Systems and Interactions

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

OUTLINE

- Why looking into “interactions”?
- How to look for “interactions”?
- Case study: pancreatic cancer
- Phenotype refinement

Why looking into “interactions”?

Define the context first

interaction

/ɪntərˈakʃ(ə)n/

noun

reciprocal action or influence.
"ongoing **interaction between** the two languages"

Similar: interplay | interchange | interactivity | interconnection | interlinkage

- communication or direct involvement with someone or something.
  "for a shy person, social interaction can be a stomach-churning, anxiety-filled experience"

  Similar: contact | relations | connection | association | communion | intercourse

- PHYSICS
  a particular way in which matter, fields, and atomic and subatomic particles affect one another, e.g. through gravitation or electromagnetism.

Biological context

- Biological interactions are the effects that organisms in a community have on one another.
- In the natural world no organism exists in absolute isolation
- The black walnut secretes a chemical from its roots that harms neighboring plants, an example of competitive antagonism
  (en.wikipedia.org)

Precision Medicine context

PREVENTION

DISEASE DIAGNOSIS

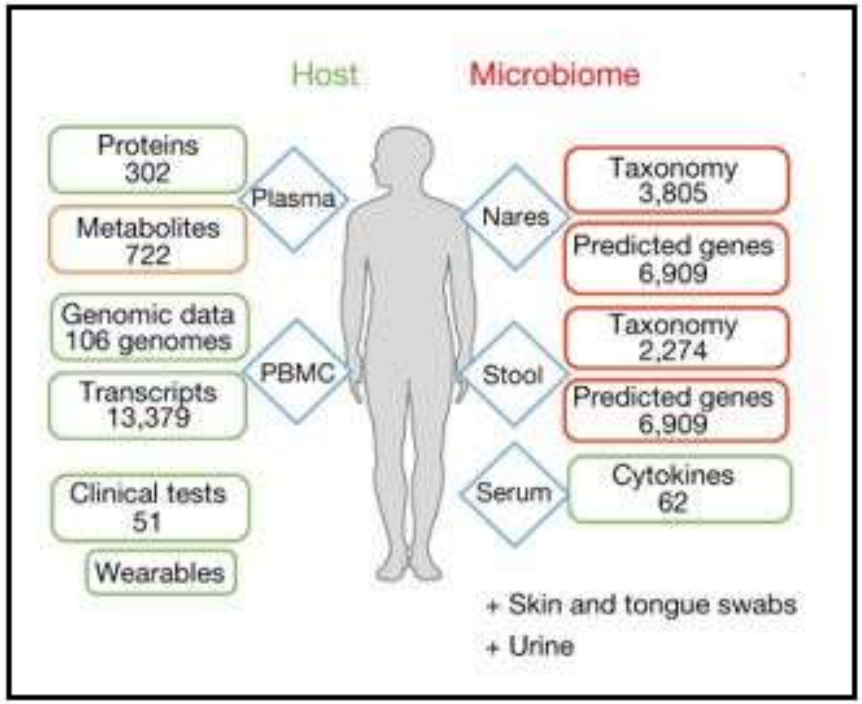
DISEASE MANAGEMENT

STATIC

DYNAMIC

Individual characterisations: beyond single views

- Prevailing procedure: population-based data integration
  - Prediction of health outcomes (from samples to individual);
  - Inter-personal variability (incl. identification of endotypes)

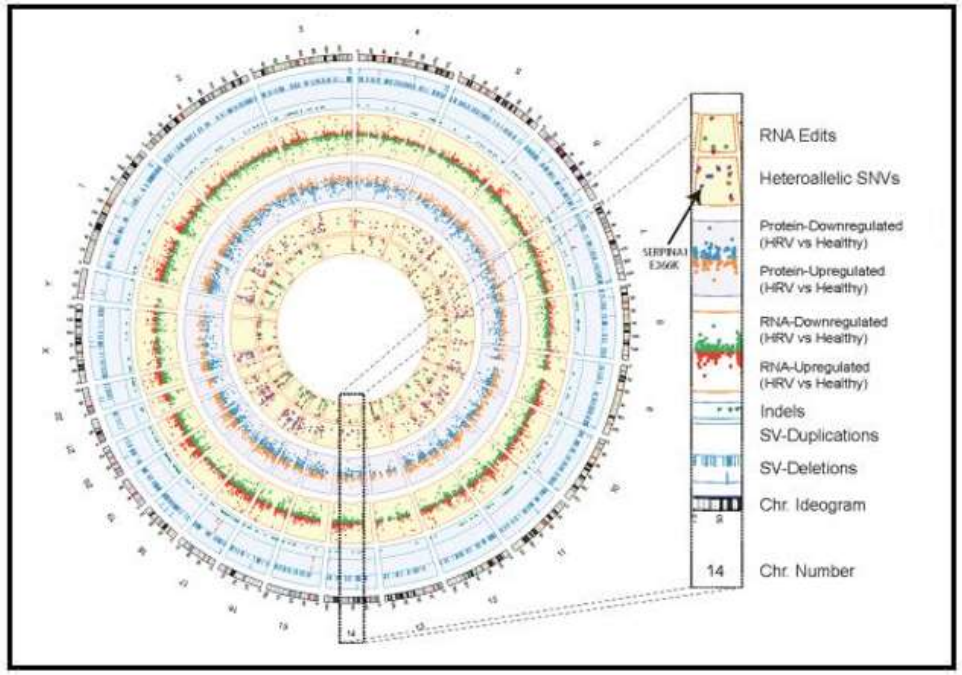


>> 1 individual

(Zhou et al. 2019)
doi : 10.1038/s41586-019-1236-x

Individual characterisation: beyond single views

- Individual as own control: time coure data
  - Dynamic multi-view picture of the individual (incl. informativity versus redundancy);
  - Early detection of disease (forecasting)

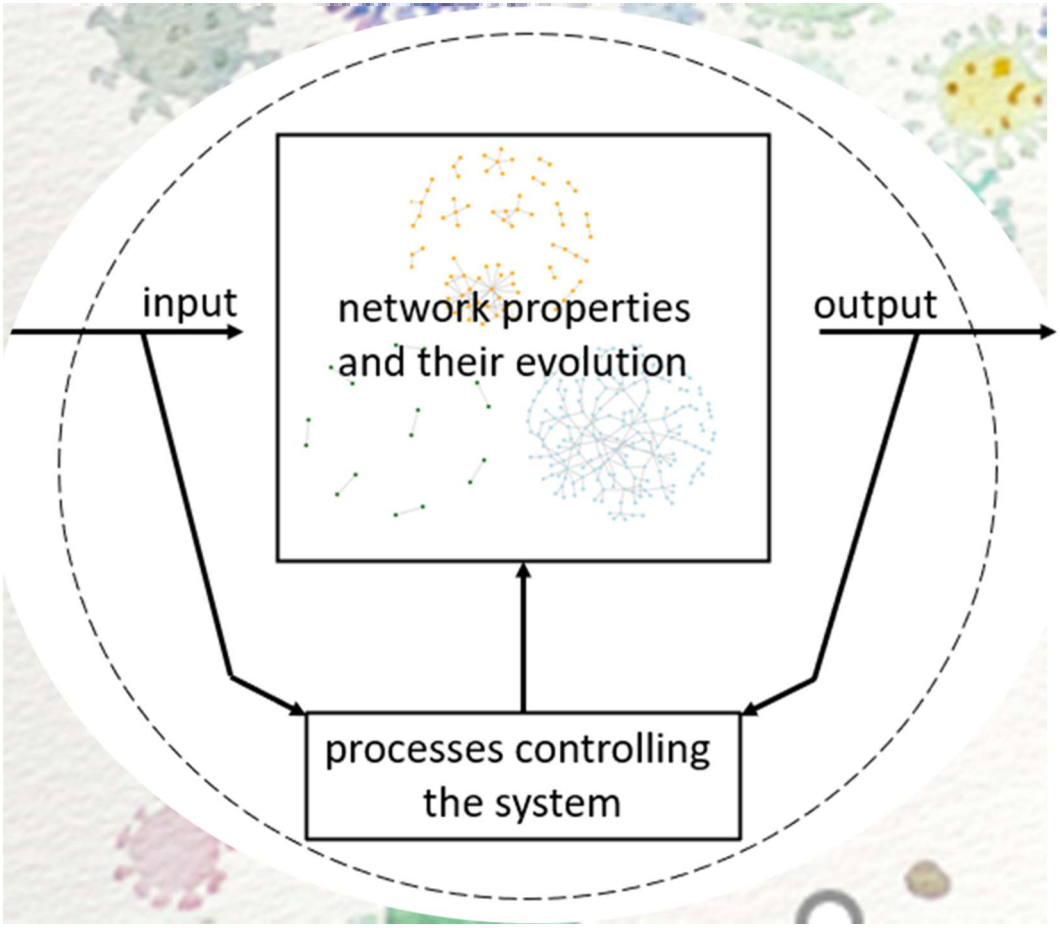


1 individual

(Chen et al. 2012)
doi: 10.1016/j.cell.2012.02.009

Individual characterisation: dependencies

(Ackoff, 1971)
doi.org/10.1287/mnsc.17.11.661

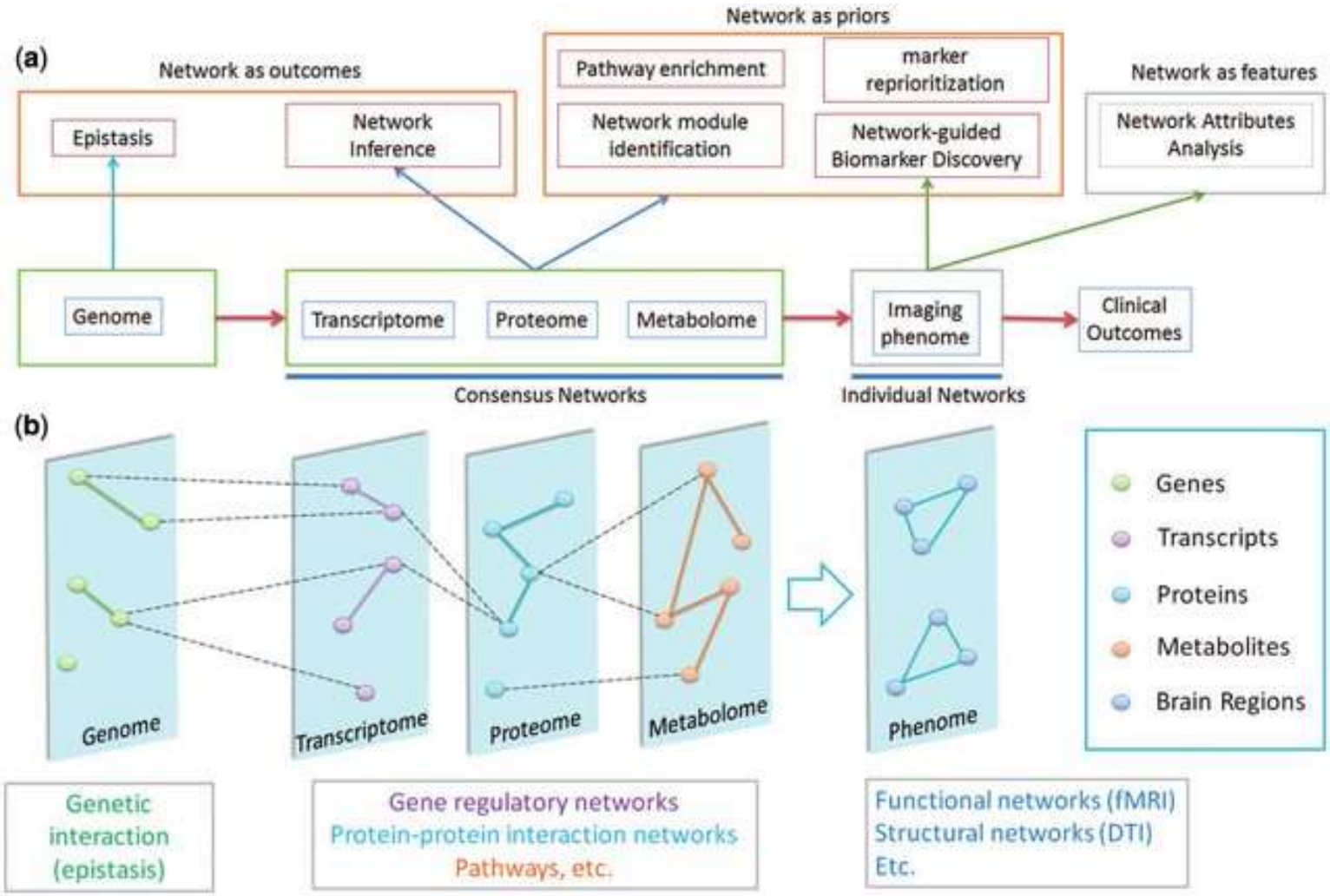


Elements of a system (a set of ≥ 2 interrelated items):

- Boundary
- Environment
- Observable interactions
- Subsystems
- Control mechanisms

Networks - unifying data integration and interactions

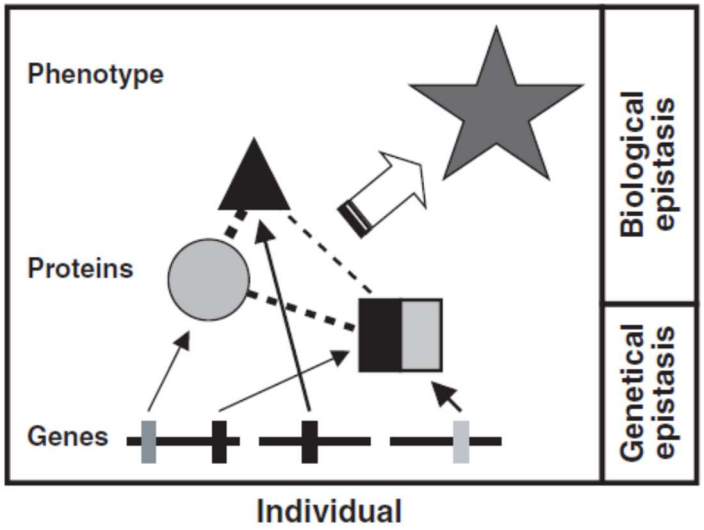
- Nodes:
  - biological features such as DNA-based molecular markers (SNPs →genes), microbial taxa (abundance of a microbial taxon), genes (expression level), metabolites (concentration), and proteins (concentration);
  - environmental (exposures) or other host features (demographics)
- Edges (connections between nodes):
  - empirically or statistically derived interactions;
    more generally: association between nodes s.a. correlation between the abundance of two taxa/dependencies
  - weights to reflect association strength
  - directions to reflect "cause and effect"



(Yan et al. 2017)

Understanding systems (individuals) == understanding interactions

- Genetical interactions (assuming a “tight” boundary)
- 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)

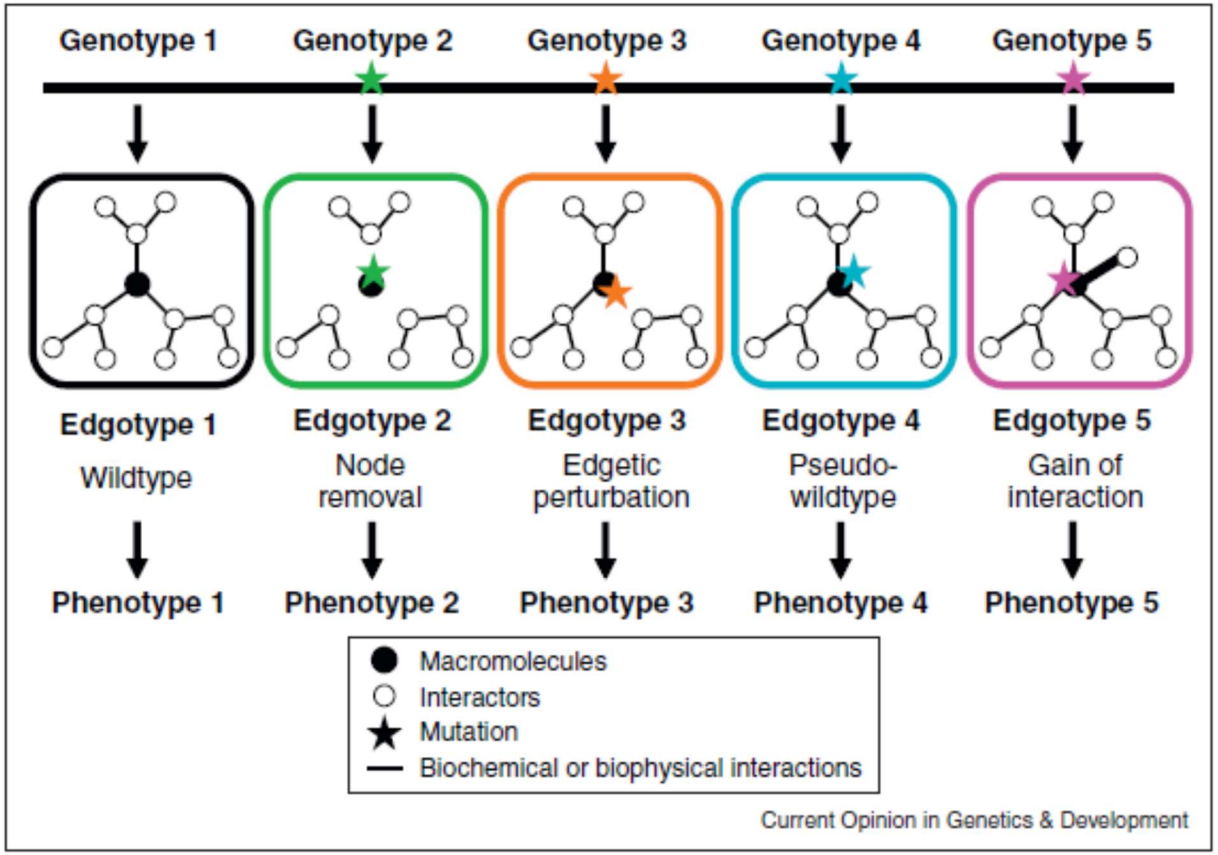


(Moore 2005)

DNA variations: increase interactome – phenotype insights

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***.

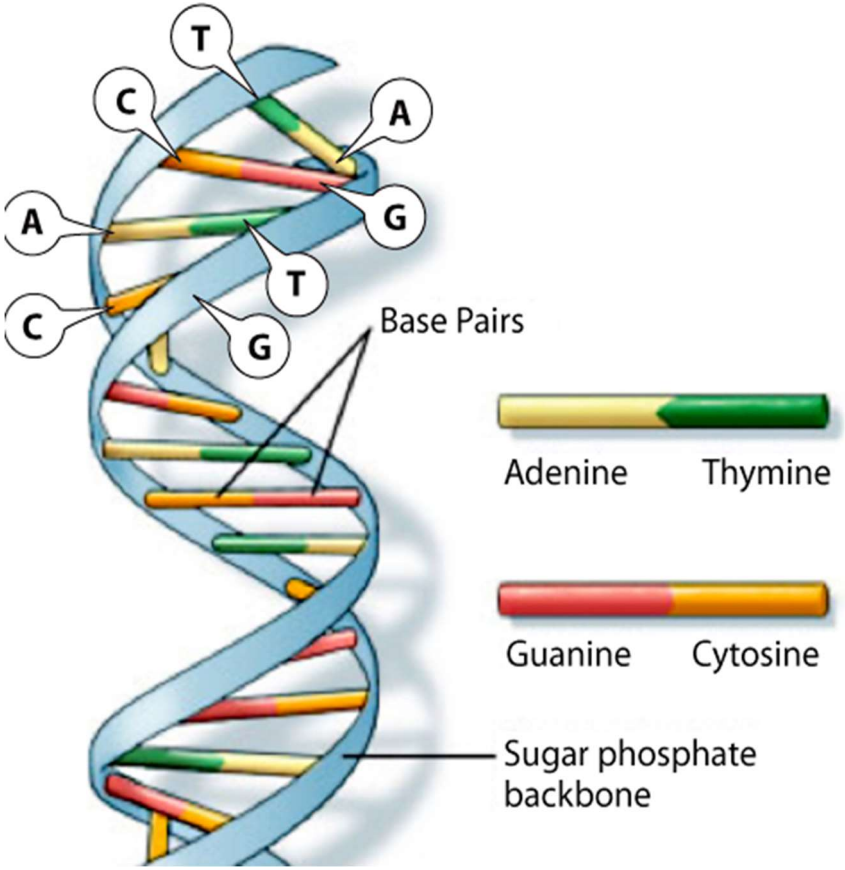
Gateway to improved interactome – phenotype insights



(Sahni et al. 2013)

How to look for " interactions"?

GWAS available DNA variations: single nucleotide polymorphisms



Epistasis

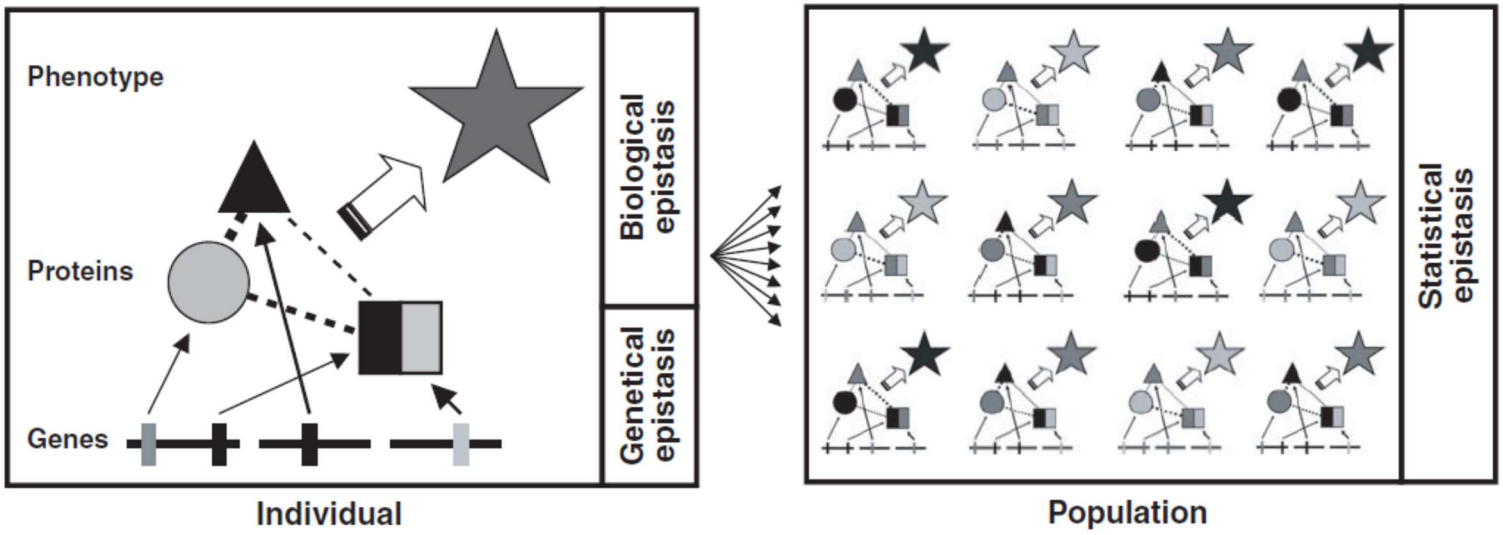
(Moore 2005; Moore and Williams 2005)

- 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).
- A later definition of epistasis (**driven by statistics**) is expressed in terms of deviations from a model of additive multiple effects (Ronald Fisher 1890-1962).

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

Analtyic models to highlight “statistical” epistasis

- 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)

(Logistic) Regression

- Alternatively, we can assume additive effects of each allele at each locus, leading to a single interaction term (instead of 4 next!)

		Locus H	
Locus G	2	1	0
2	$\beta_0 + 2\beta_G + 2\beta_H + 4\beta$	$\beta_0 + 2\beta_G + \beta_H + 2\beta$	$\beta_0 + 2\beta_G$
1	$\beta_0 + \beta_G + 2\beta_H + 2\beta$	$\beta_0 + \beta_G + \beta_H + \beta$	$\beta_0 + \beta_G$
0	$\beta_0 + 2\beta_H$	$\beta_0 + \beta_H$	β_0

- This corresponds in statistical analysis packages to the model

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_G X_1 + \beta_H X_2 + \beta X_1 X_2$$

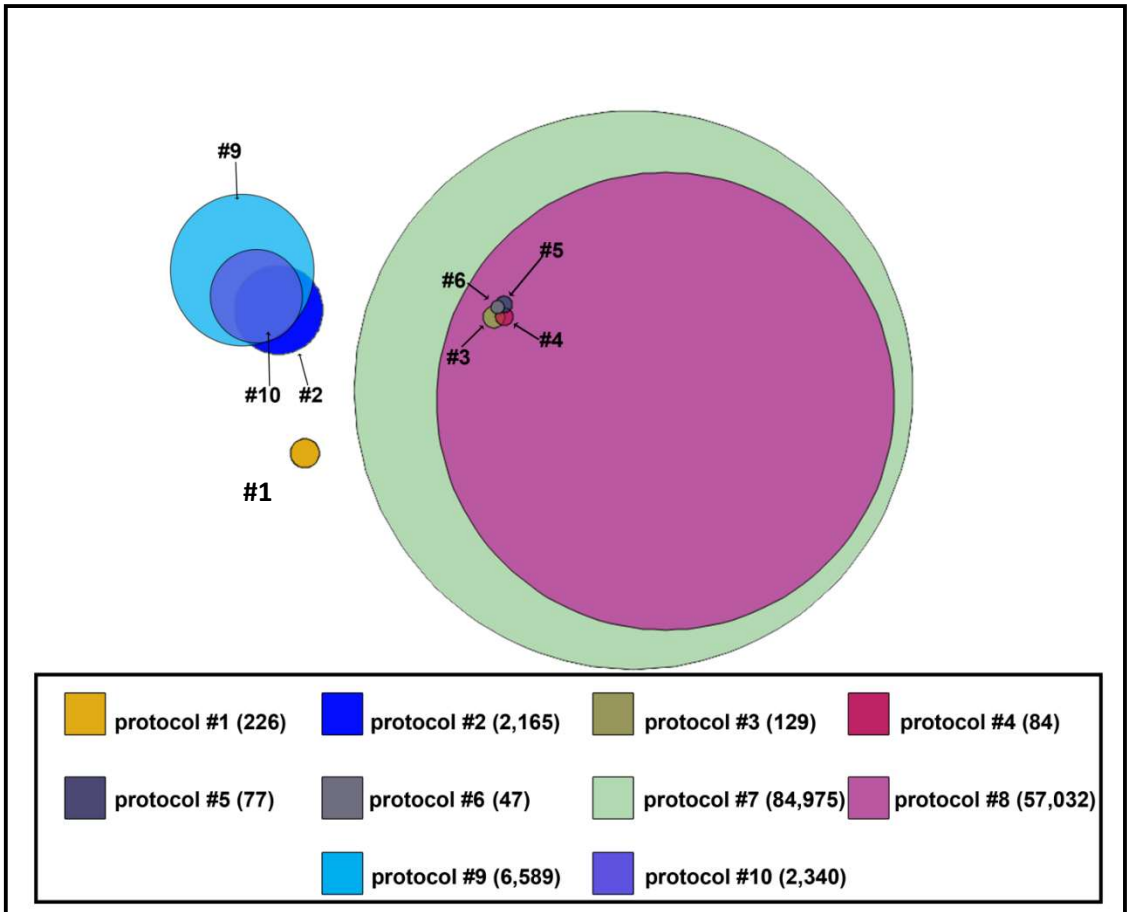
and **dosage encoding for X1 and X2**.

(Logistic) Regression

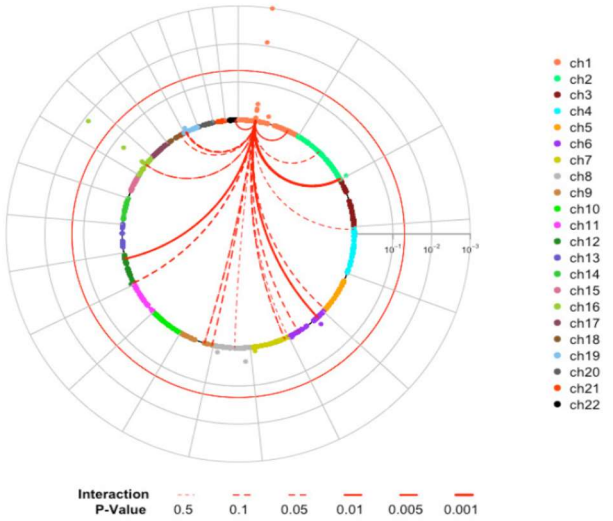
- Most general saturated (9 parameter) genotype model allows all 9 penetrances to take different values
- Log odds is modelled in terms of a baseline effect (β_0), main effects of locus *G* (β_{G1}, β_{G2}), main effects of locus *H* (β_{H1}, β_{H2}), 4 int. terms
- This corresponds in statistical analysis packages to **encoding X1, X2 (0,1,2) as a "factor"**

	Locus H		
Locus G	2	1	0
2	$\beta_0+\beta_{G2}+\beta_{H2}+\beta_{22}$	$\beta_0+\beta_{G2}+\beta_{H1}+\beta_{21}$	$\beta_0+\beta_{G2}$
1	$\beta_0+\beta_{G1}+\beta_{H2}+\beta_{12}$	$\beta_0+\beta_{G1}+\beta_{H1}+\beta_{11}$	$\beta_0+\beta_{G1}$
0	$\beta_0+\beta_{H2}$	$\beta_0+\beta_{H1}$	β_0

Importance of SNP encoding scheme (Ankylosing Spondylitis; WTCCC2 - ~2000 cases + 5000 controls)



(extended from Bessonov et al. 2016)

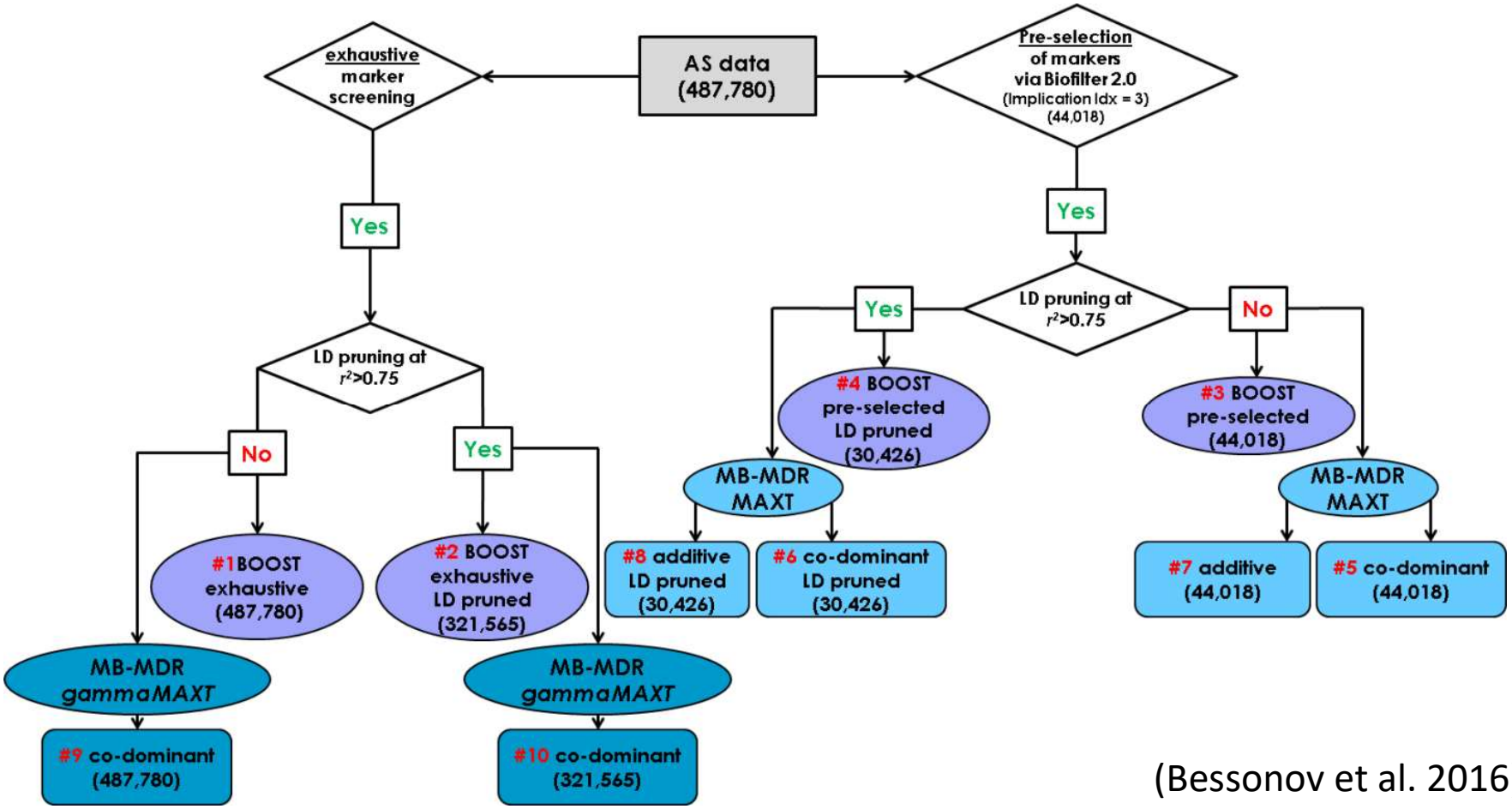


MOGPLOT

Manhattan plot (main effects) over GRAIL plot (interactions)

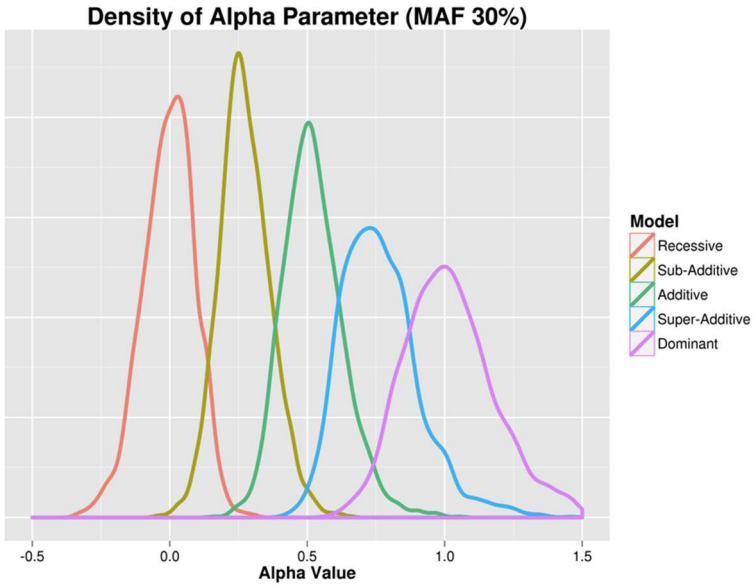
(Van Lishout et al. 2016)

Importance of SNP encoding scheme (Ankylosing Spondylitis; WTCCC2 - ~2000 cases + 5000 controls)



(Bessonov et al. 2016)

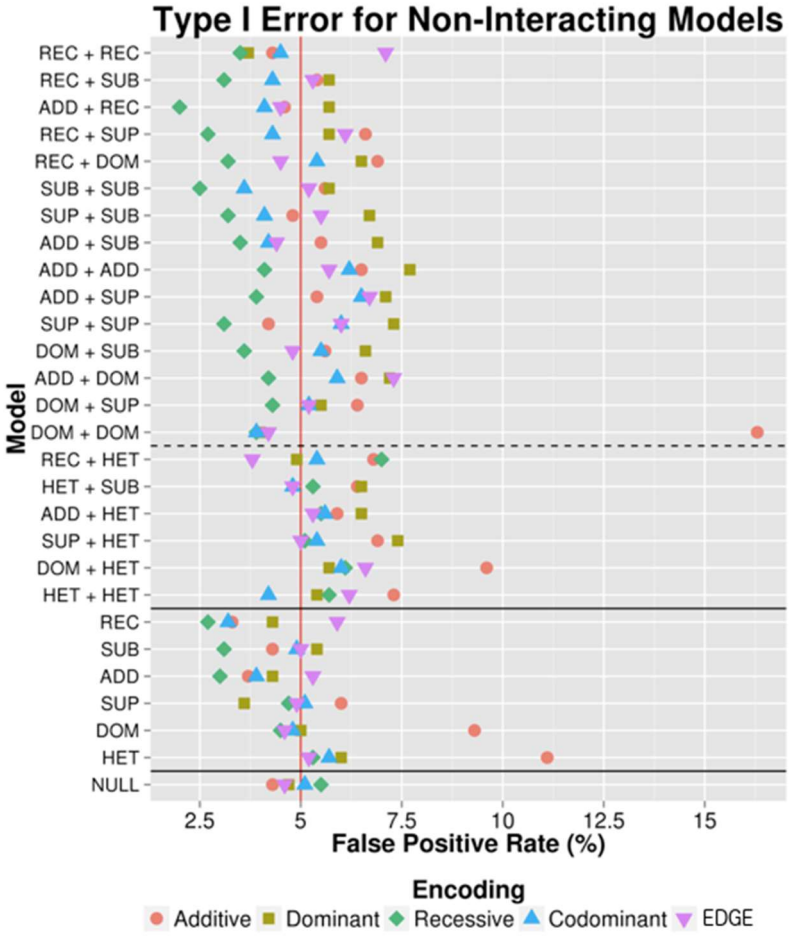
Importance of SNP encoding scheme (Hall et al. 2021)



"Biological Action"	Homozygous Referent	Heterozygous	Homozygous Alternate
EDGE	0	α	1

$$Y \sim \beta_{Het} SNP_{Het} + \beta_{HA} SNP_{HA}$$

$$\alpha = \beta_{Het} / \beta_{HA}$$



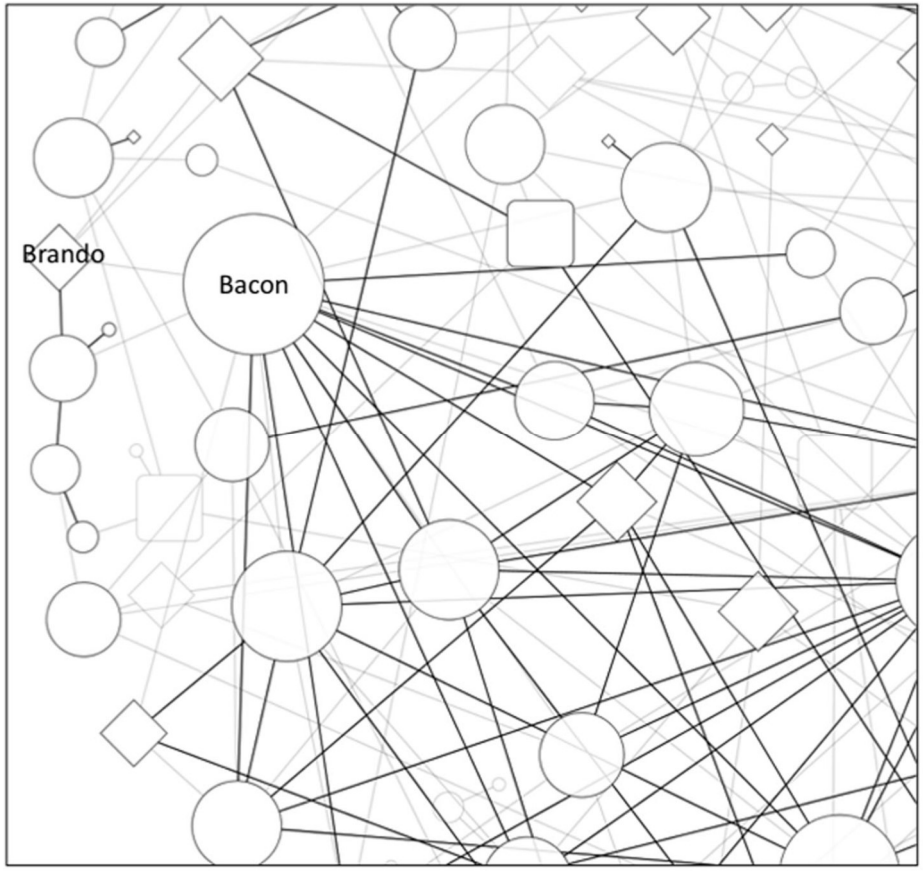
Importance of SNP encoding scheme (Hall et al. 2021)

Table 1. Examples of possible proportional genotype risk underlying genetic loci

Biological Action	Homozygous Referent (AA)	Heterozygous (Aa)	Homozygous Alternate (aa)
Recessive (REC)	0%	0%	100%
Sub-Additive (SUB	0%	25%	100%
Additive (ADD)	0%	50%	100%
Super-Additive (SUP)	0%	75%	100%
Dominant (DOM)	0%	100%	100%

Disappointing results for human complex traits [… to date with SNPs]

Expectations of the first hour



Edges represent small gene–gene interactions between SNPs.

Gray nodes and edges have weaker interactions.
Circle nodes represent SNPs that do not have a significant main effect.
The diamond nodes represent significant main effect association.

The size of the node is proportional to the number of connections.

(McKinney et al 2012)

Disappointing results for human complex traits [… to date with SNPs]

Expectations of the first hour seem to be poorly met:

- Hard to see the forest for the trees: explosion of methodological approaches
- Widely accepted protocol to perform a Genome-Wide Association Interaction Study (GWAIS) is still lacking

  Possible explanations:

  - many difficulties (technical, statistical, computational) involved in performing large-scale epistasis screening
  - and in inferring biological evidence from statistical findings

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]

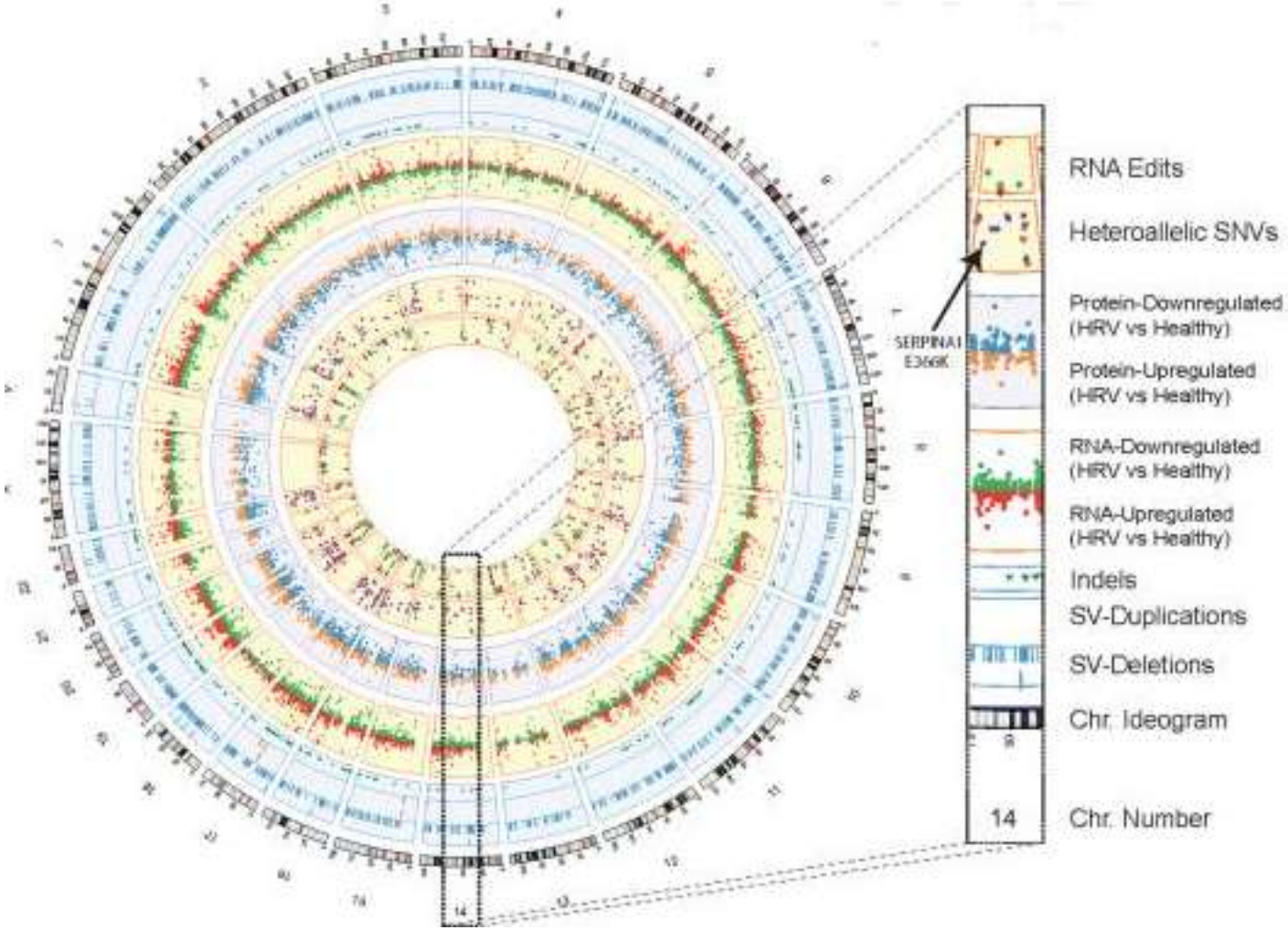
Received: 26 July 2018 / Accepted: 20 February 2019 / Published online: 6 March 2019
© The Author(s) 2019

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.

Case study

Data context: Bioinformatics data availability (Chen et al. 2012)



Disease context: complex “complex diseases”

Addressing complexity in “complex diseases” - pancreatic cancer

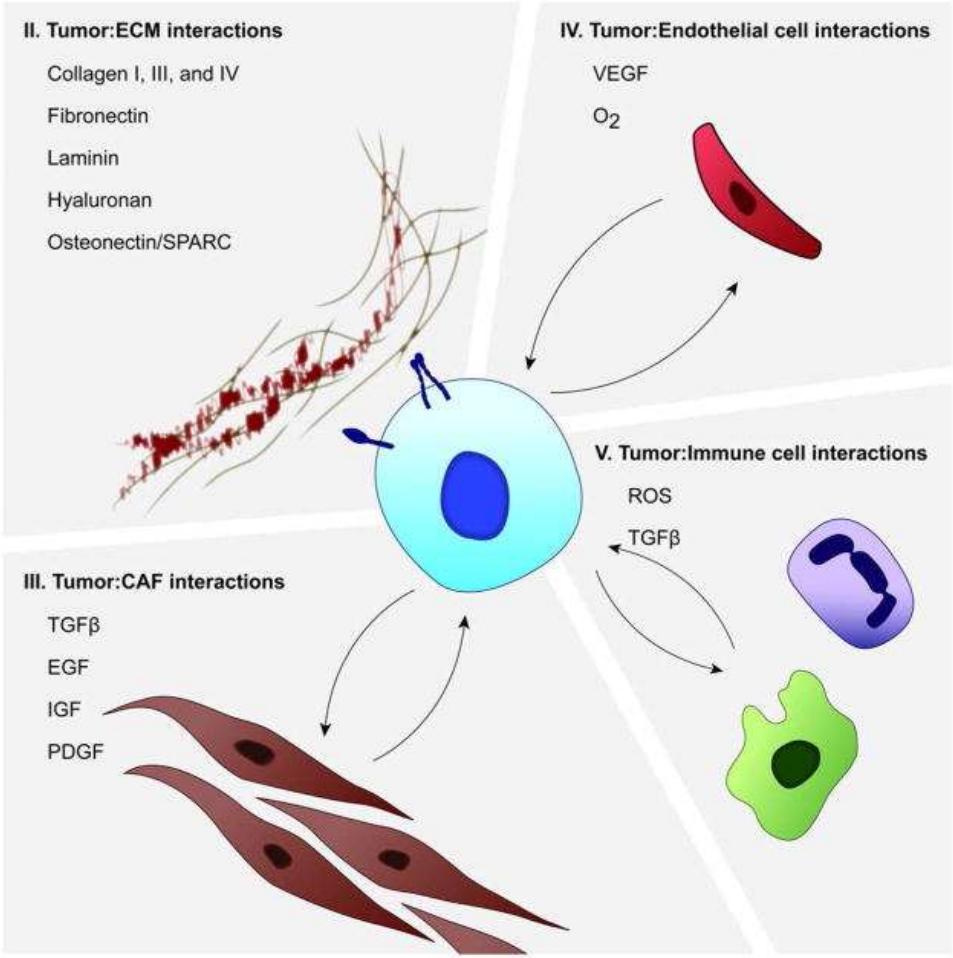
“Because effective systemic therapy capable of controlling the aggressive pancreatic cancer biology is currently lacking, the need for a better understanding of detailed mechanisms underlying pancreatic cancer development and progression is **URGENT***”*

(Xie and Xie 2015)

Examples of interactions in pancreatic cancer

Tumor-stromal interactions

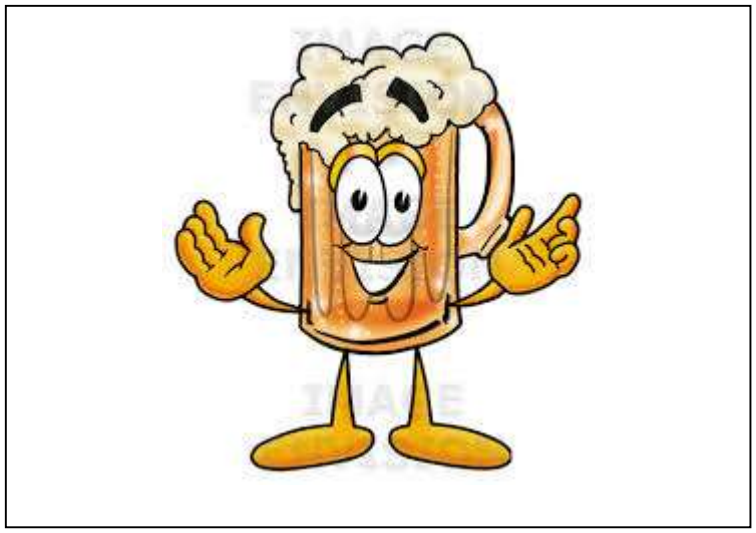
- Treatments focusing on pancreatic cancer cells alone have failed to significantly improve patient outcome over many decades
- Research efforts have now moved to understanding the pathophysiology of the stromal reaction and its role in cancer progression



(Whatcott et al. 2014)

Gene-environment interactions

(Jansen et al. 2015)



Formal definition of gene-environment interactions

- Also gene-environment interactions can be defined in a statistical or a biological way.
- A **biological gene-environment** interaction occurs when one or more genetic and one or more environmental factors participate in the same causal mechanism in the same individual (Yang and Khoury 1997; Rothman et al. 2008)
- As with gene-gene interactions, a **statistical gene-environment** interaction does not imply any inference about a specific biological mode of action. It is based on modeling a sample of individuals.

Formal definition of epistasis

- In practice, when modeling or testing, it may only be possible to detect **effect modification** from real-life data and not **interaction**, or interaction but not effect modification.
- Whereas an interaction effect for "exposures" X_1 and X_2 relies on a symmetric role for both X_1 and X_2 ,
  an effect modification relies on a conditioning argument (for instance on X_2) (VanderWeele 2009a)
- The distinction between both effect types is often concealed in regression analysis ... (Robins et al. 2000; North et al. 2005)

Comparison between gene–gene and gene–environment issues

- Conceptually many similar issues in terms of definition and mathematical modelling.
- In practice, some clear differences emerge.
- For G x E:
  - We generally have to decide which environments to measure / test; these are typically only a few (often < 100)
  - Measurement error (lifestyle) and unknown confounding
  - Risk estimation, important for screening strategies and public health interventions

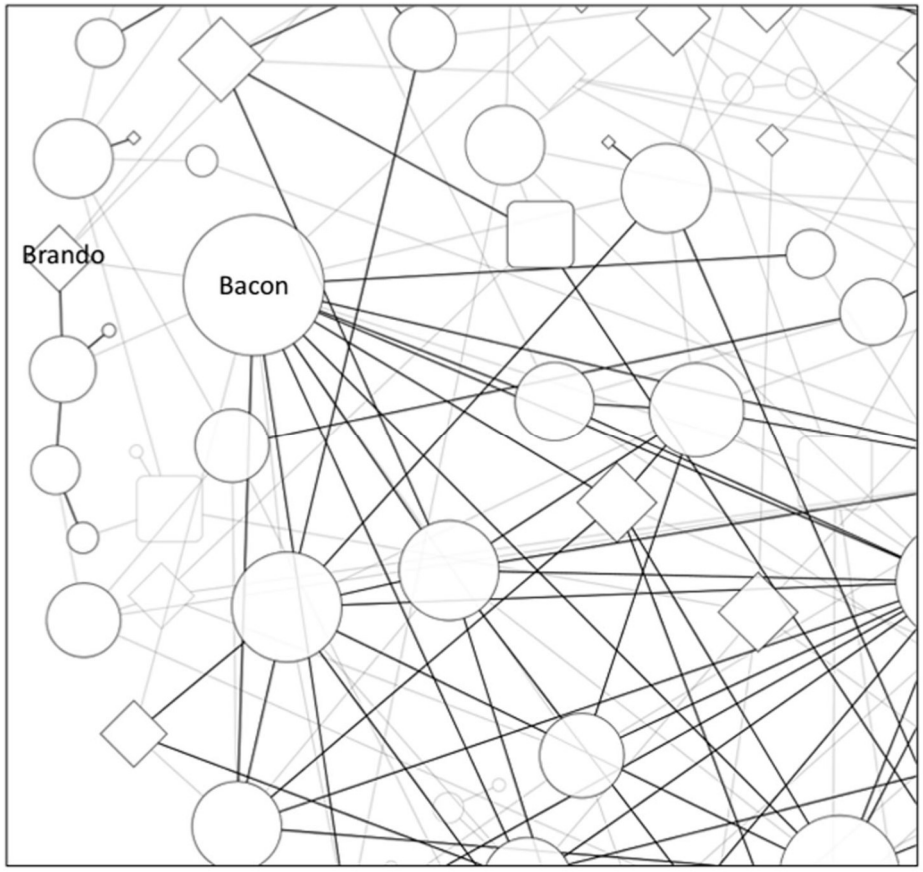
(courtesy slide EUPancreas WG2 Training School, Antwerp, 2016)

Comparison between gene–gene and gene–environment issues

- For G x G
  - Assuming we have GWAS data, we have already measured the genetic factors of interest
  - Adequate error rates (except for newer sequencing technologies)
  - (Hundred) thousands of variants
  - Higher-order interactions may reflect the complex biological wiring of complex diseases (whereas G x E often restricts attention to pairwise interactions)

(courtesy slide EUPancreas WG2 Training School, Antwerp, 2016)

Looking for higher-order interactions



Edges represent small gene–gene interactions between SNPs.

Gray nodes and edges have weaker interactions.
Circle nodes represent SNPs that do not have a significant main effect.
The diamond nodes represent significant main effect association.

The size of the node is proportional to the number of connections.

(McKinney et al 2012)

Some references

Published in final edited form as:
Hum Genet. 2012 October ; 131(10): 1591–1613. doi:10.1007/s00439-012-1192-0.

Challenges and Opportunities in Genome-Wide Environmental Interaction (GWEI) studies

Hugues Aschard[1], **Sharon Lutz**[2,*], **Bärbel Maus**[3,4,*], **Eric J. Duell**[5], **Tasha Fingerlin**[2], **Nilanjan Chatterjee**[6], **Peter Kraft**[1,7], and **Kristel Van Steen**[3,4]

Hum Genet (2014) 133:1343–1358
DOI 10.1007/s00439-014-1480-y

REVIEW PAPER

Practical aspects of genome-wide association interaction analysis

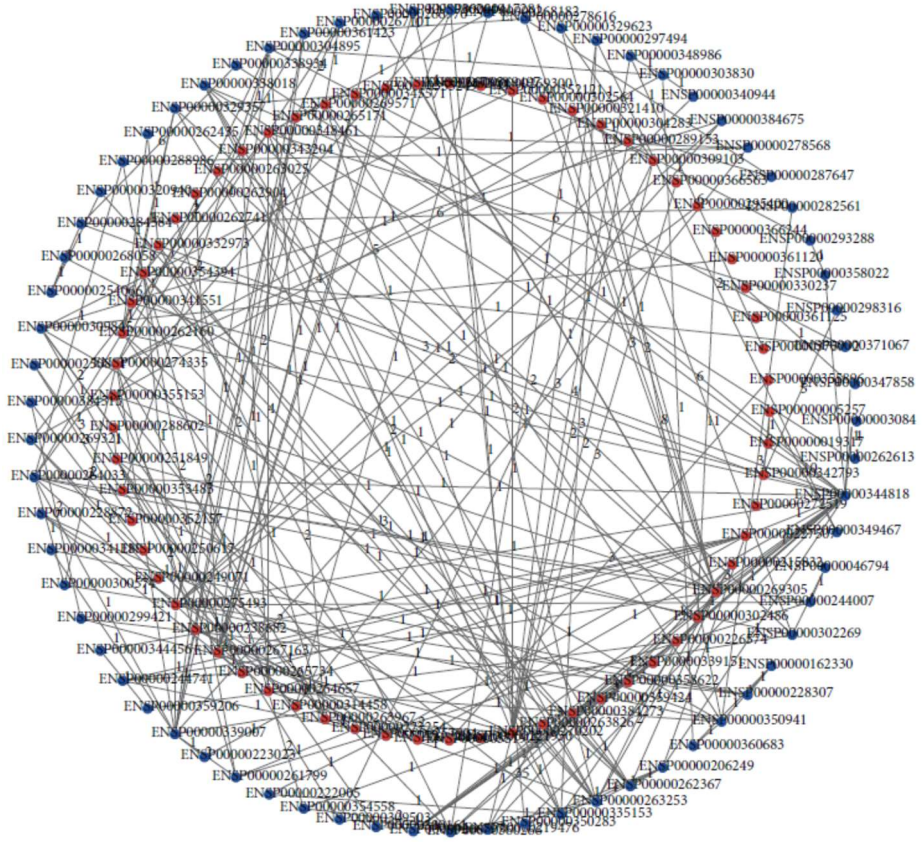
Elena S. Gusareva · Kristel Van Steen

Protein-protein interactions

(Yuan et al. 2015)

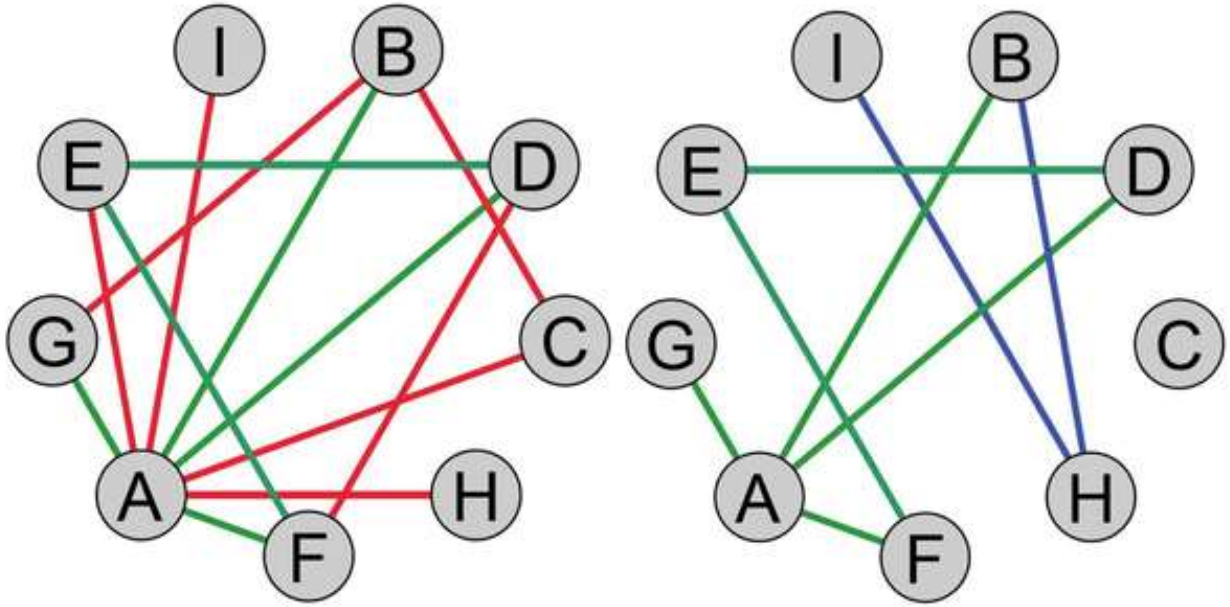
A graph consisting of 2,080 shortest paths:

- The nodes on the inner circle (red nodes) represent 65 PC-related genes.
- The nodes on the outer circle (blue nodes) represent 69 shortest path genes.
- The numbers on the edges represent the weights of the edges.



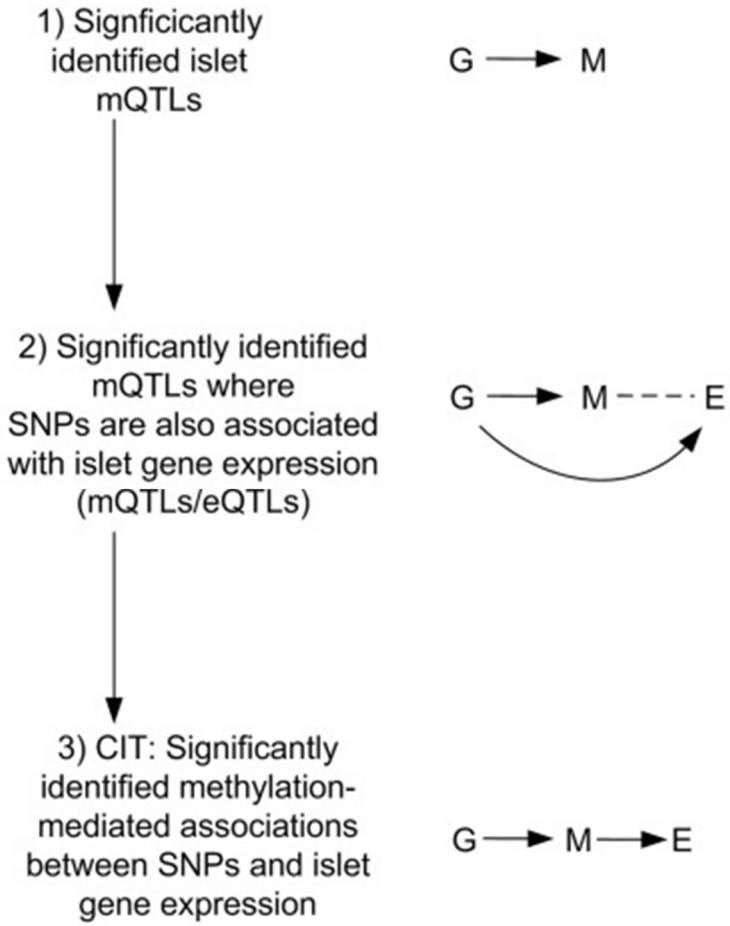
Gene-coexpression networks

(Anglani et al. 2014)



- Healthy condition on the left and disease-affected tissue on the right. Green links remain unchanged in the two phenotypes
- Red connections are loss from healthy to cancer network
- Blue edges are novel connections in the cancer tissue

Genetic-epigenetic mechanistic interactions (pancreatic islets)



cis-mQTLs:	*trans-mQTLs*:
67,438 SNP-CpG pairs (36,783 unique SNPs and 11,735 unique CpGs)	2,562 SNP-CpG pairs (1,465 unique SNPs and 383 unique CpGs)
cis-mQTLs/eQTLs:	*trans-mQTLs/eQTLs*:
2,805 SNP-CpG-mRNA combinations (243 unique SNPs, 276 unique CpGs and 46 unique mRNA transcripts)	37 SNP-CpG-mRNA combinations (22 unique SNPs, 12 unique CpGs and 8 unique mRNA transcripts)
cis-mQTLs/eQTLs:	*trans-mQTLs/eQTLs*:
28 SNP-CpG-mRNA combinations (17 unique SNPs, 14 unique CpGs and 5 unique mRNA transcripts)	4 SNP-CpG-mRNA combinations (2 unique SNPs, 4 unique CpGs and 1 unique mRNA transcript)

Gene-gene interactions using SNPs?

(Olsson et al. 2014)

GWAS Catalogue – "Pancreas Cancer"

Wolpin BM (PMID: 25086665)	2014-08-03	Nat Genet	Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer.	Pancreatic cancer	10

Initial sample description	1,582 European ancestry cases, 5,203 European ancestry controls
Initial ancestry (country of recruitment)	6785 European (U.S., Australia, France, Germany, Netherlands, Denmark, Finland, Norway, Sweden, U.K., Greece, Italy, Spain)
Replication sample description	6,101 European ancestry cases, 9,194 European ancestry controls
Replication ancestry (country of recruitment)	15295 European (Canada, U.S., France, Germany, Netherlands, Denmark, Finland, Norway, Sweden, U.K., Greece, Italy, Spain)
Platform [SNPs passing QC]	Illumina [608202]

(http://www.ebi.ac.uk/gwas/search?query=pancreas%20cancer#study)

López de Maturana et al. Genome Medicine (2021) 13:15
https://doi.org/10.1186/s13073-020-00816-4

Genome Medicine

RESEARCH Open Access

A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer

Evangelina López de Maturana[1,2†], Juan Antonio Rodríguez[3†], Lola Alonso[1,2], Oscar Lao[3], Esther Molina-Montes[1,2], Isabel Adoración Martín-Antoniano[1,2], Paulina Gómez-Rubio[1,2], Rita Lawlor[4], Alfredo Carrato[2,5], Manuel Hidalgo[6,7], Mar Iglesias[7,8], Xavier Molero[9,10], Matthias Löhr[11], Christopher Michalski[12,13], José Perea[14], Michael O'Rorke[15,16], Victor Manuel Barberà[17], Adonina Tardón[18,19], Antoni Farré[20], Luis Muñoz-Bellvís[2,21], Tanja Crnogorac-Jurcevic[22], Enrique Dominguez-Muñoz[23], Thomas Gress[24], William Greenhalf[25], Linda Sharp[26,27], Luis Arnes[28,29,30], Lluis Cecchini[2,8], Joaquim Balsells[9,10], Eithne Costello[25], Lucas Ilzarbe[2,8], Jörg Kleeff[2,13], Bo Kong[12], Mirari Márquez[1,2], Josefina Mora[20], Damian O'Driscoll[26], Aldo Scarpa[4], Weimin Ye[31], Jingru Yu[31], PanGenEU Investigators, Montserrat García-Closas[32], Manolis Kogevinas[1933], Nathaniel Rothman[32], Debra T Silverman[32], SBC/EPICURO Investigators, Demetrius Albanes[32], Alan A Arslan[34,35,36], Laura Beane-Freeman[32], Paige M Bracci[37], Paul Brennan[38], Bas Bueno-de-Mesquita[39], Julie Buring[40], Federico Canzian[41], Margaret Du[42], Steve Gallinger[43], J Michael Gaziano[44], Phyllis J Goodman[45], Marc Gunter[38], Loic LeMarchand[46], Donghui Li[47], Rachael E Neale[48], Ulrika Peters[49], Gloria M Petersen[50], Harvey A Risch[51], María José Sánchez[52,53,54,55], Xiao-Ou Shu[56], Mark D Thornquist[49], Kala Visvanathan[50], Wei Zheng[56], Stephen J Chanock[32], Douglas Easton[57], Brian M Wolpin[58], Rachael Z Stolzenberg-Solomon[32], Alison P Klein[59], Laufey T Amundadottir[60], Marc A Marti-Renom[6,7*], Francisco X Real[2,6,7,61] and Núria Malats[1,2*]

Abstract

Background: Pancreatic cancer (PC) is a complex disease in which both non-genetic and genetic factors interplay. To date, 40 GWAS hits have been associated with PC risk in individuals of European descent, explaining 4.1% of the phenotypic variance.

(Continued on next page)

Phenotype refinement

REVIEWS

Molecular subtypes of pancreatic cancer

*Eric A. Collisson[1], Peter Bailey[2], David K. Chang[2,3] and Andrew V. Biankin[2,3,4]**

Abstract | Cancers that appear morphologically similar often have dramatically different clinical features, respond variably to therapy and have a range of outcomes. Compelling evidence now demonstrates that differences in the molecular pathology of otherwise indistinguishable cancers substantially impact the clinical characteristics of the disease. Molecular subtypes now guide preclinical and clinical therapeutic development and treatment in many cancer types. The ability to predict optimal therapeutic strategies ahead of treatment improves overall patient outcomes, minimizing treatment-related morbidity and cost. Although clinical decision making based on histopathological criteria underpinned by robust data is well established in many cancer types, subtypes of pancreatic cancer do not currently inform treatment decisions. However, accumulating molecular data are defining subgroups in pancreatic cancer with distinct biology and potential subtype-specific therapeutic vulnerabilities, providing the opportunity to define a de novo clinically applicable molecular taxonomy. This Review summarizes current knowledge concerning the molecular subtyping of pancreatic cancer and explores future strategies for using a molecular taxonomy to guide therapeutic development and ultimately routine therapy with the overall goal of improving outcomes for this disease.

(Collisson et al. 2019)