Complexity and Reductionism in the Omics Era

Kristel Van Steen, PhD

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

- Terminology
- Systems Thinking in Analytics
- Systems Thinking in Translational Research
- Take-home Messages
Reductionism and Holism
**Reductionism**

**A definition**

- *Epistemological reductionism*: 
  
  “Knowledge about one scientific domain can be reduced to another body of scientific knowledge”

  Explain all biology in terms of physics and chemistry? (F Crick, 1966)

- *Ontological reductionism*:
  
  “Each particular biological system is constituted by nothing but molecules and their interactions”

  Physical matter is the only reality in nature – metaphysics? (Plato, Aristotle)
Reductionism

A definition

- *Methodological (analytical) reductionism:*
  “Complex systems or phenomena can be understood by the analysis of their simpler components”

  Divide each difficulty into as many parts as is feasible and necessary to resolve it  (Descartes 1637)
Holistic thinking

A definition

“Systems and their properties should be viewed as wholes, not as collections of parts”

- When a system is broken down in structural parts, we are breaking dynamical relationships as well
- A (subjective) choice of boundary needs to be selected to determine the “whole” one decides to be holistic about ...
Omics as a system

• Large-scale data derived from high-throughput technologies, describing cell biomolecules (Joyce and Palsson 2006)
• Reductionist approach: Multiple factor analysis (Escofier and Pagès 1988)
• Holistic approach: A working brain model
• Challenges include:
  – replication
  – validation
  – interpretation
  – obtaining clinically relevant and impactful results
“Environment” as a system

- Proximate environmental exposures (e.g., chemicals, physical agents, and microbiological pathogens). Distal exposures (e.g., social conditions, climate change, broad-scale environmental changes)

- Micro/Macro and Meso-approaches

- Environmental epidemiology (Pekkanen and Pearce 2001) and propensity scores (Westreich et al. 2011)

- Challenges include
  - confounding,
  - coarsening (Heitjan 1993),
  - dynamics
Micro/Meso/Macro

- High-dimensional mixed graphical modeling and structure learning (Cheng et al. 2013);
- Mathematical modeling (Stepanov et al. 1996)
- Multi-level statistical modeling (Duell 2006)
- Agent-based (computational) modeling (Marshall and Galea 2015)
Fact

- Virtually all human common diseases result from the interaction of genetic susceptibility factors and (modifiable) environmental factors

(Furrow et al. 2011)

Environment-Sensitive Epigenetics and the Heritability of Complex Diseases

Robert E. Furrow,*1 Freddy B. Christiansen,† and Marcus W. Feldman*

*Department of Biology, Stanford University, Stanford, California 94305, and †Department of Bioscience and Bioinformatics Research Center, University of Aarhus, DK-8000 Aarhus C, Denmark
(Van Steen and Malats 2015)

WP1: Omics

- **Data Selection**
  - Understanding the biological problem
  - Know your data
  - Preprocess: cleanse, prepare for step 4
  - Integrate: select analytic tool, validate, replicate, report
  - Interpret: reconcile statistics with biology, visualize

WP2: Environment

- Virome
- Microbiome
- Clinical history
- Lifestyle

WP3: Environmental omics

- Genomics
- Transcriptomics
- Proteomics

(Van Steen and Malats 2015)
“Environmental Omics”

- Genetics loads the gun, environment pulls the trigger? (Olden and White 2005)
- Complex routes: environment may "directly" affect intermediate phenotypes (e.g., proteome: drugs targeting specific proteomic complexes)
- Redundancy versus informativity
- Challenges include
  - dealing with heterogeneity (~ increased sample size, number of measurements)
  - small stratified medicine
So we have the **motive** and the **opportunity** for ...
“Systems genetics is the branch of systems biology referring to the integration of omics scale measurements, from genome to metabolome and to functome; the latter through transcriptome and proteome data”

(Kadarmideen et al. 2006)
Biological challenge: omics data are related

“As of 2006 there were 1,062 papers explicitly mentioning "data integration" in their abstract or title, whereas this number has more than doubled in 2013 (2,365).”

(Gomez-Cabrero et al. 2014)

Chapter 13: Perspectives on Data Integration in Human Complex Disease Analysis

(Van Steen and Malats 2015)
Medical challenge: complex “complex diseases”

WG2: « integration of omics data »
(work group leader: K Van Steen)

http://eupancreas.com
What’s in a name?

“Data integration is the process of combining data within a generic framework that encompasses organizing principles for the interaction of different types of systems.”

(Van Steen et al. 2015)

• This definition does not explicitly refer to statistical, bioinformatics or computational tools but to any approach that fits within a transdisciplinary viewpoint.
• It includes data fusion as well as more fancy and more elaborate forms of combining evidences from different data sets or sources.
Analytic Systemics
## Data integration: Analytic toolbox

<table>
<thead>
<tr>
<th>Component-Based Path-Modeling</th>
<th>Diffusion Kernels on Graphs</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Diagram" /></td>
<td><img src="image2.png" alt="Diagram" /></td>
</tr>
<tr>
<td>(Vinzi et al. 2014)</td>
<td>(Kondor and Lafferty 2002)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multi-Layer Networks</th>
<th>SNF</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.png" alt="Diagram" /></td>
<td><img src="image4.png" alt="Diagram" /></td>
</tr>
<tr>
<td>(Kavelä et al. 2013)</td>
<td>(Wang et al. 2014)</td>
</tr>
</tbody>
</table>
Integrative omics analysis strategies
Systemic thinking = out-of-the-box thinking
**MB-MDR (SNPxSNP)**

**Step 1:** Data Organization

**Step 2:** Labeling and Dimensionality Reduction; test value $W_{\text{max}}(i,j)$

**Step 3:** Permutation-based significance assessment; Model selection

For all $i, j$
MB-MDR (SNP×SNP) → extended MB-MDR


[aggregating based on similarity measures to deal with DNA-seq data]
The extended MB-MDR framework (Fouladi et al. 2015 – DNA-seq)

- **Phase 1**: Select sets of interest (ROI) / Prepare the data

- **Phase 2**: Clustering individuals according to features (e.g., common and rare variants, epigenetic markers, ... and kernel methodology)

- **Phase 3**: Application of classic MB-MDR on new constructs
The extended MB-MDR framework

Diffusion kernel PCA
to perform omics integrated gene-based sample clustering

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

(Fouladi et al. 2015, 2016+)
The reward ... multiple application areas

- Gene-based association analysis
  (~GWiS - Huang et al 2011)
- Gene-gene statistical interactions
  (~ GGG – Ma et al. 2013)
- Gene-gene statistical interaction networks
  (~ correlation-based networks/differential network analysis, machine learning based or “forest”-based network construction)
- Integrating different types of omics data
  (genetic + epigenetic variants)
Translational Systemics
Personalized Medicine

A definition

“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 personalized medicine

Personalized Medicine

Learn by recognizing relevant patterns

Public Health Environmental omics

Redefine patient state

Bioinformatics-driven disease management
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“?

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

Maus et al. 2013

Disease heterogeneity - Disease subtypes
What does it mean to be “Diseased”?

Highlighting nonlinear patterns in population genetics datasets

Gregorio Alanis-Lobato1,2*, Carlo Vittorio Cannistraci3*, Anders Eriksson1,4, Andrea Manica4 & Timothy Ravasi1,2

Received 30 September 2014
Accepted 8 January 2015

1Integrative 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, 2Division of Medical Genetics, Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093 USA, 3Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Technische Universität Dresden, Tatzberg 47/49, 01307 Dresden, Germany, 4Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, England.

(Alanis-Lobato et al. 2015)
Fine structure

Combine with EM clustering

(Chaichoompu – ongoing)

IP2CAPS

Genotype data

LD block analysis by PLINK

Haplotype inference by BEAGLE

LD based haplotype encoding

Extracting SNPs outside LD blocks

continue on clustering step
Systemic thinking in personalized medicine

Personalized Medicine

Learn by recognizing relevant patterns

Bioinformatics-driven disease management

Redefine patient state

Public Health Environmental omics
Take-home Messages
Systemic thinking

“Do not accept that either a holistic or reductionist view can be taken. Combining the strengths of both is the way to go!”
Out-of-the-box thinking – data integration

- A series of challenges will need to be overcome:
  - protocol development for standardizing data generation and pre-processing or cleansing in integrative analysis contexts,
  - development of computationally efficient analytic tools to extract knowledge from dissimilar data types to answer particular research questions,
  - the establishment of validation and replication procedures, and tools to visualize results.

- Toy examples on smaller systems can be instrumental in understanding what matters in the context of a complex “integromics” world, but will not be sufficient ...
Out-of-the-box thinking – personalized medicine

“It’s far more important to know what person the disease has than what disease the person has.”

(Hippocrates, 460-370 BC)
Acknowledgements