

1

Introduction to computational and systems biology

ION PETRE
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF TURKU
FALL 2019

Introduction to computational and systems biology 30.10.2019

2

Goal

- ▶ To introduce a number of topics at the border between Biology, Mathematics and Computer Science
 - ▶ explain the rationale and the benefits of the interactions
 - ▶ introduce a number of challenges and success stories in the area
- ▶ Run a student-driven, medicine-focused systems biology project
 - ▶ Start with a medical problem
 - ▶ Build a network model
 - ▶ Visualize, analyze the model
 - ▶ Interpret the conclusions
- ▶ Introductory course
 - ▶ does not go into deep details in any of the topics
 - ▶ advanced courses on some of the topics in the course exist
 - ▶ algorithms in computational biology
 - ▶ Modeling
 - ▶ bioinformatics
 - ▶ molecular computing

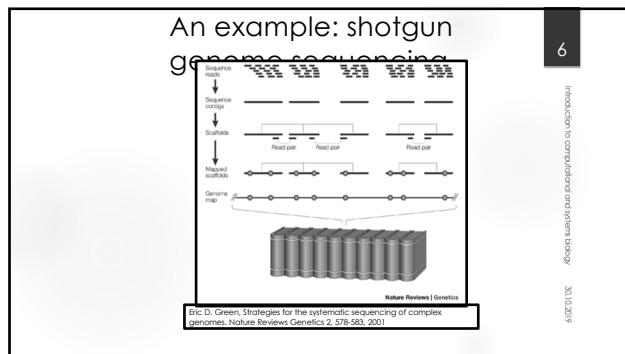
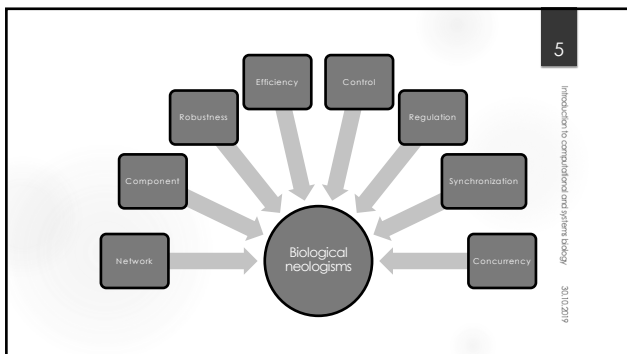
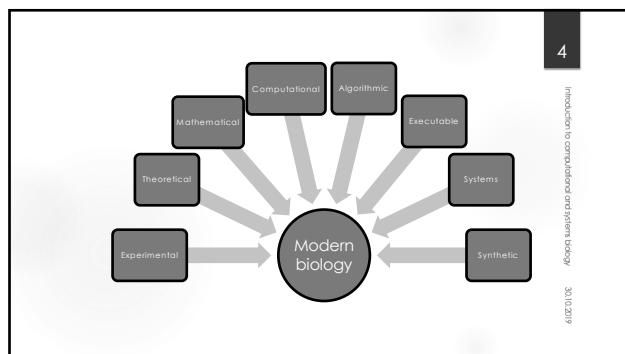
Introduction to computational and systems biology 30.10.2019

3

Biology needs computing

- ▶ Need for a formal language
 - ▶ needed for translating biological problems to computational ones
 - ▶ needed for reasoning about interactions, global effects
 - ▶ needed for representing knowledge, hypothesis
 - ▶ needed in the transition towards a quantitative science
 - ▶ needed as a communication standard
- ▶ Big data
 - ▶ Huge amounts of data, multiple sources, noise
- ▶ Computing needed in many different forms
 - ▶ data storage, retrieval, representation, analysis
 - ▶ algorithmics
 - ▶ modeling
 - ▶ abstract reasoning
- ▶ The view of life as computing
 - ▶ leads to biology as a science focused on reverse-engineering life
 - ▶ life processes as information processing devices: inputs, output, internal states, communication, concurrency, efficiency, robustness, etc.

Introduction to computational and systems biology 30.10.2019



Another example: whole

Theory

A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jacobson R, Kim H, Jorgensen C, Song H, Davis M, Mariani M, Sankaranarayanan S, Jorgensen M, Bergmann R, et al. 2010. Nature Reviews Genetics 11: 119-131.

- Whole cell model of *Mycoplasma genitalium* (human urogenital parasite) whose genome contains 525 genes.
- 28 submodels integrated into a "super-model"; several math formalisms
- Biggest computational challenges:
 - Model fitting: based on 900 publications, 1900 parameter measurements
 - Integrate the 28 submodels into a unified model

Introduction to computational and systems biology 30.10.2019

High interest in building complex quantitative models in neuroscience

- ▶ The **Human Brain Project** (2013-2023) is a Future and Emerging Technologies (FET) Flagship project of the European Commission
 - ▶ Aim: build and simulate complete model of the human brain to better understand its functions
 - ▶ Total budget: 1.2 billion euros
- ▶ The **BRAIN Activity Map Project** (2013-2023) is a US initiative
 - ▶ Aim: map the activity of every neuron in the human brain
 - ▶ Seen as the next high-impact project after the human genome project
 - ▶ Total budget: at least 3 billion dollars

Introduction to computational and systems biology 30.10.2019

Another example: synthetic biology

How scientists created the first artificial life

1. Decode DNA from a bacterium (single-celled organism), in this case *Mycoplasma mycoides*
2. Synthetically create the DNA of the bacterium in the lab and add a "watermark" to distinguish it from real DNA
3. Transplant the artificial DNA into a living bacterium (in this case *Mycoplasma capricolum*) with its own authentic DNA
4. Allow the bacterium, which now contains artificial and authentic DNA, to divide and create "daughter" bacteria, some of which contain artificial DNA and others that contain authentic DNA
5. Add an antibiotic that kills the bacteria with authentic DNA but not the bacteria with artificial DNA
6. Allow the artificial bacteria to produce proteins

RESULT: The artificial DNA produce proteins from the original bacterium, the *Mycoplasma mycoides*, qualifying as the world's first artificial cell.

Source: <http://www.mbi.yale.edu/>

Introduction to computational and systems biology 30.10.2019

▶ The following few slides are from Avi Ma'ayan – Introduction to network analysis in systems biology (Science Signalling 2011)

- ▶ <http://stke.sciencemag.org/content/4/190/tr5.full>
- ▶ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3184357/>

Introduction to computational and systems biology 30.10.2019

Network Analysis

Lecture 1 Representation of biological systems as networks

in Systems Biology

Avi Ma'ayan, Ph.D.

Department of Pharmacology and Systems Therapeutics
Systems Biology Center New York (SBCNY)
Mount Sinai School of Medicine
New York, NY

Introduction to computational and systems biology 30.10.2019

Two Fundamental Ways to Abstract Biochemical Reactions

A

S → P

The function of protein A is its action on S to form P

B C

X A Y

V Z

The function of A is the context of its interactions with other proteins in the cell

Figure 2 The evolution of the meaning of protein function. The traditional view is illustrated on the left, and the post-genomic view on the right.

Eisenberg et al. Nature 405:823 (2000)

Introduction to computational and systems biology 30.10.2019

Different Levels of System Representation

13

Introduction to computational and systems biology 30.10.2019

- A- gene ontology
- B- protein-protein interactions (undirected graphs)
- C- signaling network diagrams (mixed graphs, directed/undirected)
- D- ODE modeling of signaling pathways (directed and weighted)
- E- PDE modeling of signaling pathways considering space (directed, weighted and nodes can move or be at different compartments)

Ma'ayan et al. *Annu Rev Biophys Biomol Struct.* 34:319-349 (2005)

Graph Theory - Basic Concepts

14

Introduction to computational and systems biology 30.10.2019

G – graph
 V – vertices/nodes
 E – edges/links
 A – arcs/directed edges/arrows

Labelled graph	Adjacency matrix
	$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$

Planar Graphs: when there are no edge crossing

Bipartite Graphs: two sets of nodes; links only between members of each set

<http://en.wikipedia.org>

Metabolic Networks

15

Introduction to computational and systems biology 30.10.2019

Bourqui et al. *BMC Systems Biology* 1:29 (2007)

- Two types of nodes: enzymes and substrates
- Reactions can be directional or bidirectional
- Bipartite graph, reactions are not connected and substrates are not connected

Berg et al. *Biochemistry*
 New York: W. H. Freeman and Co., c2002

Cell Signaling Pathways

16

Introduction to computational and systems biology 30.10.2019

Nodes are proteins, metabolites, lipids, second messengers, or peptides

Interactions designate information flow, can be activation or inhibition, and are direct and physical

Ma'ayan A, et al. *Sci Signal.* 2:cm1 (2009)

Cell Signaling Networks

17

Introduction to computational and systems biology 30.10.2019

Signaling pathways are not isolated and can be merged into large networks

Ma'ayan et al. *Science* 310, 1078 (2005)

Indirect Signaling Interactions from Literature

18

Introduction to computational and systems biology 30.10.2019

Pseudo-nodes are used as place holders to fill-in unknown links and components

Li et al. *PLoS Biol.* 4:e312 (2006)

Kinase-Substrate Network

19

Protein kinase-substrate networks are directed bipartite graphs that connect kinases to their substrates through protein phosphorylation

Tan et al. *Sci Signal.* 2009 Jul 28;2(81):ra39

Introduction to computational and systems biology 30.10.2019

Example of Gene Regulation Networks

20

Stem cell differentiation regulation

- Nodes are genes and transcription factors
- Interactions can be directional or bidirectional
- Interactions can be activation or inhibition

MacArthur et al., *PLoS ONE* 3: e3086 (2008)

Introduction to computational and systems biology 30.10.2019

Another Example of a Gene Regulation Network

21

Drosophila Segment Polarity Expression Pattern

- Nodes are genes, transcription factors or signaling components
- Interactions are directional and can be activation or inhibition

Albert R, Othmer HG. *J Theor Biol.* 2003 223(1):1-18.

Introduction to computational and systems biology 30.10.2019

Disease Gene Networks

22

Human Disease Network

Goh et al. *Proc Natl Acad Sci USA.* (2007) 104:8685-90

Each node corresponds to a distinct disorder, colored based on the disorder class. The size of each node is proportional to the number of genes in the corresponding disorder, and the link thickness is proportional to the number of genes shared by the disorders connected by the link.

Introduction to computational and systems biology 30.10.2019

Drug-Target Networks

23

Drugs can be connected to their known protein targets

Ma'ayan et al. *Mt Sinai J Med* (2007) 74:27

Yildirim et al. *Nat Biotechnol.* (2007) 25:1110

Introduction to computational and systems biology 30.10.2019

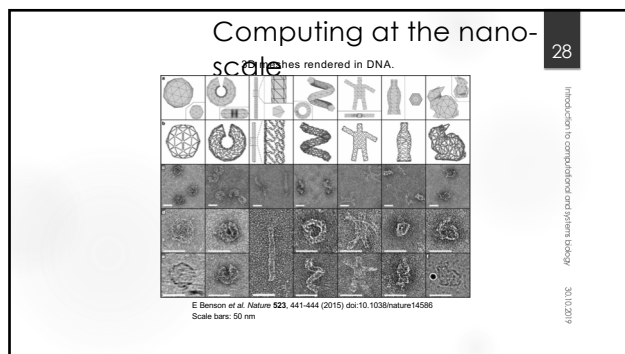
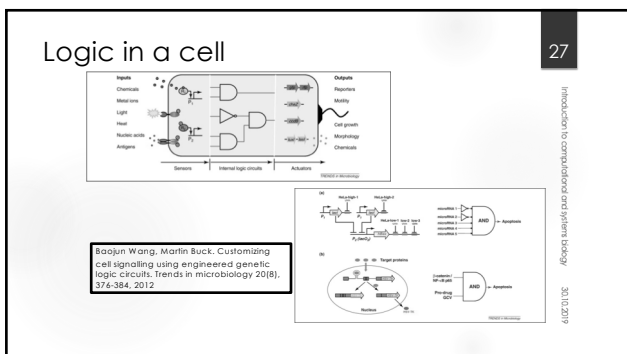
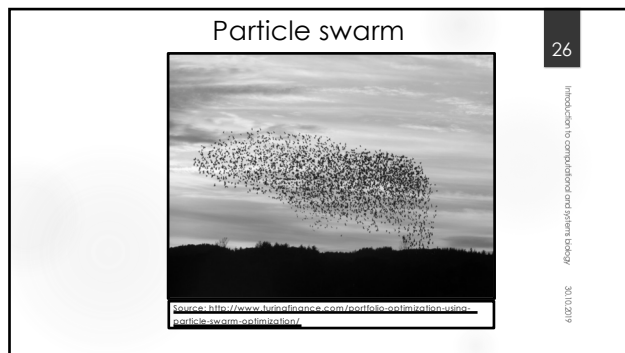
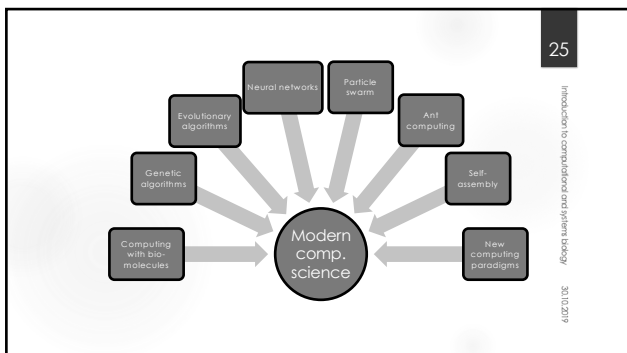
Math and Computer science need biology

24

- ▶ Less talked about than the other direction of the interaction
- ▶ The view of life as computing
 - ▶ new definitions of what computing is
- ▶ Many of the approaches in AI have a conceptual origin in biology
- ▶ Parallelism and concurrency on a massive-scale
 - ▶ A mainstream area in CS is the study of parallelism and concurrency
- ▶ Biology is a rich reservoir of parallelism and concurrency problems, on a massive scale, much larger than that seen in engineering
- ▶ Tools developed in CS need to be refined, new ones needed

- ▶ Biology as hardware for computing
 - ▶ Computing with molecules
 - ▶ Nano-science and -engineering
 - ▶ Self-assembly
 - ▶ Functional materials
 - ▶ Personalized medicine
 - ▶ Faster computers

Introduction to computational and systems biology 30.10.2019



- 29
- Introduction to computational and systems biology 30.10.2019
- ### Content (tentative list)
- ▶ Elements of molecular biology
 - ▶ Elements of biotechnology
 - ▶ Molecular programming
 - ▶ Computational self-assembly
 - ▶ Bioinformatics tools and databases
 - ▶ Biological networks
 - ▶ Algorithms in computational biology
 - ▶ Gene sequencing
 - ▶ Sequence alignment
 - ▶ Gene prediction
 - ▶ Genome rearrangements
 - ▶ Protein sequencing