Through the looking-glass

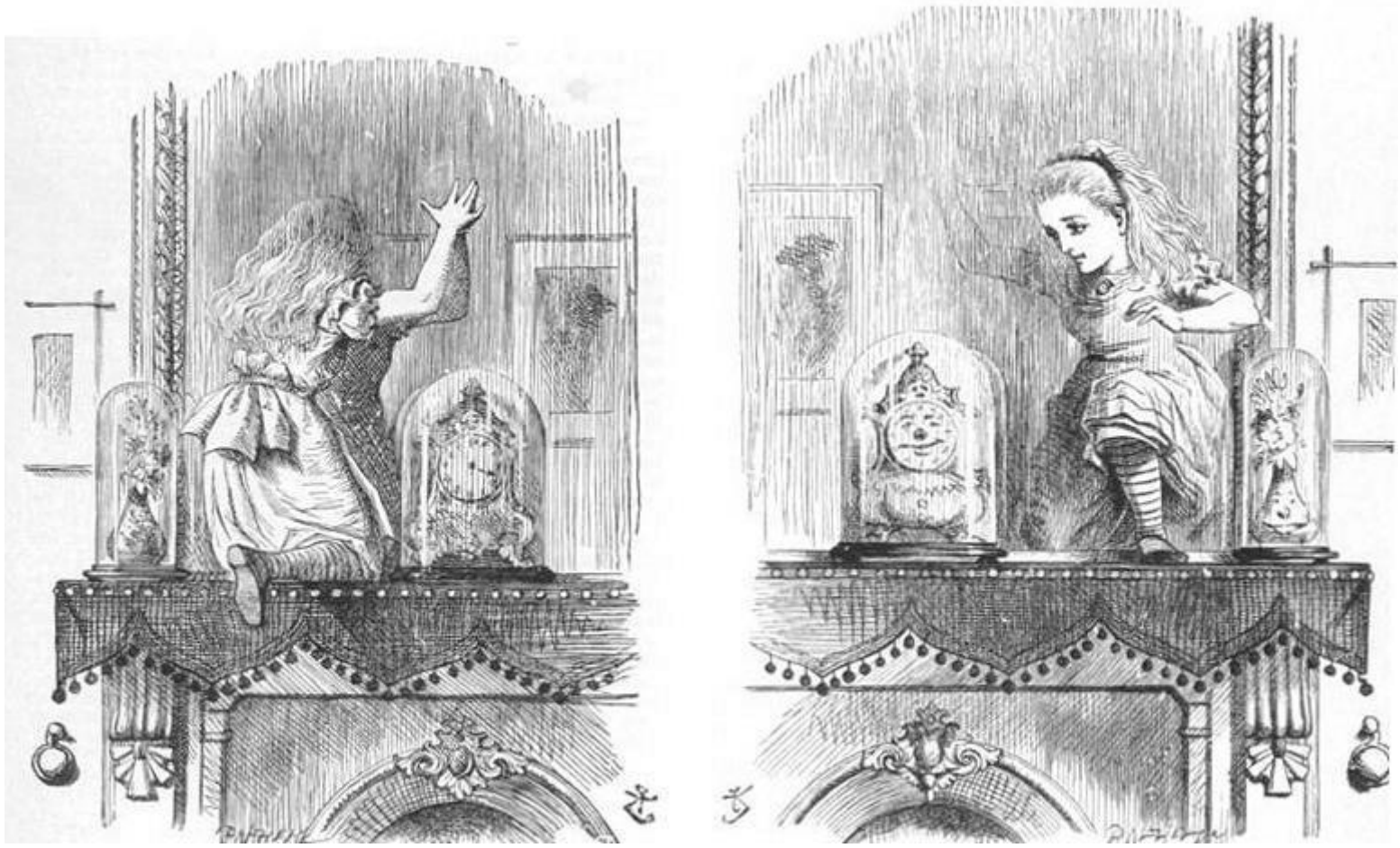
Interactions revealed!

Kristel Van Steen, PhD[2] (*)

(*) WELBIO, GIGA-R Medical Genomics (BIO3), University of Liège, Belgium

Department of Human Genetics, KULeuven, Leuven, Belgium

Through the looking-glass



(Lewis Carroll, 1871 – illustration by Sir John Tenniel)

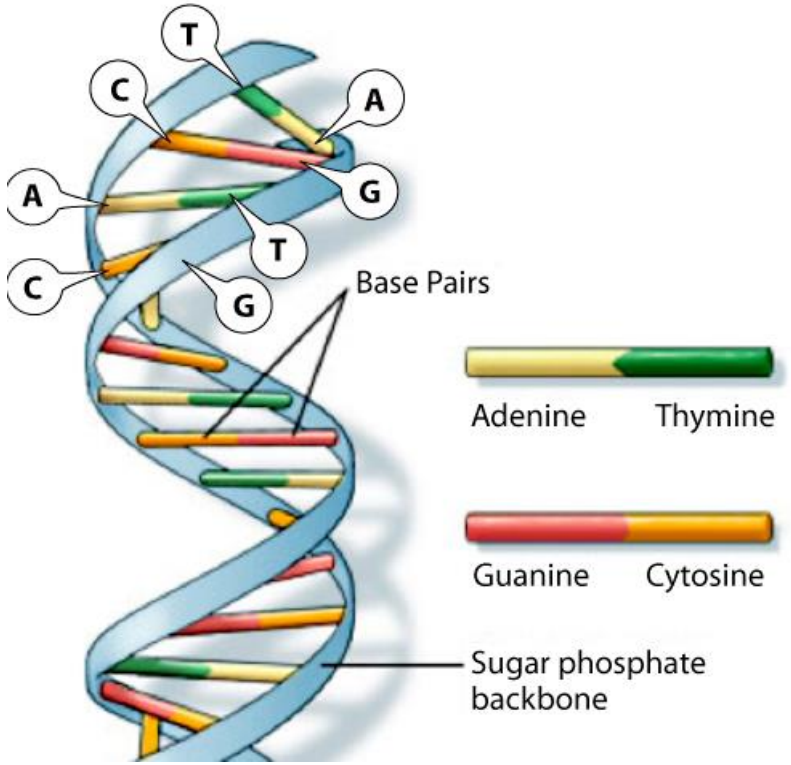
Outline

- Why unveiling interactions?
- How to identify interactions?
- What is the context?
- Are we ready for translational systemics?
- Take-home messages

Why unveiling interactions?

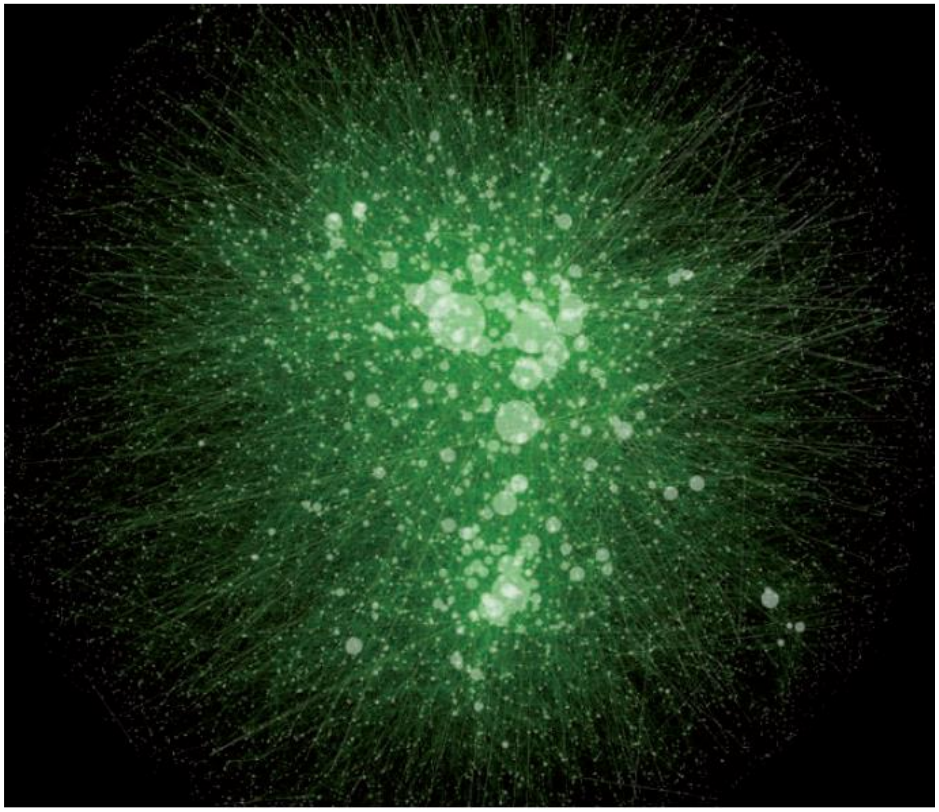
Differences between human genomes

- Any two people plucked at random off the street are on average 99.9 percent the same, DNA-wise (> 3 million differences).
- Most genome variations are relatively small and simple, involving only a few bases—an A substituted for a T here, a G left out there, a short sequence such as CT added somewhere else (somatic versus germline).



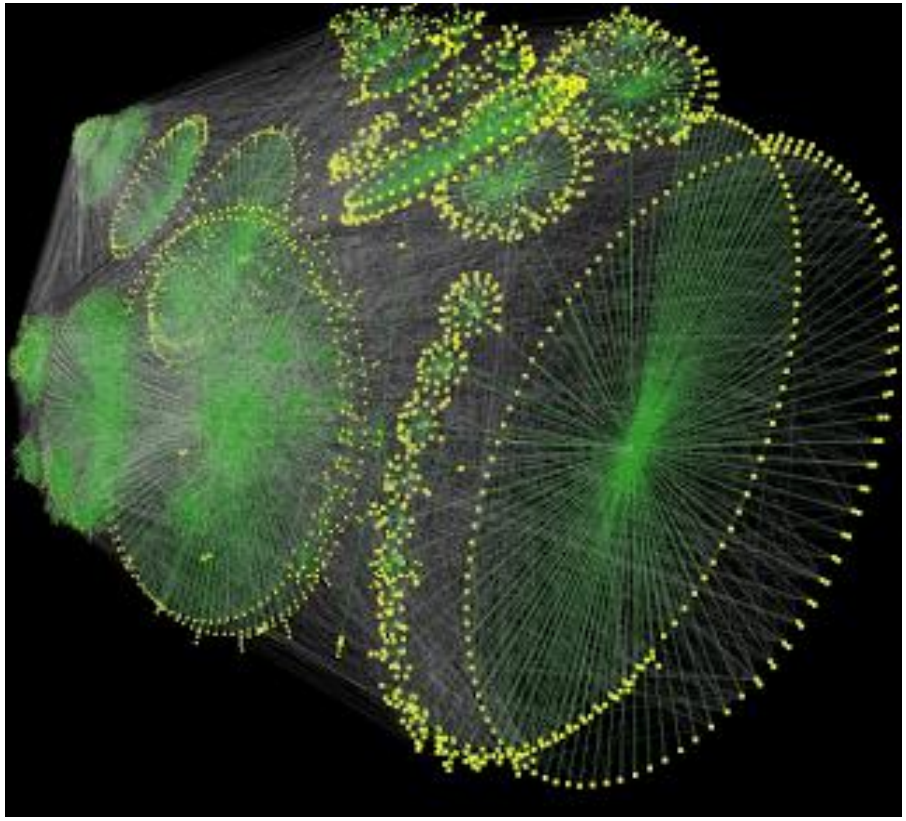
(U.S. National Library of Medicine)

Interactome differences between organisms



Human interactome

(Bonetta 2010)



Fruit fly interactome

(owww.molgen.mpg.de)

The “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. As a consequence, the interactome should be examined or considered in particular contexts.

Human interactomes

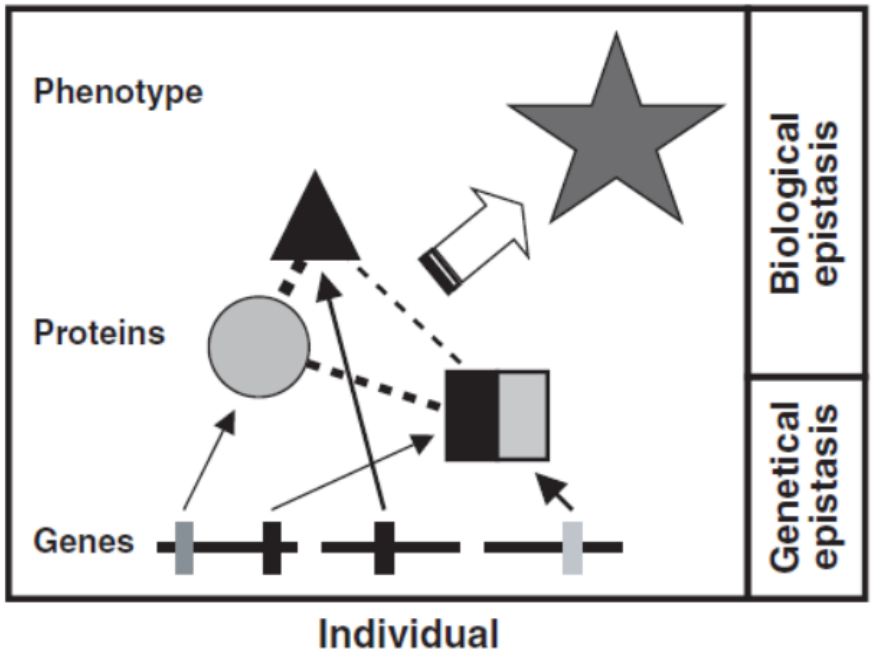
- Evolution seeks to keep our blood pressure, glucose levels and other important physiological and metabolic systems in a healthy range.
- For a phenotype to be buffered against the effects of mutations, it must have an underlying genetic architecture that is comprised of networks of genes that are redundant and robust.
- The existence of these networks creates dependencies among the genes in the network: “gene-gene interactions” or epistasis.
- This suggests that epistasis should not be a limited phenomenon and may have implications for personal genetics.

(Moore 2005, Moore and Williams 2009,
Mackay and Moore 2014)

How to detect interactions?

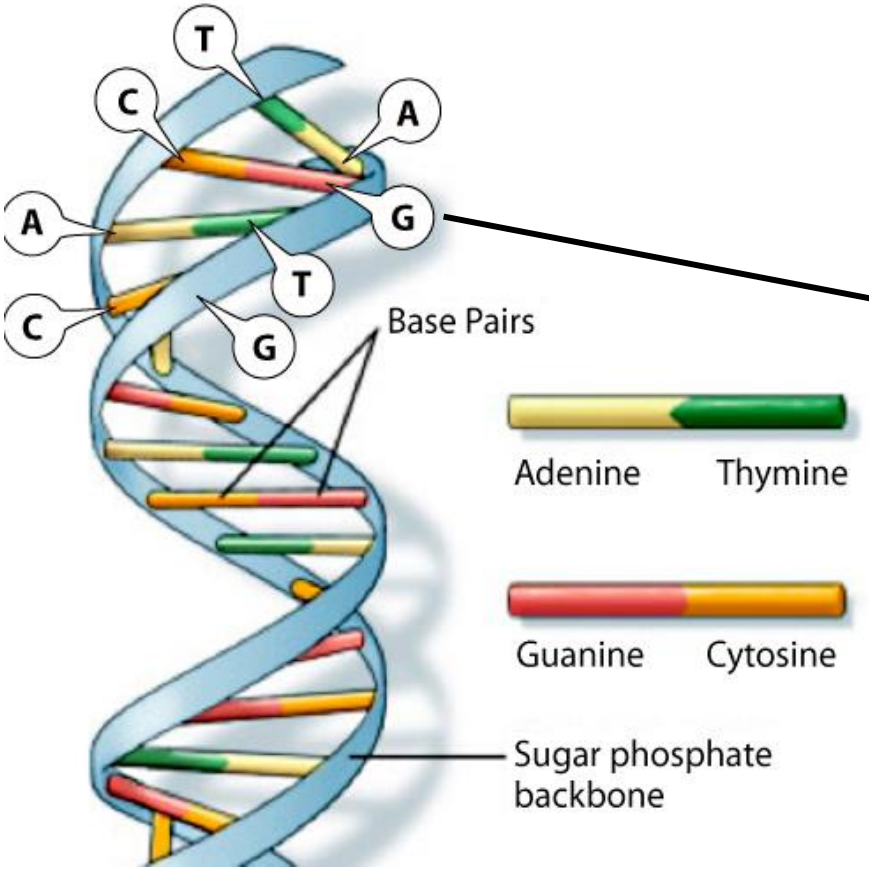
DNA-DNA interactions: biological viewpoint

- Two or more DNA variations may interact either directly to change transcription or translation levels, or indirectly by way of their protein product (to alter disease risk separate from their independent effects)



(Moore 2005)

Common genetic variations



Single Nucleotide Polymorphisms (SNPs)	Frequency in general population
G	95%
A	5% > 1%

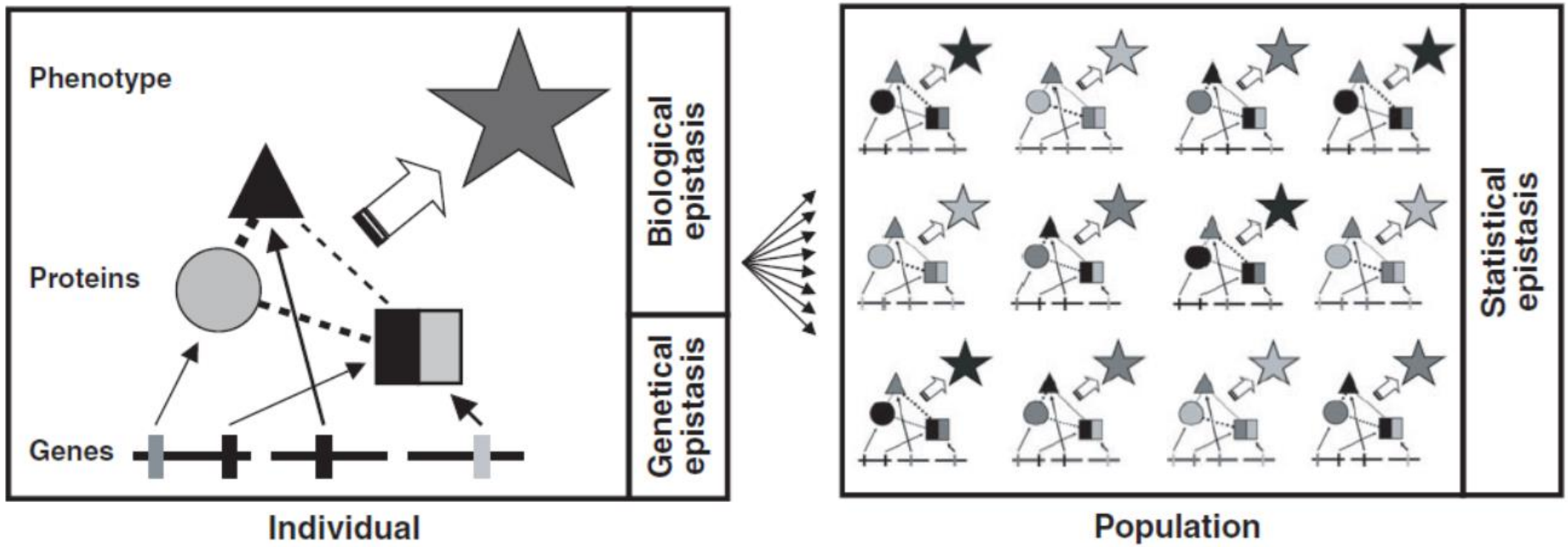
DNA-DNA interactions: multiple viewpoints

- **Biological epistasis** is the result of physical interactions among biomolecules within gene regulatory networks and biochemical pathways in an individual such that the effect of a gene on a phenotype is dependent on one or more other genes

- **Statistical epistasis** is defined as deviation from additivity in a mathematical model where the relationship between multilocus genotypes and phenotypic variation in a population is not predictable based solely on the actions of the genes considered singly.

(Moore and Williams 2005)

DNA-DNA interactions: aligning viewpoints



(Moore 2005)

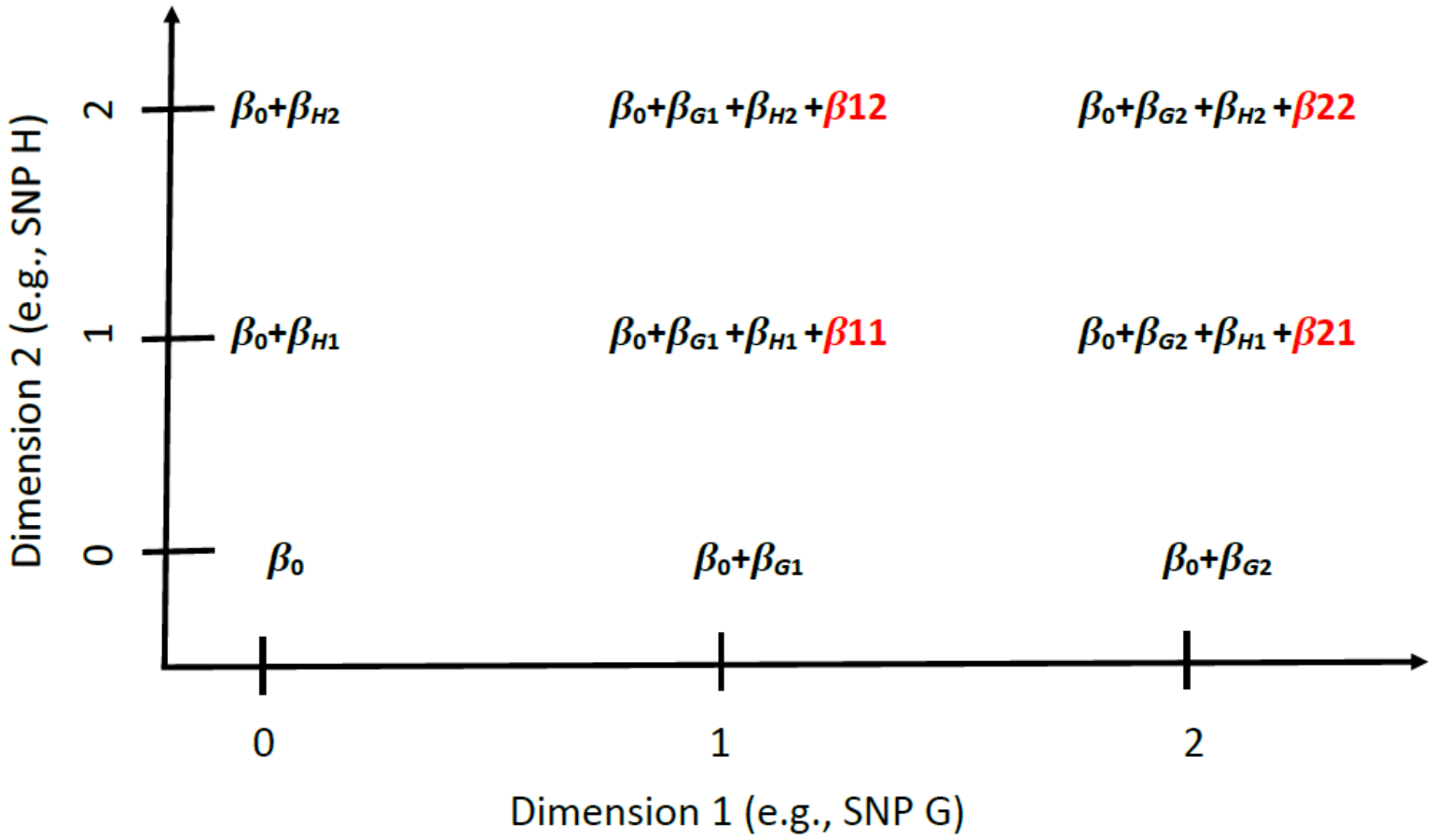
Little correspondence

- From the literature:
  - Siemiatycki and Thomas (1981) Int J Epidemiol 10:383-387
  - …
  - Moore and Williams (2005) BioEssays 27:637–646
  - Phillips (2008) Nat Rev Genet 9:855-867
  - Clayton DG (2009) PLoS Genet 5(7): e1000540
  - Wang, Elston and Zhu (2010) Hum Hered 70:269-277
  - …

  - **Van Steen et al (2012) Brief Bioinform. 13(1):1-19.**
  - **Aschard et al (2012) Hum Genet 131(10):1591-1613.**
  - **Gusareva and Van Steen (2014) Hum Genet 133(11):1343-58.**

- Statistical interactions DO imply joint involvement

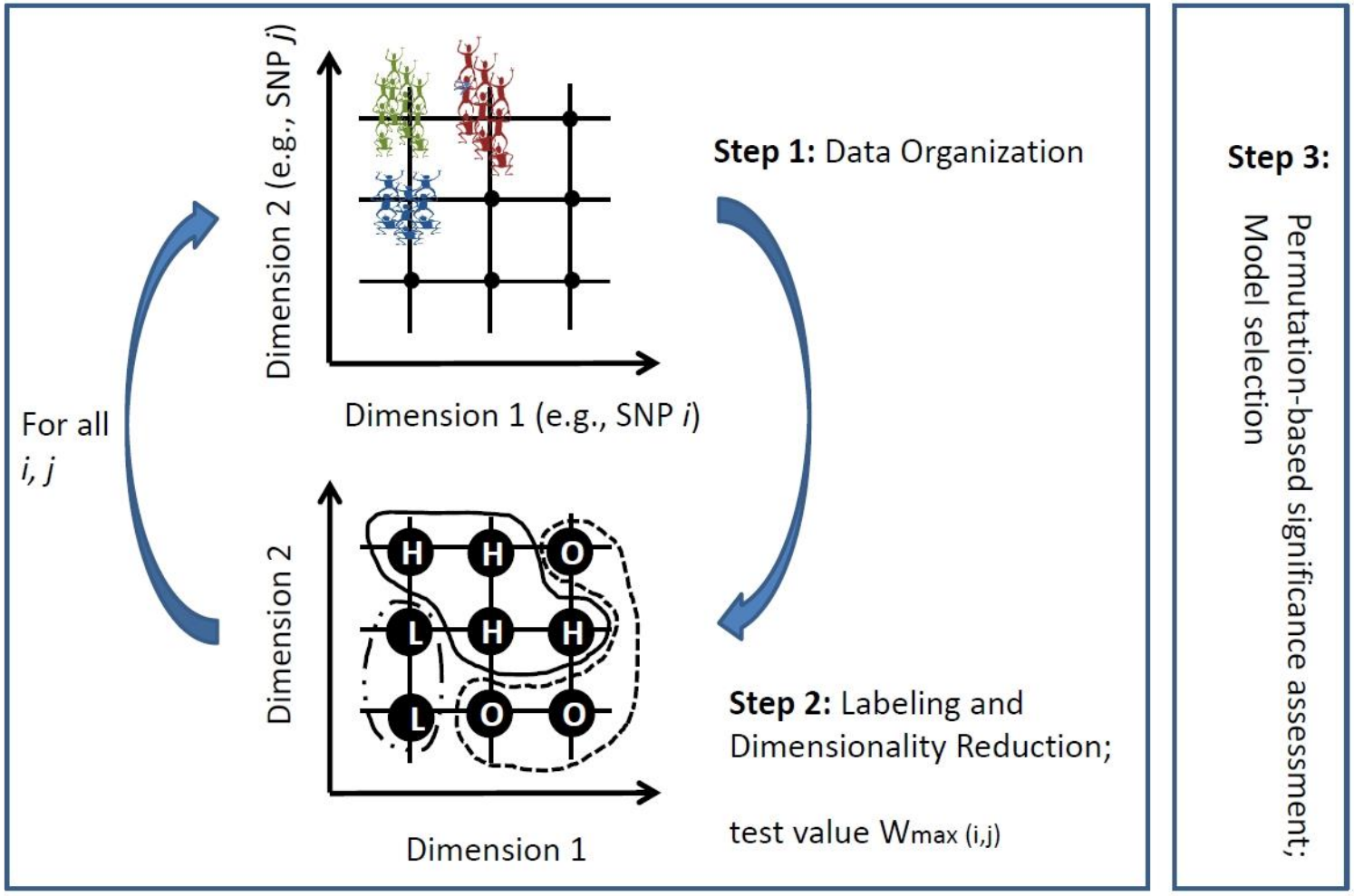
DNA-DNA interactions: statistical epistasis in regression



[classical regression framework; max 8 df]

Model-Based Multifactor Dimensionality Reduction

Model-Based Multifactor Dimensionality Reduction (MB-MDR)



Performance

Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
False Positives (%)											
MB	**MDR**	**MB**	**MDR**	**MB**	**MDR**	**MB**	**MDR**	**MB**	**MDR**	**MB**	**MDR**
6	9	4	5	6	17	5	13	5	21	5	23
Power (%)											
MB	**MDR**	**MB**	**MDR**	**MB**	**MDR**	**MB**	**MDR**	**MB**	**MDR**	**MB**	**MDR**
100	99	100	100	100	95	100	93	93	62	97	73
MB-MDR (MB): p_c = 0.1, T = *H* vs *L* test; MDR: default options, screening over 1-5 order models											

(Cattaert et al. 2011)

Model 1, $p = 0.5$

	BB	Bb	bb
AA	0	0.1	0
Aa	0.1	0	0.1
aa	0	0.1	0

Model 3, $p = 0.25$

	BB	Bb	bb
AA	0.08	0.07	0.05
Aa	0.1	0	0.1
aa	0.03	0.1	0.04

Model 5, $p = 0.1$

	BB	Bb	bb
AA	0.07	0.05	0.02
Aa	0.05	0.09	0.01
aa	0.02	0.01	0.03

Model 2, $p = 0.5$

	BB	Bb	bb
AA	0	0	0.1
Aa	0	0.05	0
aa	0.1	0	0

Model 4, $p = 0.25$

	BB	Bb	bb
AA	0	0.01	0.09
Aa	0.04	0.01	0.08
aa	0.07	0.09	0.03

Model 6, $p = 0.1$

	BB	Bb	Bb
AA	0.09	0.001	0.02
Aa	0.08	0.07	0.005
aa	0.003	0.007	0.02

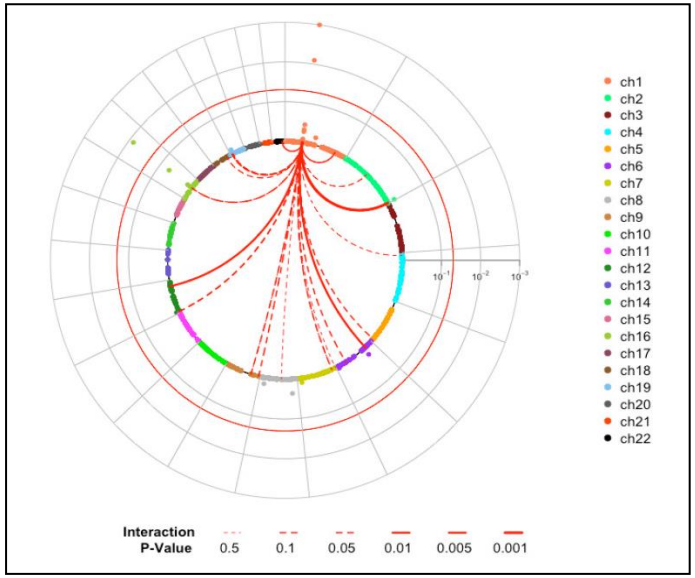
Learning from data

- **Calle**, M. L., Urrea, V., Vellalta, G., Malats, N. & Van Steen, K. (2008a) Model-Based Multifactor Dimensionality Reduction for detecting interactions in high-dimensional genomic data. Technical Report No. 24, Department of Systems Biology, Universitat de Vic, http://www.recercat.net/handle/2072/5001 **[technical report, first mentioning MB-MDR]**
- **Calle** M, Urrea V, Malats N, Van Steen K. (2008) Improving strategies for detecting genetic patterns of disease susceptibility in association studies – Statistics in Medicine 27 (30): 6532-6546 **[MB-MDR with Wald tests and MAF dependent empirical test distributions]**
- **Calle** ML, Urrea V, Van Steen K (2010) mbmdr: an R package for exploring gene-gene interactions associated with binary or quantitative traits. Bioinformatics Applications Note 26 (17): 2198-2199 **[first MB-MDR software tool, in R]**
- **Cattaert** T, Urrea V, Naj AC, De Lobel L, De Wit V, Fu M, Mahachie John JM, Shen H, Calle ML, Ritchie MD, Edwards T, Van Steen K. (2010) FAM-MDR: a flexible family-based multifactor dimensionality reduction technique to detect epistasis using related individuals, PLoS One 5 (4). **[first implementation of MB-MDR in C++, with improved features on multiple testing correction and improved association tests + recommendations on handling family-based designs]**

- **Cattaert** T, Calle ML, Dudek SM, Mahachie John JM, Van Lishout F, Urrea V, Ritchie MD, Van Steen K (2010) Model-Based Multifactor Dimensionality Reduction for detecting epistasis in case-control data in the presence of noise (*invited paper*). Ann Hum Genet. 2011 Jan;75(1):78-89 **[detailed study of C++ MB-MDR performance with binary traits]**
- **Mahachie John** JM, Cattaert T, De Lobel L, Van Lishout F, Empain A, Van Steen K (2011) Comparison of genetic association strategies in the presence of rare alleles. BMC Proceedings, 5(Suppl 9):S32 **[first explorations on C++ MB-MDR applied to rare variants]**
- **Mahachie John** JM, Cattaert T, Van Lishout F, Van Steen K (2011) Model-Based Multifactor Dimensionality Reduction to detect epistasis for quantitative traits in the presence of error-free and noisy data. European Journal of Human Genetics 19, 696-703. **[detailed study of C++ MB-MDR performance with quantitative traits]**
- **Van Steen** K (2011) Travelling the world of gene-gene interactions *(invited paper).* Brief Bioinform 2012, Jan; 13(1):1-19. **[positioning of MB-MDR in general epistasis context]**
- **Mahachie John** JM , Cattaert T , Van Lishout F , Gusareva ES , Van Steen K (2012) Lower-Order Effects Adjustment in Quantitative Traits Model-Based Multifactor Dimensionality Reduction. PLoS ONE 7(1): e29594. doi:10.1371/journal.pone.0029594 **[recommendations on lower-order effects adjustments]**

Replication

*"Leaving aside for the moment **what replication means** or should mean in the context of GWAIS, even for the currently so-called replicated genetic interactions it is unclear to what extent **a false positive has been replicated** due to the adopted methodological strategy itself or whether the replication of epistasis is not solely attributed to main effects (such as HLA effects) not properly accounted for."*

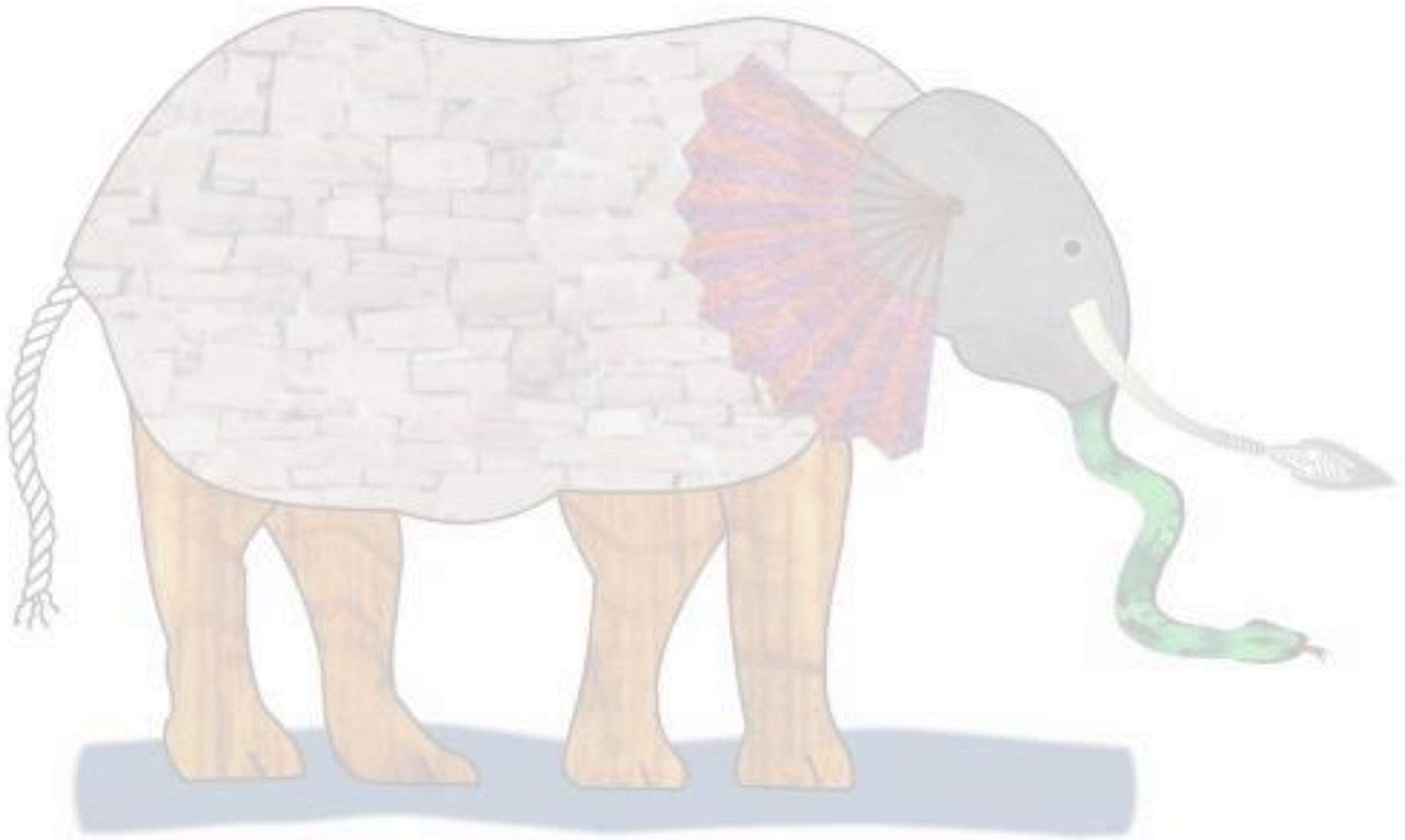


(Ritchie and Van Steen, 2016 – under review)

(MoG-Plot, BIO3 lab – Van Steen)

- **Mahachie John** JM, Van Lishout F, Gusareva ES, Van Steen K (2013) A Robustness Study of Parametric and Non-parametric Tests in Model-Based Multifactor Dimensionality Reduction for Epistasis Detection. BioData Min. 2013 Apr 25;6(1):9**[recommendations on QT analysis]**
- **Van Lishout** F, Mahachie John JM, Gusareva ES, Urrea V, Cleynen I, Theâtre E, Charloteaux B, Calle ML, Wehenkel L, Van Steen K (2012) An efficient algorithm to perform multiple testing in epistasis screening. BMC Bioinformatics. 2013 Apr 24;14:138 **[C++ MB-MDR made faster!]**
- **Gusareva ES,** Van Steen K (2014) Practical aspects of genome-wide association interaction analysis. Hum Genet 133(11):1343-58 **[GWAI analysis protocol]**
- **Bessonov K**, Gusareva ES, Van Steen K (2015**)** A cautionary note on the impact of protocol changes for Genome-Wide Association SNP x SNP Interaction studies: an example on ankylosing spondylitis. Hum Genet - accepted **[non-robustness of GWAI analysis protocols]**
- **Van Lishout F,** Gadaleta F, Moore JH, Wehenkel L, Van Steen K (2015) gammaMAXT: a fast multiple-testing correction algorithm – Nov 20;8:36. doi: 10.1186/s13040-015-0069-x. eCollection 2015. **[C++ MB-MDR made SUPER fast]**
- **Fouladi R**, Bessonov K, Van Lishout F, Van Steen K (2015) Model-Based Multifactor Dimensionality Reduction for Rare Variant Association Analysis. Hum Hered 79(3-4):157-67 **[aggregating based on similarity measures to deal with DNA-seq data]**

What is the “context”?



Context: systems medicine

- Human diseases are “complex systems”: network of interconnected components that may impact each other in non-linear ways

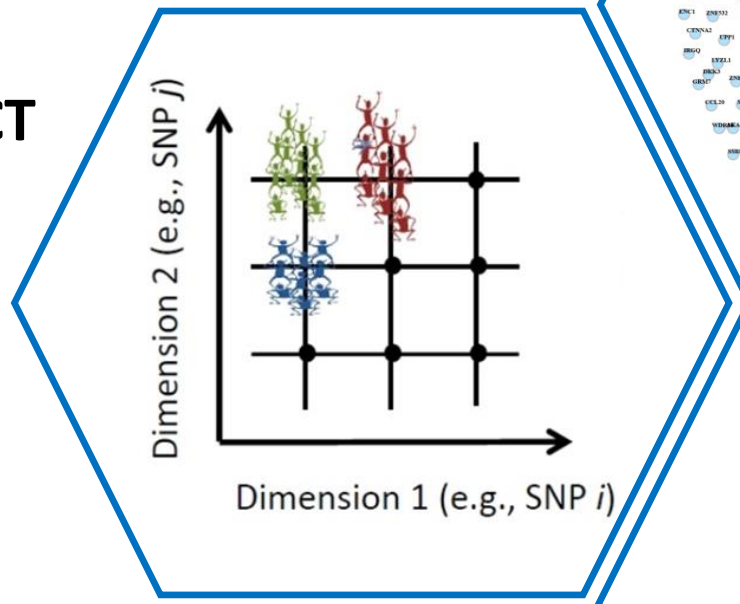
Holistic or systemic view

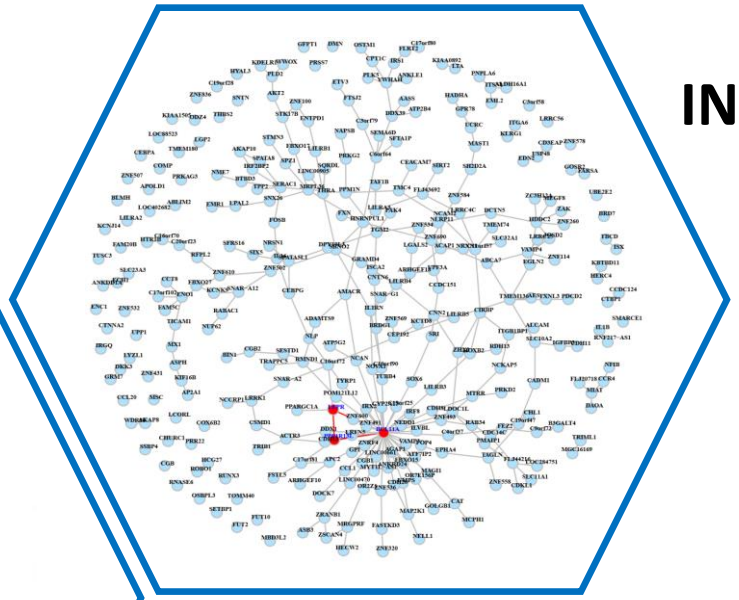
- When “systems” and their properties are not viewed as wholes but broken into parts, (dynamical) relationships will be broken as well.

Pragmatic approach

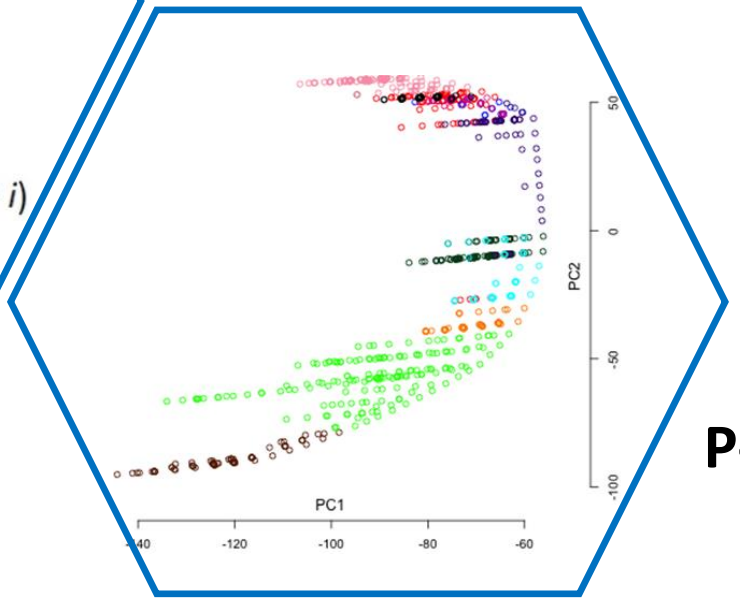
- A boundary needs to be chosen to determine the “whole” one wants to be holistic / systemic about.

DEST*in*CT



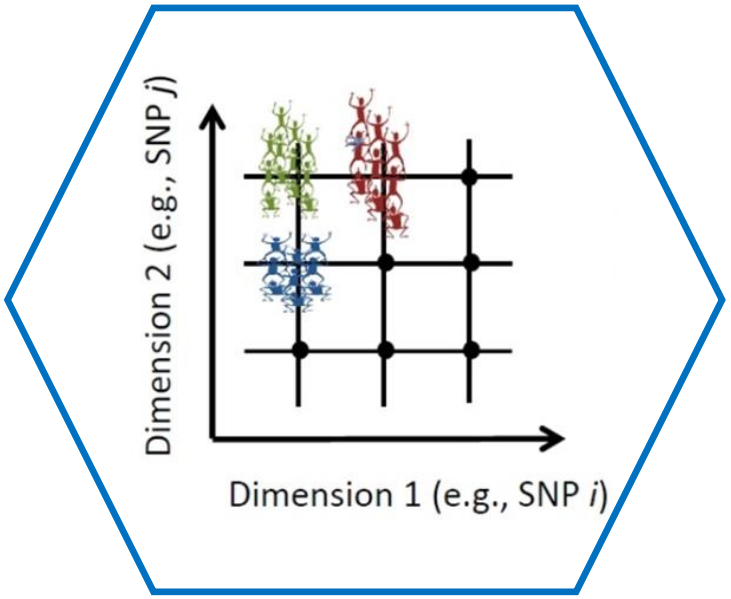


INTEGROMIX

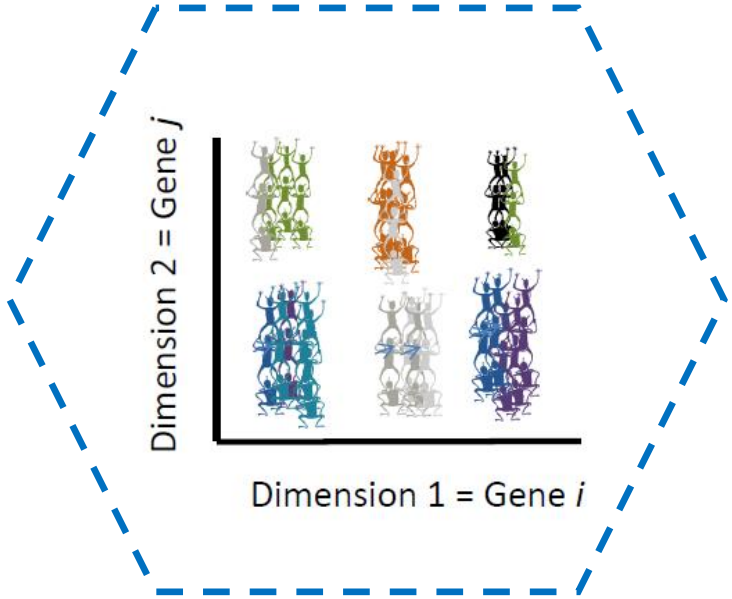


P-STRUCT

DEST*in*CT → INTEGROMIX



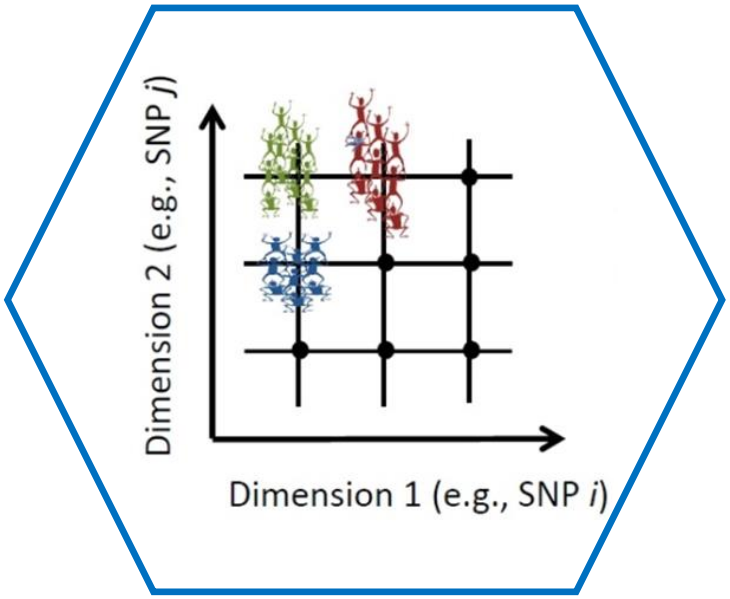
MB-MDR in integrative context



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

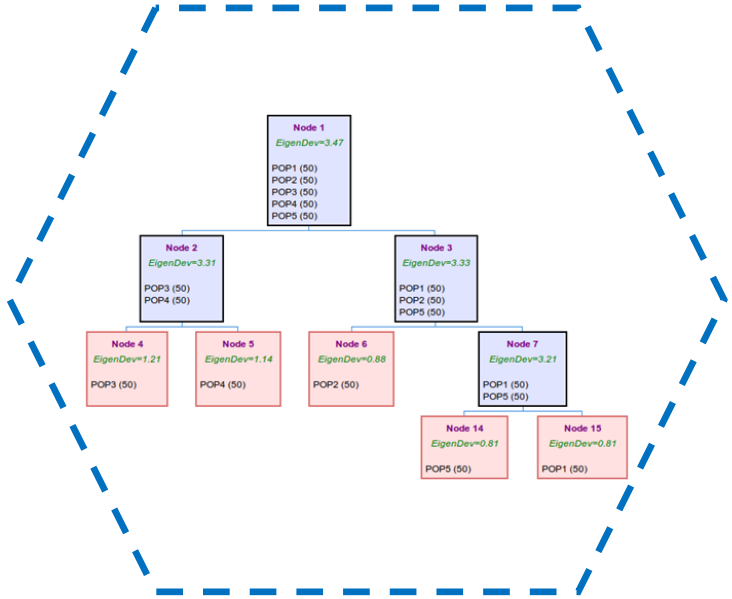
(Fouladi et al. 2015, 2016 - in preparation)

DEST*in*CT → P-STRUCT



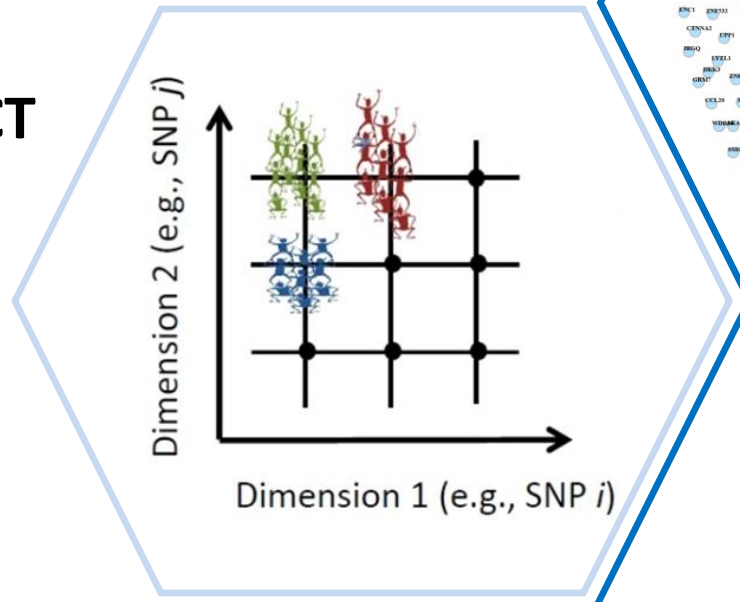
MB-MDR for structured populations

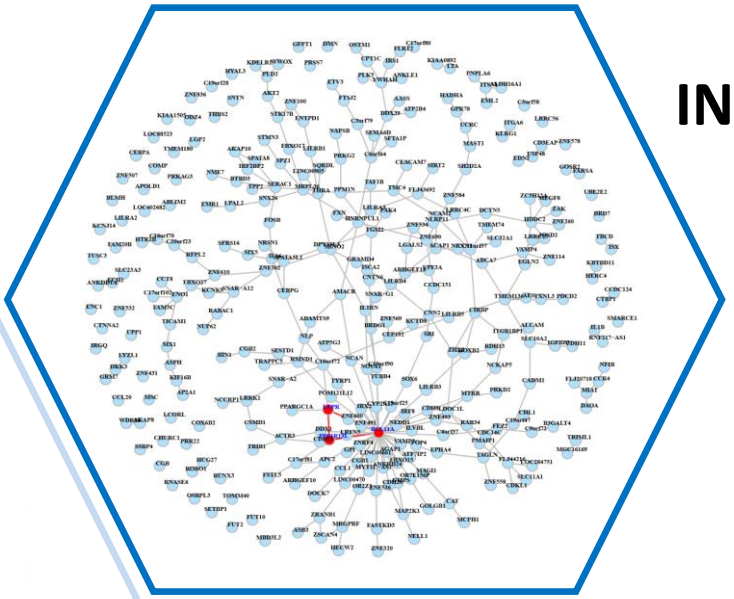
- Continuous axes of confounding
- ipPCA (IP2CAPS)
- Hypothesis-specific genomic control



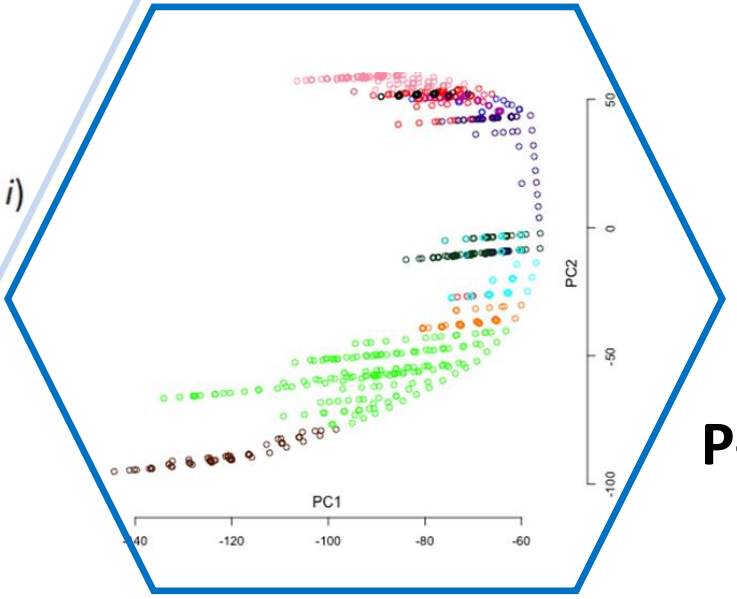
(Chaichoompu et al. 2016 and
Abegaz et al. 2016 - in preparation)

DEST*in*CT





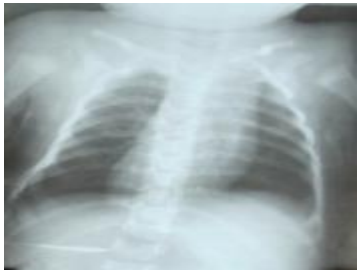
INTEGROMIX



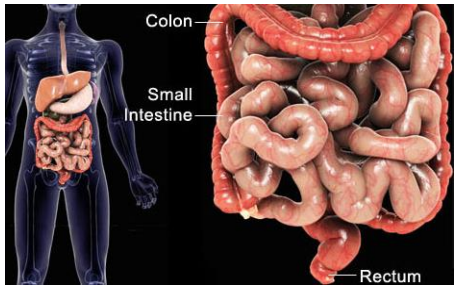
P-STRUCT

Are we ready for translational systemics?

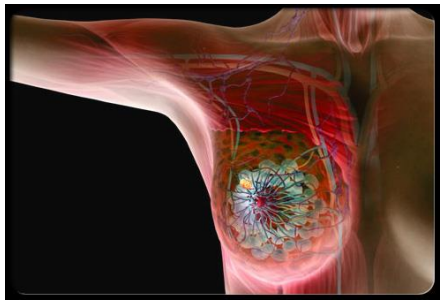
Translating methods to real-life applications



Inflammatory Disorders of Airways; Nature Genetics, 2005 (IF5: 32.138)

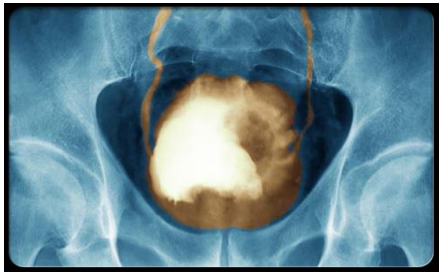


Inflammatory Bowel Disease; Gastroenterology 2002-2016 (IF5: 13.811)

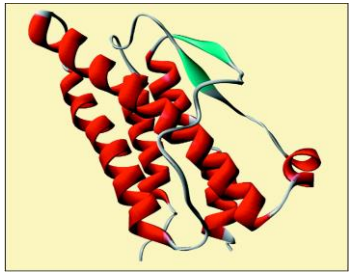


Breast Cancer; Journal of Clinical Oncology, 2004 (IF5: 17.158)

Alzheimer Disease; Neurobiology of Aging, 2014 (IF5: 5.22)

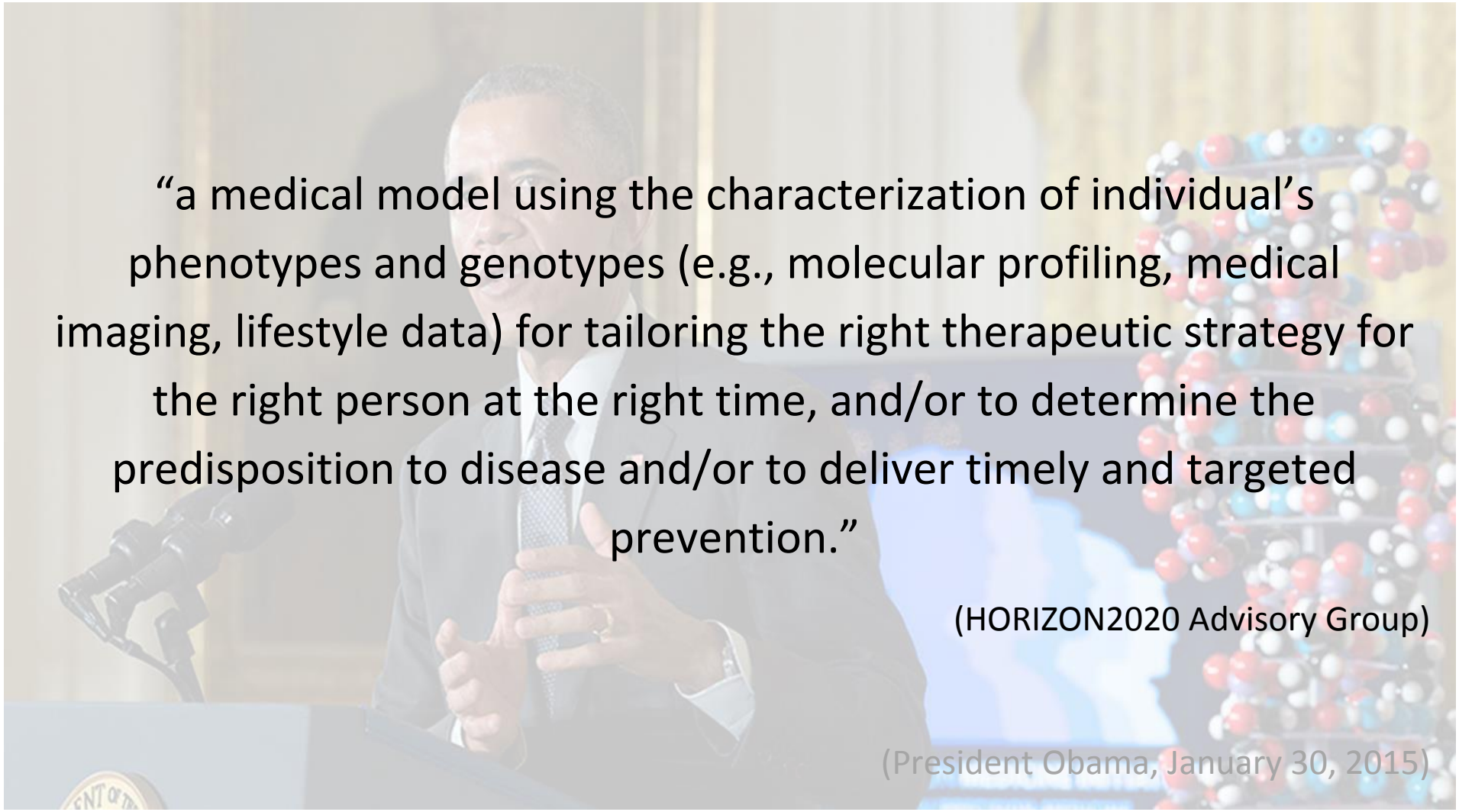


Bladder Cancer; PloS Genetics, 2015 (IF5: 8.56)



Quality of Life; The Lancet Oncolocy, 2002 (IF5: 26.239)

Personalized Medicine

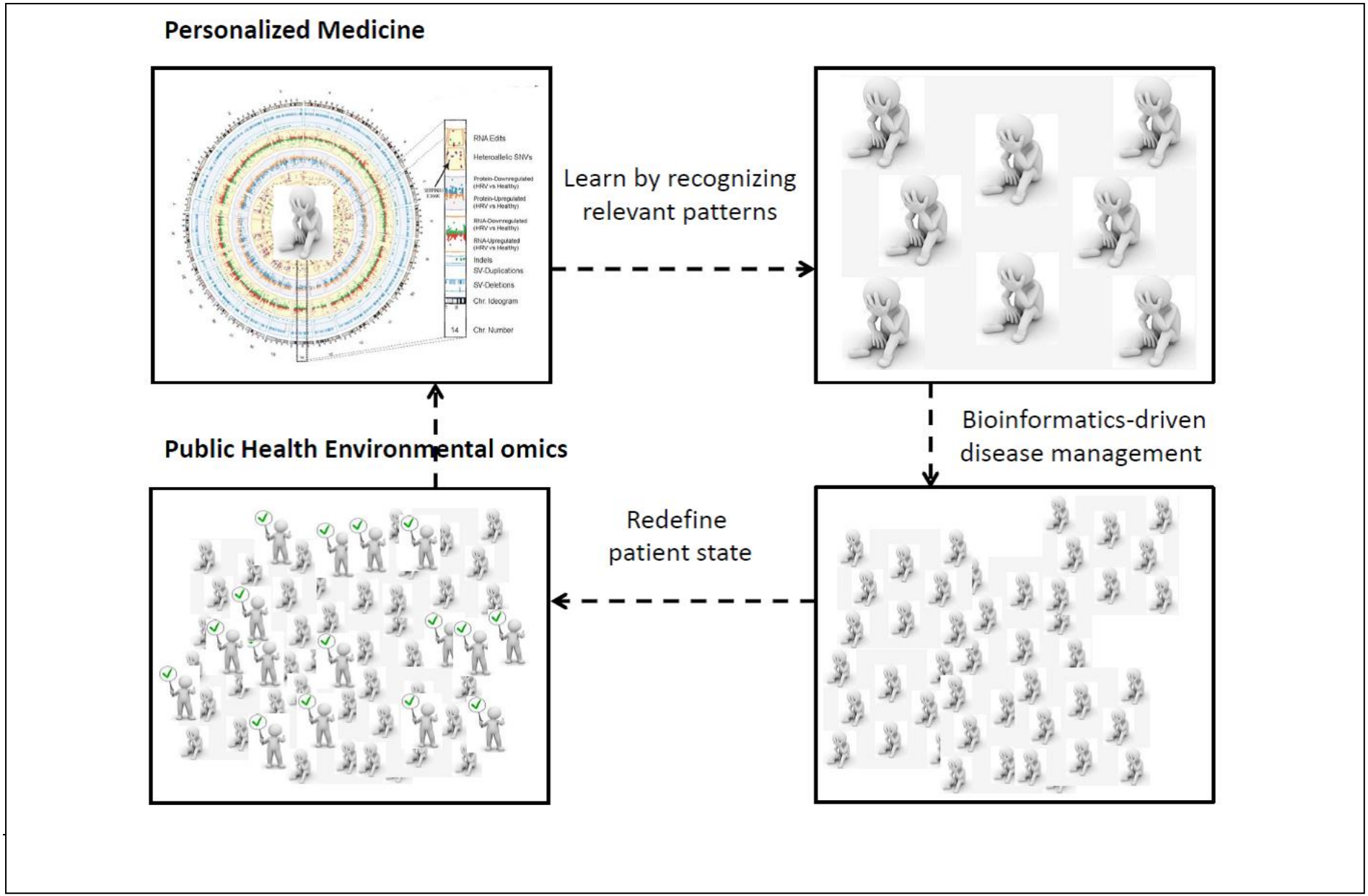


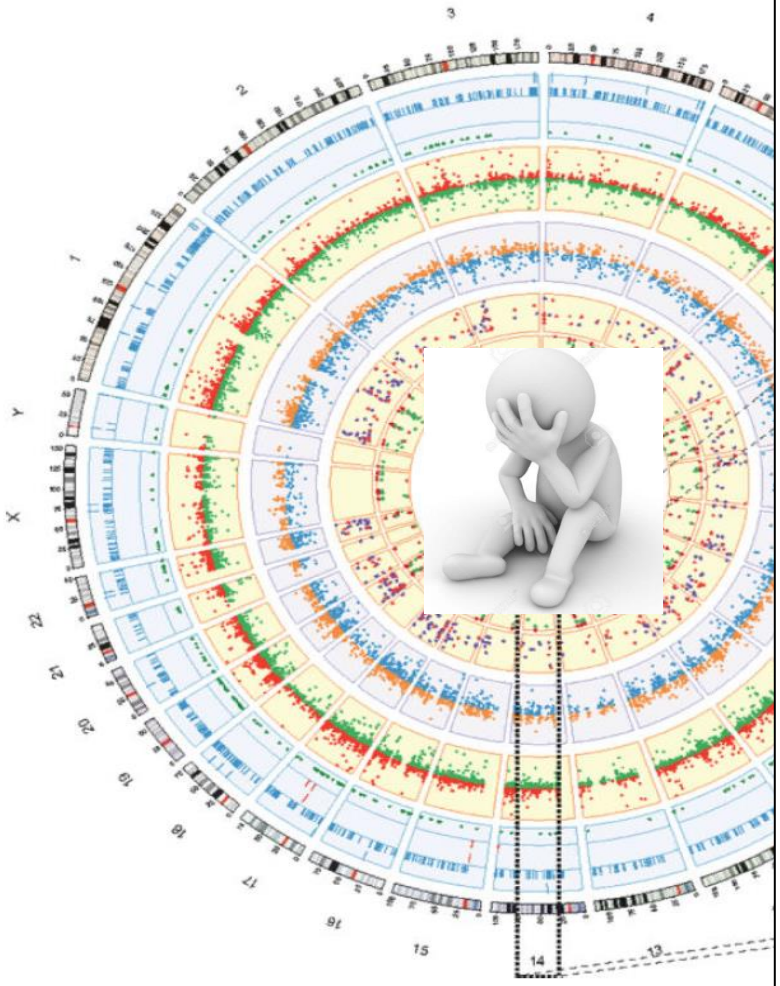
"a medical model using the characterization of individual's phenotypes and genotypes (e.g., molecular profiling, medical imaging, lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely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

(HORIZON2020 Advisory Group)

(President Obama, January 30, 2015)

Systemic thinking in precision medicine





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

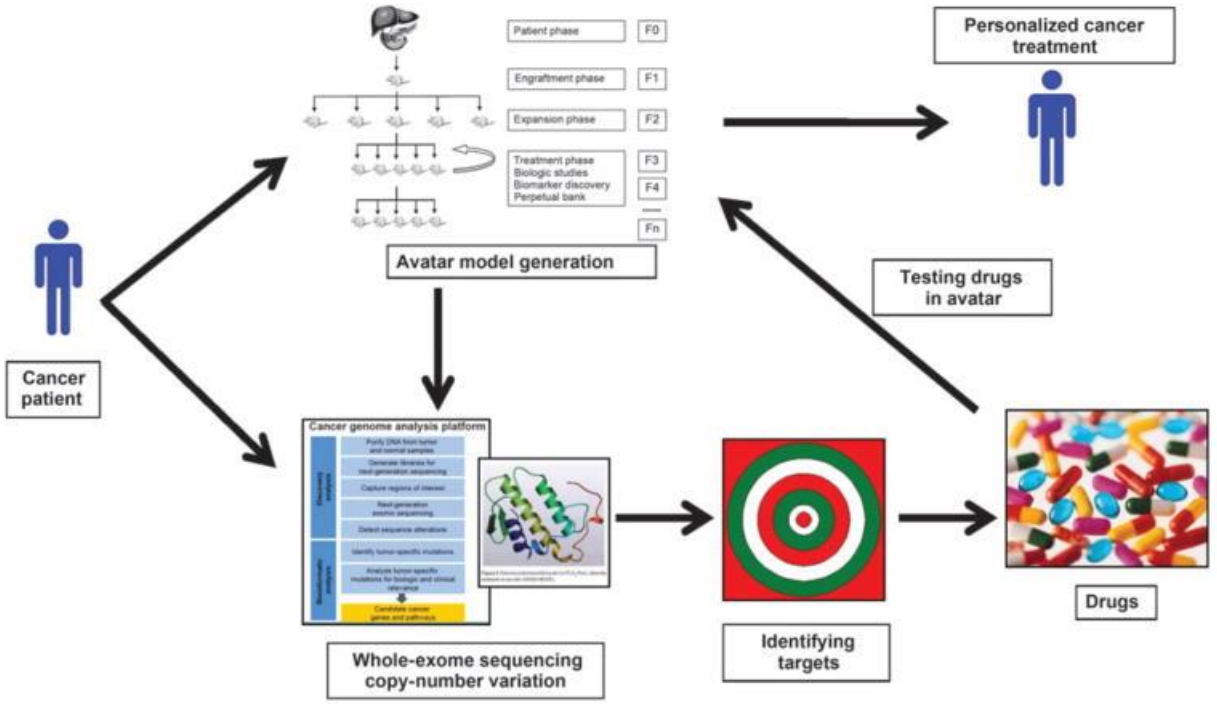
"Not in the form we are doing it – *iPOP (Integrated Personalized Omics Profiling)*.

… 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."

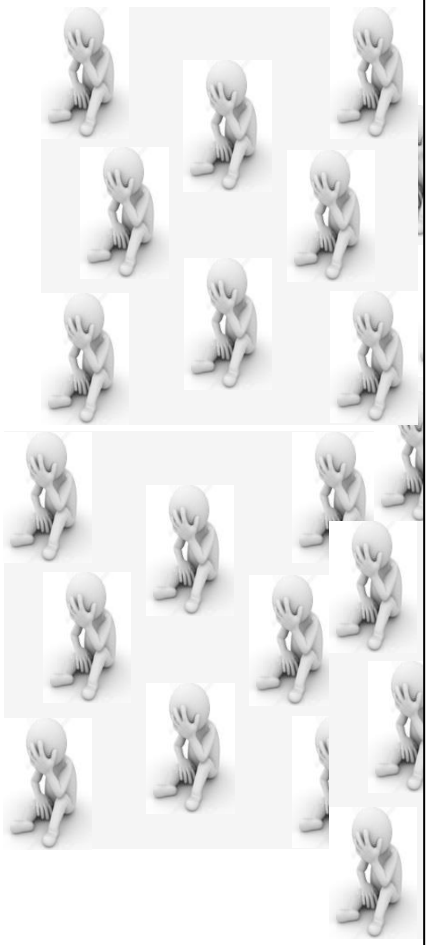
(Snyder 2014)

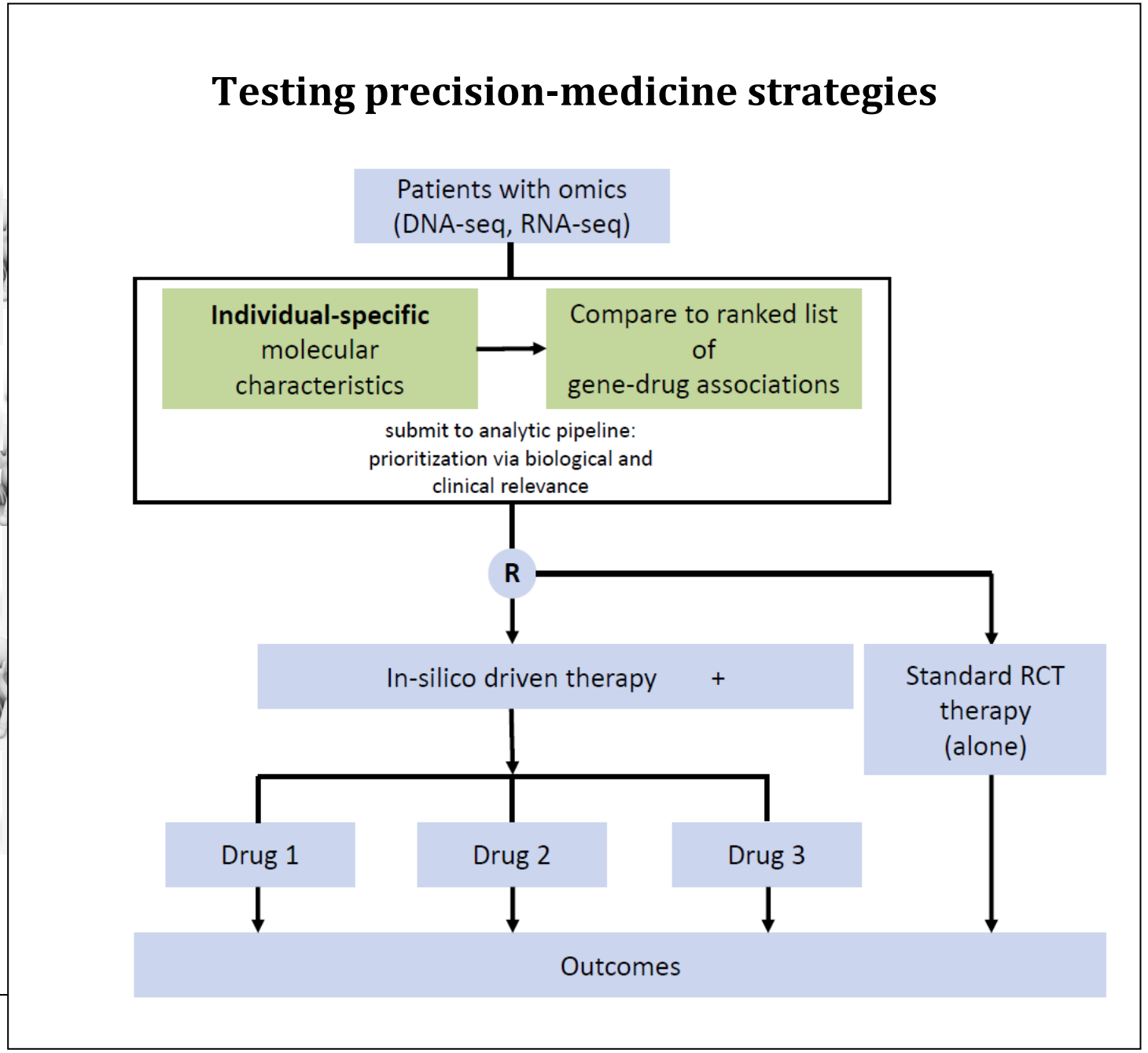
Redundancy - Informativity

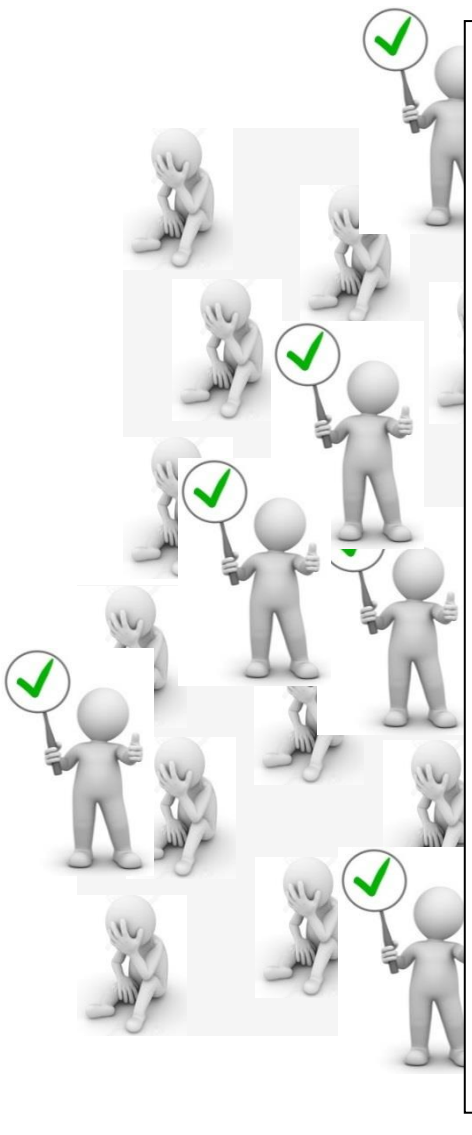
Integrating sequencing and avatar mouse models



(Garralda et al. 2014)




Testing precision-medicine strategies
Patients with omics (DNA-seq, RNA-seq)
Individual-specific molecular characteristics
Compare to ranked list of gene-drug associations
submit to analytic pipeline: prioritization via biological and clinical relevance
R
In-silico driven therapy +
Standard RCT therapy (alone)
Drug 1
Drug 2
Drug 3
Outcomes



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

OPEN ACCESS Freely available online

PLOS ONE

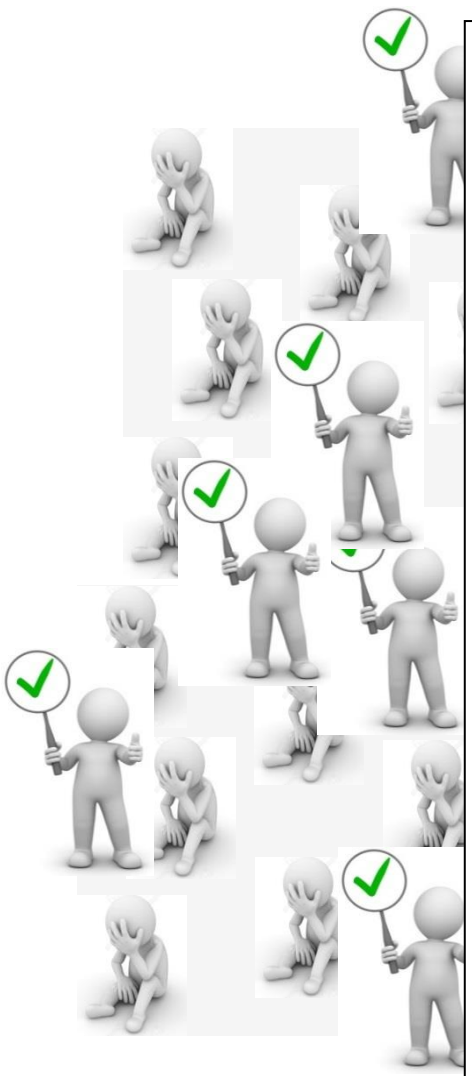
Molecular Reclassification of Crohn’s Disease: A Cautionary Note on Population Stratification

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)

Disease heterogeneity - Disease subtypes



What does it mean to be „Diseased“?

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
MACHINE LEARNING
POPULATION GENETICS

Received
30 September 2014

Accepted
8 January 2015

Highlighting nonlinear patterns in population genetics datasets

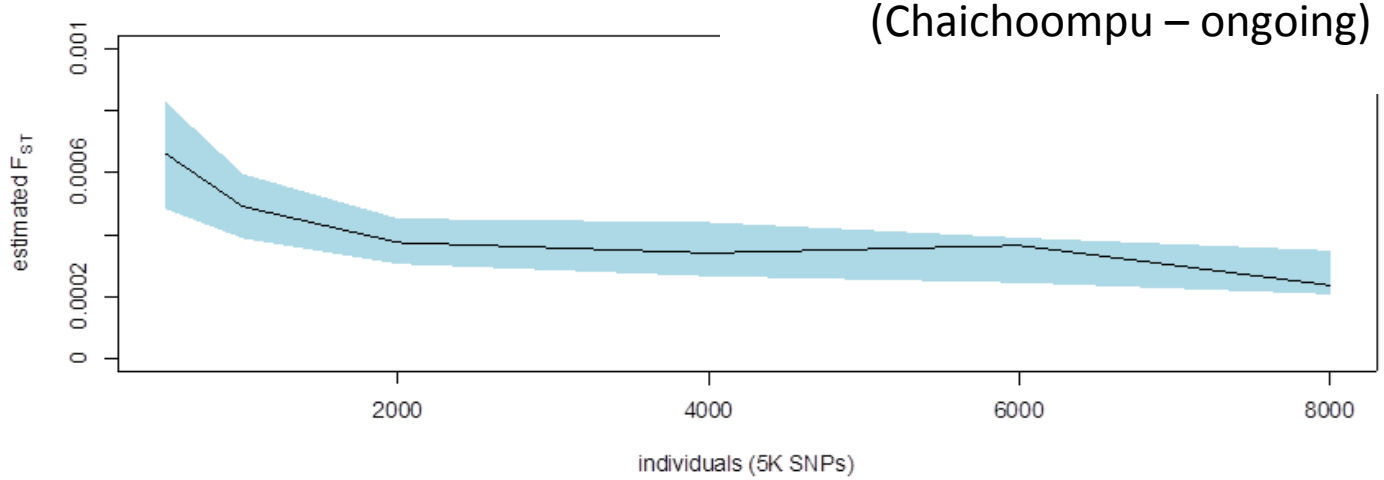
Gregorio Alanis-Lobato[1,2]*, Carlo Vittorio Cannistraci[3]*, Anders Eriksson[1,4], Andrea Manica[4] & Timothy Ravasi[1,2]

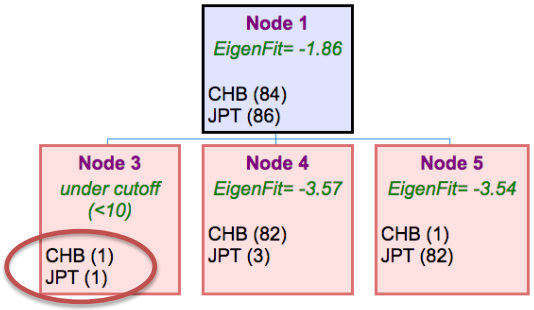
[1]Integrative Systems Biology Laboratory, Biological and Environmental Sciences and Engineering Division, Computer, Electrical and Mathematical Sciences and Engineering Division, Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Ibn Al Haytham Bldg. 2, Level 4, Thuwal 23955-6900, Kingdom of Saudi Arabia, [2]Division of Medical Genetics, Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093 USA, [3]Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Technische Universität Dresden, Tatzberg 47/49, 01307 Dresden, Germany, [4]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, England.

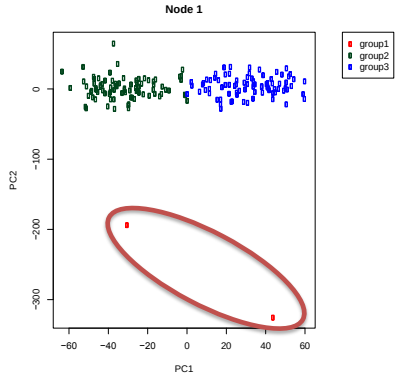
(Alanis-Lobato et al. 2015)

Fine structure

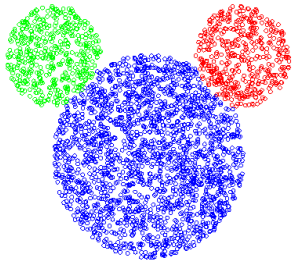
(Chaichoompu – ongoing)



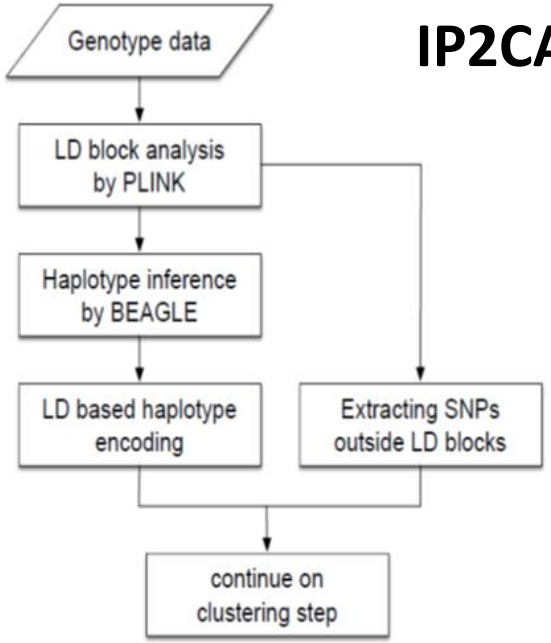




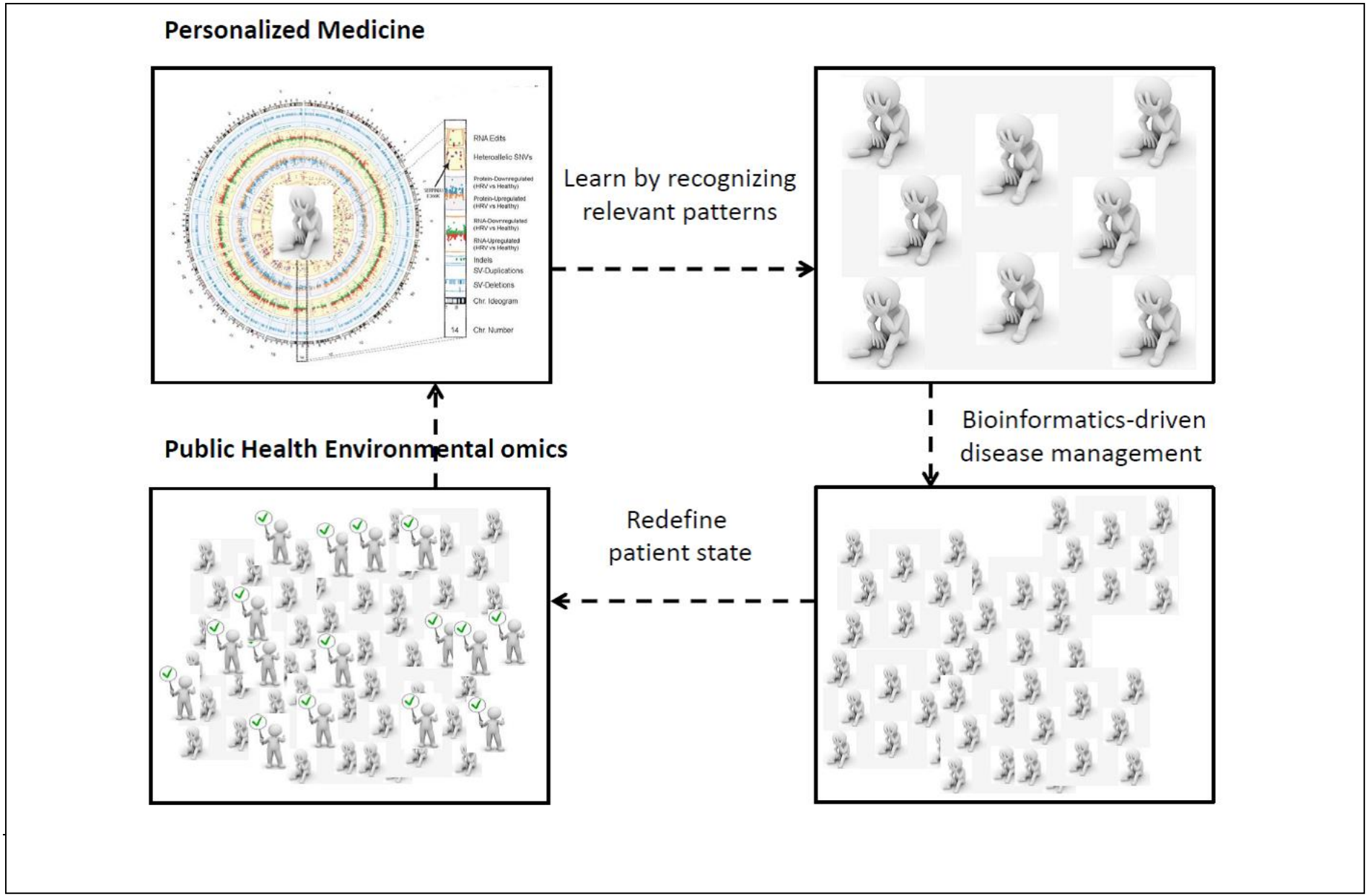
Combine with EM clustering



IP2CAPS

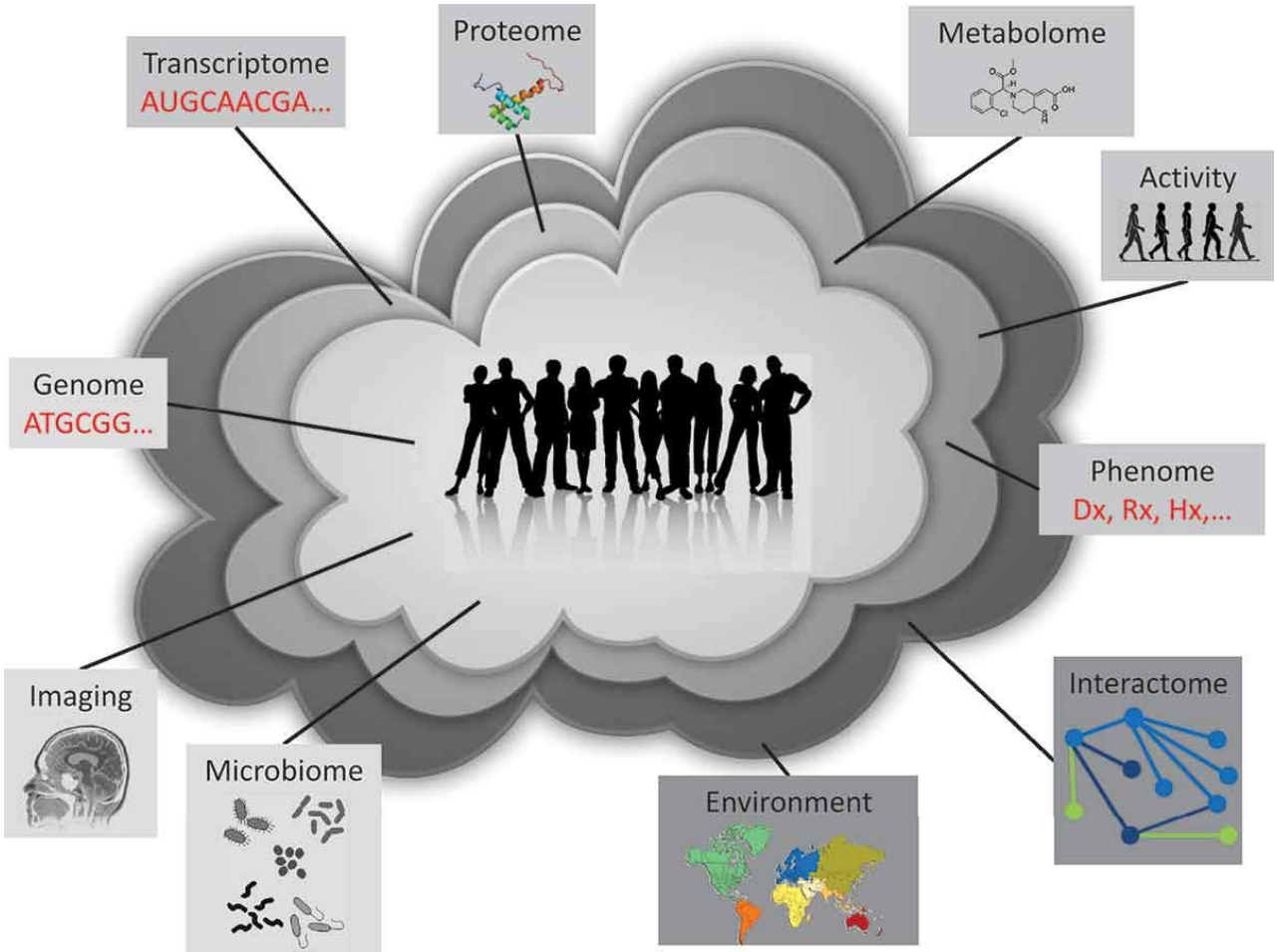


Systemic thinking in precision medicine



What have we learned?

Precision Medicine: defining individual patterns of disease



(Huang et al. 2016)

Take home messages

- Context is important, although grasping this context is not always an obvious thing to do [patient/population substructures]
- Analytic developments involving interactions can benefit a lot from a transdisciplinary viewpoint rather than interdisciplinary viewpoints [hybrid machine learning and statistics]
- Open problems relate to: validation (causality), replication → individual risk prediction, heterogeneity handling (confounding), meta-analysis
- Small systems may be nice as toy examples but are by no means a reflection of the complexity in real life (at least not yet …):

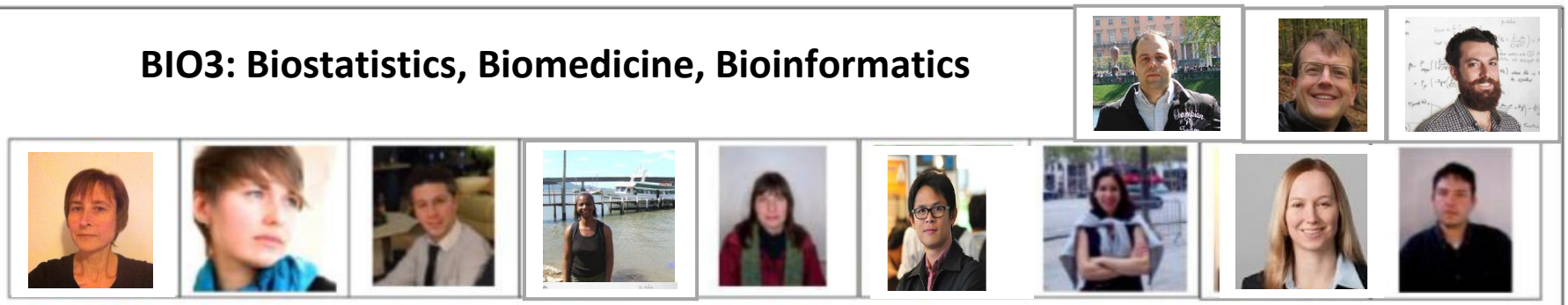
Embrace Complexity

Through the looking-glass

Alice doesn't play by the conventional rules of a little girl during the 1800s; she's up for whatever comes her way and is willing to take a chance on the unexpected with brilliant results.

(Lewis Carroll, 1871)

BIO3: Biostatistics, Biomedicine, Bioinformatics



welbio

Machine Learning for Personalized Medicine

fnrs LA LIBERTÉ DE CHERCHER

EU PANCREAS

télévie

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