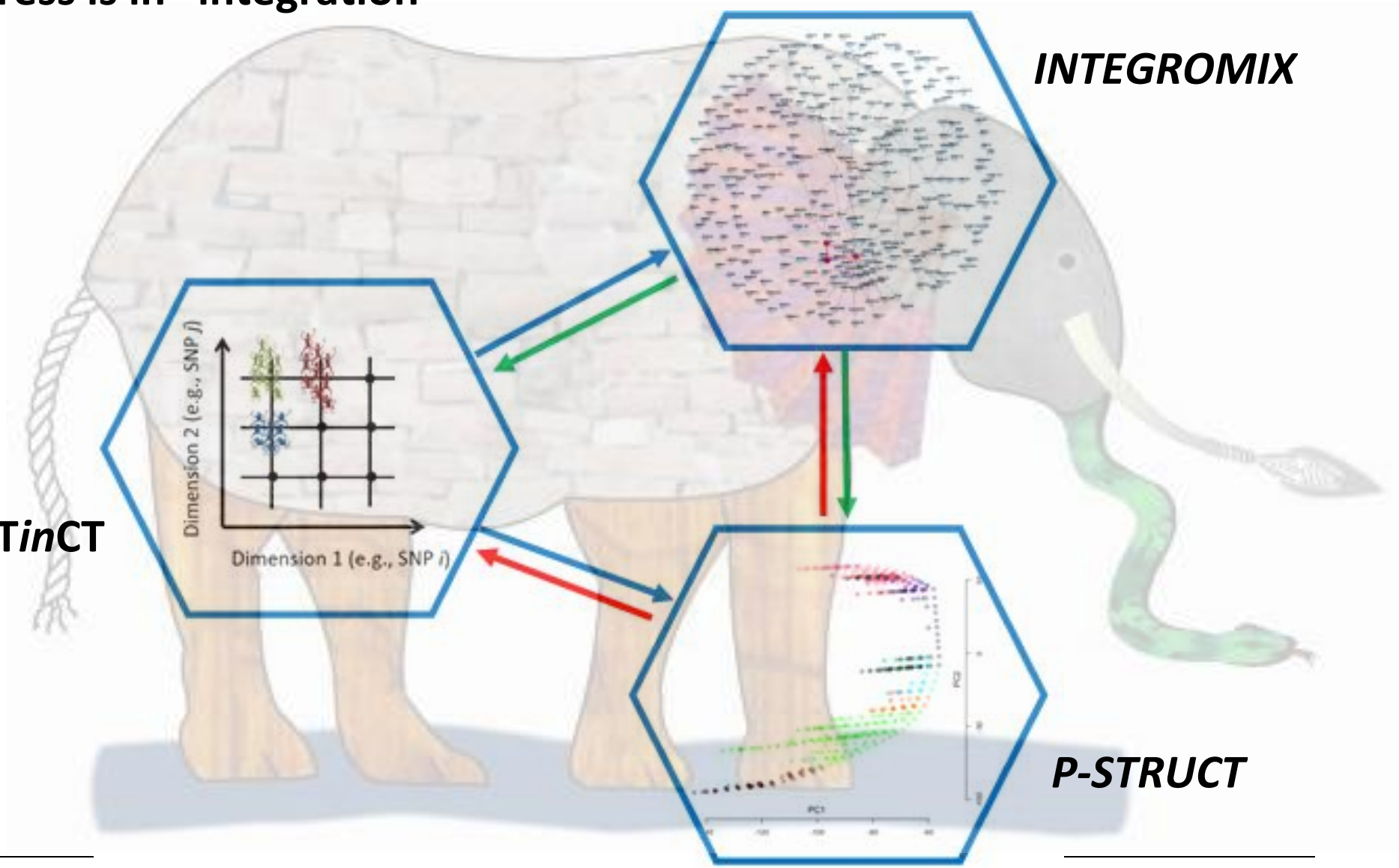
Towards CTs in PM

Patient entry criteria - Information is in “the edges”



(moving towards individual networks)

Progress is in “integration”



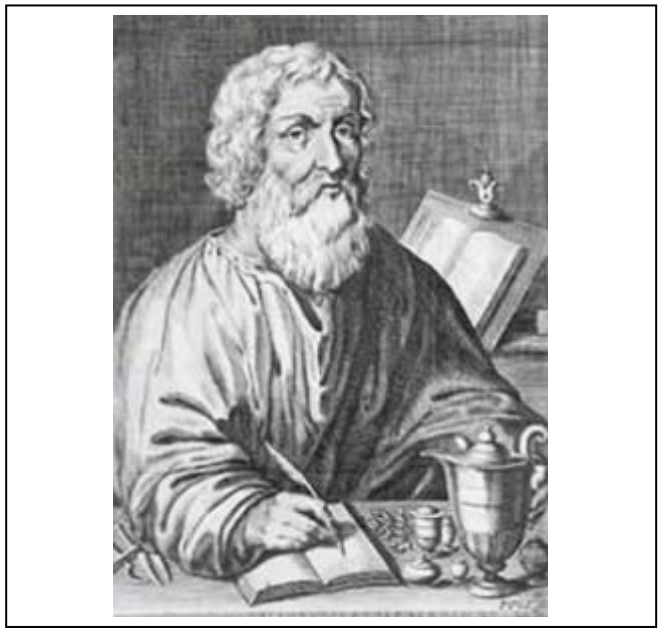
Don’t forget about presumably healthy populations

- To benchmark
- To target for interventions: risk prediction

- Again lessons can be learned from work on “interactions”
  - Collaborators extended **MB-MDR** to generate **prediction rules**
  - The new algorithm (available in R) can use information hidden in interactions more efficiently than two other state-of-the-art algorithms; it clearly **outperforms Random Forest and Elastic Net** if interactions are present.
  - The performance of these algorithms is comparable if no interactions are present

(Gola et al. 2019)

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



Hippocrates (460-370 BC)

Acknowledgements

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

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