



King's Research Portal

DOI:

[10.1016/j.coisb.2017.05.001](https://doi.org/10.1016/j.coisb.2017.05.001)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Gomez-Cabrero, D., & Tegnér, J. (Accepted/In press). Iterative Systems Biology for Medicine – time for advancing from network signature to mechanistic equations. *Current Opinion in Systems Biology*.
<https://doi.org/10.1016/j.coisb.2017.05.001>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

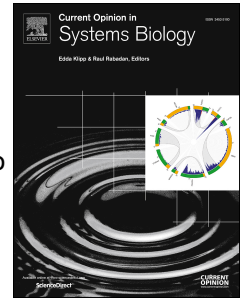
Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Accepted Manuscript

Iterative Systems Biology for Medicine – time for advancing from network signature to mechanistic equations

David Gomez-Cabrero, Jesper Tegnér



PII: S2452-3100(17)30005-7

DOI: [10.1016/j.coisb.2017.05.001](https://doi.org/10.1016/j.coisb.2017.05.001)

Reference: COISB 60

To appear in: *Current Opinion in Systems Biology*

Received Date: 19 April 2017

Accepted Date: 3 May 2017

Please cite this article as: Gomez-Cabrero D, Tegnér J, Iterative Systems Biology for Medicine – time for advancing from network signature to mechanistic equations, *Current Opinion in Systems Biology* (2017), doi: 10.1016/j.coisb.2017.05.001.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Iterative Systems Biology for Medicine – time for advancing from network signature to mechanistic equations

David Gomez-Cabrero^{1,2,3,4,5,6,8}, Jesper Tegnér^{1,2,3,4,7,8}

¹ Unit of Computational Medicine, Department of Medicine, Karolinska Institutet, Stockholm, 171 77, Sweden

² Center for Molecular Medicine, Karolinska Institutet, Stockholm, 171 77, Sweden

³ Unit of Clinical Epidemiology, Department of Medicine, Karolinska University Hospital, Stockholm, L8, 17176, Sweden

⁴ Science for Life Laboratory, Solna, 17121, Sweden

⁵ Mucosal and Salivary Biology Division, King's College London Dental Institute, London, SE1 9RT, United Kingdom.

⁶ Unit of Bioinformatics, NavarraBiomed, Departamento de Salud-Universidad Pública de Navarra, Pamplona-31008, Navarra, Spain

⁷ Biological and Environmental Sciences and Engineering Division, & Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955–6900, Kingdom of Saudi Arabia

⁸ Authors contributed equally.

Abstract:

The rise and growth of Systems Biology following the sequencing of the human genome has been astounding. Early on, an iterative wet-dry methodology was formulated which turned out as a successful approach in deciphering biological complexity. Such type of analysis effectively identified and associated molecular network signatures operative in biological processes across different systems. Yet, it has proven difficult to distinguish between causes and consequences, thus making it challenging to attack medical questions where we require precise causative drug targets and disease mechanisms beyond a web of associated markers. Here we review principal advances with regard to identification of structure, dynamics, control, and design of biological systems, following the structure in the visionary review from 2002 by Dr. Kitano. Yet, here we find that the underlying challenge of finding the governing mechanistic system equations enabling precision medicine remains open thus rendering clinical translation of systems biology arduous. However, stunning advances in raw computational power, generation of high-precision multi-faceted biological data, combined with powerful algorithms hold promise to set the stage for data-driven identification of equations implicating a fundamental understanding of living systems during health and disease.

In our view, systems biology has now become an accepted paradigm in biological research [1]. This is in part reflected in the sheer number and quality of publications utilizing systems approaches acknowledging and embracing the complexity of biology [2–4]. Recently, the application of such frameworks to clinical challenges has led to the emergence of what could be referred to as systems or network medicine [1,5,6]. It is therefore timely to ask *to what extent real progress has been achieved* - and to critically assess the nature of conceptual and technical hurdles remaining in meeting the needs from a medical standpoint. Here, we use the structure of the very influential position paper (close to 4000 citations) by Kitano in 2002 [7] to assess achievements and challenges on the basis of the research agendas put forward. Next, on the basis this analysis we argue that despite conceptual and technical advances, there remains a fundamental gap between finding associated features (biomarkers) of a given process versus the more challenging task obtaining a causal (e.g. mechanistic) understanding of the process. This, in our view an ultimate gap, becomes even more glaring in a medical context, since there we would like to ask therapeutic questions such as what happens if we do X to a (human) system. At the end of the day, X is an intervention based on causal understanding in the sense that “if X is executed” then “the relevant processes become properly modified”. We conclude this opinion paper with the sentiment that the time is ripe for bridging this gap and algorithmic tools in combination with richer data and more powerful computational platforms have the potential to operationally address the inherent challenges in wordings such as ‘relevant’ and ‘properly’ above.

Systems-based analysis a la Kitano.

Since the sequencing of the human genome, there has been a shift in biomedical research from reductionism towards a holistic view in the sense of acknowledging the complexity and myriad of parallel and interconnected processes, including the multiple spatio-temporal scales involved in almost any biological phenomena. Interestingly, technological advances rather than theory itself have largely driven this shift of perspective. It has generated a multitude of novel methodologies (or creative applications of existing methodologies), many of them labeled under the fields of Systems Biology [7] or Systems Medicine [8,9]. While multiple complementary definitions of Systems Biology do exist [10,11], we frame our discussion using the landmark paper from Prof. Kitano in 2002 [7]. Prof. Kitano provided a comprehensive concept, and what could be referred to as a normative account, in turn translated into an operational *pipeline* defining Systems Biology as a methodology to understand biological systems. Specifically, an iterative standpoint was formulated such that a cycle of research combining dry-lab and wet-lab efforts would generate, validate or reject a hypothesis, and finally incorporated the outcomes of the analysis in the state-of-the-art amenable for a new iteration of the cycle. In this, Prof. Kitano emphasized four necessary vital avenues of investigations that jointly would admit system-level understanding: (1) *system structures* (for instance the network of interactions), (2) *system dynamics* (mathematical description and analysis), (3) *control method* (identification of the biological targets that can modulate or control the state of the cell) and a (4) *design method* (aiming to construct systems de novo to make use or to validate properties identified or hypothesis generated). Remarkably, in hindsight the 2002 Kitano’s vision has turned out to be truly predictive in that we have witnessed remarkable progress in all those four areas, yet at different pace, and in part evolving in separate communities. For example, the emergence of the young dynamic field of synthetic biology can be viewed as response to the need for design, which in turn can be traced back to Feynman’s classic dictum on what you can’t

create you don't understand. At this juncture, we could conceptually ask whether these four areas are necessary, sufficient, or both to achieve systems understanding [12]. To shed light on this issue we will first briefly review progress in respective area, finding that the aforementioned gap between biomarkers and mechanisms cuts across all four areas.

Structure, Dynamics, Control, and Design – progress and gaps.

In engineering, or more specifically control theory, system identification is defined as a method for developing mathematical and computer-based models that represent the characteristics of that system from measurements of the system inputs and outputs [13]. Traditionally, linear systems have been in focus and the mathematical model captures the transfer function between input and output, thus not necessarily incorporating neither the underlying biophysical components nor the non-linear dynamics governing the interactions between the components over time. In contrast, in biology we aspire to identify not only the structure of cellular networks but also their dynamics, in order to achieve engineered control of the system [14,15]. This motivates the division of labor between finding the structure, dynamics, and control respectively as originally conceptualized by Kitano. The *identification of System Structures* can be attained by data-driven reverse-engineering approaches [16], either augmented by prior knowledge as a structural scaffold or by direct experimental analysis requiring structural learning directly from the data [17]. With the advent of high-throughput technologies – including both microarray and Next-Generation Sequencing technologies, reverse-engineering approaches have been a major research area in systems biology since the original 2002 publication. Pure data-driven reverse-engineering methods have as a rule only used time-series and/or perturbation experiments to uncover associations - not necessarily causal - between features e.g a transcription factor and the expression of the corresponding target genes [18]. Such relationships can readily be represented using different modelling formalisms, such as Mutual information [11], Boolean networks, Bayesian networks (BNs) [12], Petri nets [19], constraint-based models, differential equations [20], rule-based models [21], cellular automata or agent-based models [22], all being parts of a growing toolkit for data-driven reverse-engineering approaches. Yet, causal parameterizing remains challenging due to uncertainties in model structure and parameters [23]. A second line of reasoning is to define a prior network structure or scaffold through a literature review. Examples include modeling of atherosclerosis modeling [20], brain functioning [24], or immune system [25] to name a few examples among many. Alternatively, the prior structural template can be collected from systematic experiments, as in the case of Protein-Protein interactions and the generation of the Proteome-Scale Map of the Human Interactome Network [17]. From the three approaches, experimental and data-driven approaches are in our view to become even more prevalent due to the exponential growth of data in public repositories [26–28] and the decrease cost in sequencing [29]. The knowledge-based approach appears to be at a turning point in the sense that “classical” text-mining methodologies [30,31] have not, in our view, provided a significant edge when compared with other approaches, whereas recent advances using DeepLearning [32] methodologies hold promise to disrupt current state-of-the-art in text mining similarly to recent achievements in genomic analysis [33,34]. In summary, these advances in network biology have enriched the notion of biomarker from a single or very few features to include a larger set of (inter-connected) features (i.e. a network signature) associated with disease or a biological process. In contrast, to achieve a mathematical description of a biological system from first or derived principles has been challenging ever since the pioneering work

of Hodgkin-Huxley and still constitutes a fundamental barrier. If we are unable to have what could be referred to as fundamental guiding principles to model dynamic biological processes and how these are controlled (e.g. least action principle in physics) and given that we are dealing with large inter-connected living systems the challenge is how to formulate unbiased mathematical models under these modeling conditions. Regardless of how the interactions between elements are represented, the importance of **system dynamics** remains as illustrated by the central dogma of Systems Biology: “*system dynamics that gives rise to the functioning and function of cells*” [35]. One of the best examples is the pioneering work by Tyson and colleagues on the cell-cycle in different systems [36,37]. For this problem, it is imperative to model and understand the **Systems Dynamics**, i.e. what governs the switch from one phase to another, as it is dependent on the dynamics between many variables, which is different in each of the phases of the cell cycle. A wiring diagram is simply not sufficient. Such a systems dynamics analysis requires: (a) an accurate definition of the system structure [23], and analysis of how variations of the structure may affect the dynamics [38,39], (b) a mathematical parameterization, linear or non-linear [40], (c) a mathematical analysis such as the bifurcation analysis [41] or the identification of slow manifolds [12,42], and (d) sensitivity analysis of structure and parameters in the model using the system dynamics as a readout [43,44]. From experience within the community, we would like to argue that essential factors for success includes dealing with a “small confined” system (e.g. the cell cycle or nerve impulse propagation) combined with physical insights enabling a simplified or reduced (phenomenological) description of the system at hand. Yet, despite this, in our hands any such analysis targeting system dynamics sooner or later is challenged by the conundrum on: “*how to generate hypothesis under uncertainty in both structure and dynamics?*”. This becomes exponentially challenging the larger system is and the less experimental constraints we can in principle impose on the state-variables and parameters governing the system dynamics. In our view, because of these difficulties the issue of uncertainty has mainly been addressed in the area of **Control Method** where ‘what if’ kind of questions are as a rule being asked aiming to uncover what controls, as a rule, a small confined system. Broadly speaking, meeting this challenge, investigators have developed two different approaches: ordinary differential equations (ODE) based dynamical models versus structural network-based models. Using ODE, the challenge of generating hypothesis under uncertainty is an active field of research [45], and methods are emerging for exploring the space of *feasible models* [23,39] explaining existing data, and generating robust hypothesis by their study [20,46,47]. In the case of network-models, the majority of models or wiring diagrams, such as Protein-Protein interactions [17,48], gene-gene interactions or co-morbidity oriented models [49], suffer from the lack of dynamics because the modeling framework is an interactome-based description of the system. However, interactome or structural models are in many cases considered as systems biology models – or as a “*gateway to systems biology*”[50] – because they provide some insights into control aspects by identifying for instance: (a) candidates for master regulators [41] in highly connected genes or genes with a high betweenness centrality value [51] (measures the centrality of a node in a network by its prevalence in the shortest paths between other nodes); (b) identifying groups of highly connected nodes associated to a specific functional role or disease [48]; or, among others, (c) identifying recurrent motifs [52] in the network that may be associated to robustness [7]. Yet, a wiring diagram is not sufficient to address what happens if those nodes or edges are modified, i.e. causal interventions are as a rule beyond the scope of such models. In our experience, and very importantly, both approaches,

ODE and Network-based models, can be complementary. A first-order approach is simply to use network-approaches combined with knowledge-based methods to select the most relevant features (nodes and edges) to be incorporated in a dynamical model [53]. A major bottleneck for such an approach is that, in most cases, the information used for variable selection may not suffice for the generation of accurate mathematical models [54]. We like to argue that a critical but fair assessment, is that while dynamical system analysis and causal control analysis has been ongoing in the modeling community those efforts have not been successfully adopted in systems medicine [55] or impacted the clinical community. This is not surprising since this state-of-affairs reflect the current gap between network biomarkers and a true causal understanding of a system [12].

The fourth area of **Design Method**, pointed out by Kitano, entails the use of combined in-silico and wet-lab strategies to design new systems [56] or to-modify existing systems on-demand [57]. In this regards, Synthetic Biology plays a fundamental role in Systems Biology because it opens the door to investigate basic design principles in biological systems [58,59]. Interestingly, some investigators consider Synthetic biology as a logical end point of Systems Biology [10]. Several milestones have been achieved since 2002, including the generation of a minimal bacterial cell, an 531-kb genome encoding 438 proteins and 35 RNAs [60]. Additionally, several techniques allows several degrees of manipulation, such as silencing RNA [61], shRNA and CRISPR-Cas9 [62,63] perturbation methodologies. Moreover, several advances in synthetic biology are already approaching clinical utility [64]. An impressive line-up of success stories has enabled interventions and reprogramming of systems without requiring complete systems understanding.

Timely to close the gap between network biomarkers and causal mechanisms?

In the previous section, we have briefly enumerated some advances since 2002 in Systems Biology (and Systems Medicine) with special reference to structure, dynamics, control, and design. Yet, there are several practical, technical, and conceptual hurdles impeding smooth transition and iterations of the Kitano cycle. These include for instance: (i) insufficient amount and quality of data supporting not only accurate reconstruction of the network structure of a system but importantly the generation of accurate mechanistic models [16]; (ii) many of the methodologies proposed were prohibitory computationally expensive. Additionally, (iii) the research community has uncovered many regulatory layers, beyond PPI or TF-gene interactions that are required to be included in any faithful biological model. For example, epigenetic regulations [65,66], chromatin accessibility [67] or transcriptional control exerted by miRNAs [68], thus making the problem of formulating a dynamical model even more difficult. Hence, the astute reader may well ask *why should we believe that it is timely to advance beyond network biomarkers to dynamical models capable of addressing causal or therapeutic interventions in “large” systems?*

In our view, we have attained a technological stage where hypothesis driven research in Systems Biology is becoming (or can become) a reality. We now highlight the most relevant elements that could spur such a development:

- **Computational speed:** in the year of 2000, Pentium 4 was the new Intel Processor, and in contrasts, 2015 the latest Intel processor was not only 84x times faster [69] but also significantly cheaper. Furthermore, recent advances in the utilization of GPUs have fueled large-scale parallel computations [70]. These developments clearly support massive ensemble approaches [38] and computational-driven parameter explorations [39] that previously were restricted to supercomputer centers.

Importantly, in addition to the continuous increase in computational power following Moore's law, high-performance computing centers centralized in large institutions and cloud-based systems, such as Amazon Web Services (<https://aws.amazon.com/hpc/>), are complementing and eventually replacing small and medium computational solutions to a large extent. It is clear that research teams have access to affordable powerful computational resources at an unprecedented scale compared to a decade ago. In summary, the sheer size and the number of models that can be explored in any given application is now enhanced by several orders of magnitude.

- Size and variety of data-types: the advent of NGS technologies initially provided the tools for whole genome sequencing [71] and a better estimation of mRNA expression [72] but soon, applications were extended to all DNA or RNA associated regulatory layers such as Transcription Factor Binding or Histone marks profiles by ChIP-seq [73], miRNA sequencing [68], Chromatin Accessibility [74] or DNA Methylation [65] among others. Through the combination of several NGS technologies we have in principle access to a complete regulatory view of a biological system or disease, and existing large-scale collaborative projects such as TCGA, ENCODE, IHEC or STATegra [75] are examples in that direction. Hence, the capacity to capture the relevant correlated features in a given process has improved dramatically.
- Increased biological resolution in data: Microarray and NGS technologies were originally capable of generating molecular profiles by averaging across several millions of cells per sample. For example, the expression profile estimated from millions of cells, masking cell-to-cell variation, has been a serious concern when analyzing heterogeneous tumors. Fortunately, recent advances in single-cell technologies, such as C1 Fluidigm, Drop-seq or 10X Genomics, are completely changing the game [76–79]. Most importantly, single-cell technologies are not exclusively targeting RNA-seq but other omics such as DNA Methylation [80], Chromatin Conformation [81,82], and efforts are currently conducted to profile several omics data-types from the same cell [83]. A second major value of single-cell techniques is that we can profile large number of single cells, within budgetary limits, which in the context of sample-hungry reverse-engineering algorithms is a major advantage. In essence, single cell techniques hold the promise of delivering a fundamental description of cells and their dynamics, thus setting the stage for modeling larger systems as well as accelerating designer or synthetic biology approaches.
- Method development: Yet, despite the technical advances referred to above we would like to emphasize that there is a deep need for developing powerful algorithms and computational modeling techniques in close alignment with the new types of high-precision data while making use of the new computational raw power. We have witnessed such kind of synergistic method evolution during the last two decades in the area of network analysis [84–86]. Furthermore, there has been advances in mechanistic modeling, analysis [45] and visualization [87]. From our point of view, the most relevant and timely challenge is *still* the systematic generation of in-silico derived causal hypothesis in larger systems in the face of uncertainty which is addressed in detail in [45].
- Workforce: Systems Biology clearly requires multi-skilled groups, covering multiple fields from mathematics to biology and even more so when entering the medical

domain. While over the years courses and communities [88,89] responding to these needs have developed, there is still a need for funding soft infrastructures including education to further nurture the advances and communication across knowledge silos to properly integrate and support advanced data-analysis close to biology and medicine.

ACCEPTED MANUSCRIPT

Conclusions:

Systems Biology has evolved over the last two decades to become a practical methodology taking center stage in the analysis of complex biological processes. Much has developed along the lines of Kitano's prophetic vision. Here we have made the point out why the remaining gaps and bottlenecks existing at the beginning of the 21st century are ready to be bridged. The overarching remaining challenge is level the field towards a quantitative and dynamical account of biological processes during health and disease. Dramatic improvements in raw computational power, rich high-precision multi-level data in conjunction with new advanced algorithms hold the promise to bridge the gap between associated biomarkers and mechanistic elucidation of the governing processes. It is therefore timely that systems-based approaches, using a combined and integrative in-silico and experimental approach, address this gap and in doing so setting the stage to advance systems biology into targeting medical challenges in high precision [90,91]. Key here for drug development and personalized medicine is our ability to identify causal mechanisms enabling robust precise therapeutic interventions beyond enumerating single biomarkers or network signatures. Last, but critical, as enablers of such work crossing traditional boundaries is the need for multi-skilled research teams, or collaborations, that make use of all tools and resources [91].

As a conclusion, we are truly excited to be part of this era in the analysis of living systems during health and disease, where systems-based research and techniques can further revolutionize the all-changing biomedical research field, possibly even beyond what Kitano could imagine after the sequencing of the human genome

Captions:

Figure 1. Reviewing Hypothesis-driven research in systems biology. The inner circle reproduces Kitano's classical Iterative System Biology workflow. The outer boxes denote gaps and current advances and tools mitigating those gaps.

Acknowledgement:

The authors would like to acknowledge helpful comments from Drs. Narsis Kiani and Saeed Shoiaie.

Funding: J.T. was supported by a CERIC (Center of Excellence for Research on Inflammation and Cardiovascular disease) grant, Vetenskapsrådet Medicine and Health (Dnr 2011-3264), Torsten Söderberg Foundation, FP7 STATegra, and AFA Insurance and Stockholm County Council.

References:

1. Commission E: *FROM SYSTEMS BIOLOGY TO SYSTEMS MEDICINE*. 2010.
 2. Bassett DS, Sporns O: **Network neuroscience**. *Nat. Neurosci.* 2017, **20**:353.
- * Detail of the emerging trends and state-of-the-art challenges in network analysis in neuroscience by reviewing the novel empirical and theoretical frameworks.
3. Schadt EE, Sachs A, Friend S: **Embracing Complexity, Inching Closer to Reality**. *Sci. Signal.* 2005, **2005**:pe40-pe40.
 4. West M, Ginsburg GS, Huang AT, Nevins JR: **Embracing the complexity of genomic data for personalized medicine**. 2006, doi:10.1101/gr.3851306.
 5. Wang RS, Maron BA, Loscalzo J: **Systems medicine: Evolution of systems biology from bench to bedside**. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2015, **7**:141–161.
 6. Barabási A-L, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease**. *Nat. Rev. Genet.* 2011, **12**:56–68.
- * Authors reviewed the use of networks in disease modeling and its use in the identification of novel disease-associated modules and genes. The manuscript includes the basic concepts of network theory in the context of medicine and existing challenges and opportunities.
7. Kitano H: **Systems biology: A brief overview**. *Sci. (New York, NY)* 2002, **295**:1662–1664.
- ** Visionary overview of Systems Biology, Systems Medicine and the hypothesis driven approach that combines both wet and dry laboratories. The current manuscript uses Kitano's work as a reference where to compare the current state of Systems Biology.
8. Strunz S, Wolkenhauer O, la de Fuente A: **Systems Medicine**. *Methods Mol. Biol.* 2016, **1386**:353–374.
 9. Hood L, Flores MA, Brogaard KR, Price ND: *Systems Medicine and the Emergence of Proactive P4 Medicine : Predictive , Preventive , Personalized and Participatory*. Elsevier Inc.; 2013.
 10. Breitling R: **What is systems biology?** *Front. Physiol.* 2010, **1 MAY**:1–5.
 11. Chuang H-Y, Hofree M, Ideker T: **A decade of systems biology**. *Annu. Rev. Cell Dev. Biol.* 2010, **26**:721–44.
 12. Tegnér J, Zenil H, Kiani N, Ball G, Gomez-Cabrero D: **A Perspective on Bridging Scales and Design of Models using Low-Dimensional Manifolds & Data-driven Model Inference**. *Philos. Trans. A* 2016, **accepted**.
 13. Ljung L: *System Identification*. Wiley Online Library; 1999.
 14. Lazebnik Y: **Can a biologist fix a radio? - Or, what I learned while studying apoptosis**.

- Biochem.* 2004, **69**:1403–1406.
15. Antebi YE, Nandagopal N, Elowitz MB: **An operational view of intercellular signaling pathways.** *Curr. Opin. Syst. Biol.* 2017, **1**:16–24.
 16. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Aderhold A, Allison KR, Bonneau R, et al.: **Wisdom of crowds for robust gene network inference.** *Nat. Methods* 2012, **9**:796–804.
 17. Rolland T, Taşan M, Charlotheaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, et al.: **A Proteome-Scale Map of the Human Interactome Network.** *Cell* 2014, **159**:1212–1226.
 18. Lacasa L, Luque B, Ballesteros F, Luque J, Nuño JC: **From time series to complex networks: the visibility graph.** *Proc. Natl. Acad. Sci. U. S. A.* 2008, **105**:4972–4975.
 19. Koh CH, Nagasaki M, Saito A, Li C, Wong L, Miyano S: **MIRACH: efficient model checker for quantitative biological pathway models.** *Bioinformatics* 2011, **27**:734–5.
 20. Gomez-Cabrero D, Tegner J: **Workflow for biological computational modeling with uncertainty.** *Interface Focus* 2011, [no volume].

* Our initial efforts in generating robust hypothesis when analyzing mathematical descriptions of disease models under uncertainty.

21. Selivanov V a, Votyakova T V, Pivtoraiko VN, Zeak J, Sukhomlin T, Trucco M, Roca J, Cascante M: **Reactive oxygen species production by forward and reverse electron fluxes in the mitochondrial respiratory chain.** *PLoS Comput. Biol.* 2011, **7**:e1001115.
22. Holcombe M, Adra S, Bicaş M, Chin S, Coakley S, Graham AI, Green J, Greenough C, Jackson D, Kiran M, et al.: **Modelling complex biological systems using an agent-based approach.** *Integr. Biol. (Camb).* 2012, **4**:53–64.
23. Cedersund G, Samuelsson O, Ball G, Tegner J, Gomez-Cabrero D: **Optimization in Biology Parameter Estimation and the Associated Optimization Problem.** In *Uncertainty in Biology: A Computational Modeling Approach.* Edited by Geris L, Gomez-Cabrero D. Springer International Publishing; 2015:177–197.
24. Gutkin B, Pinto D, Ermentrout B: **Mathematical neuroscience: from neurons to circuits to systems.** *J. Physiol. Paris* 2003, **97**:209–19.
25. Louzoun Y: **The evolution of mathematical immunology.** *Immunol. Rev.* 2007, **216**:9–20.
26. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res.* 2002, **30**:207–10.
27. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall K a, Phillippy KH, Sherman PM, Holko M, et al.: **NCBI GEO: archive for functional genomics data sets--update.** *Nucleic Acids Res.* 2013, **41**:D991-5.
28. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M, et al.: **ArrayExpress update--trends in database growth and links to data analysis tools.** *Nucleic Acids Res.* 2013, **41**:D987-90.
29. National Human Genome Research Institute: **The Cost of Sequencing a Human Genome [Internet].** 2016, [no volume].
30. Harmston N, Filsell W, Stumpf MPH: **What the papers say: text mining for genomics and systems biology.** *Hum. Genomics* 2010, **5**:17–29.
31. Maier D, Kalus W, Wolff M, Kalko SG, Roca J, Marin de Mas I, Turan N, Cascante M,

- Falciani F, Hernandez M, et al.: **Knowledge Management for Systems Biology; a general and visually driven framework applied to translational medicine.** *BMC Syst. Biol.* 2011, **5**:38.
32. LeCun Y, Bengio Y, Hinton G: **Deep learning.** *Nature* 2015, **521**:436–444.
- ** Deep Learning are the machine learning next-generation algorithms, making use of smart algorithmic developments in neural networks and the recent years increased computational power. Authors provide a well-designed overview of the methodologies associated and examples in image and language processing.
33. Kelley DR, Snoek J, Rinn JL: **Basset: Learning the regulatory code of the accessible genome with deep convolutional neural networks.** *Genome Res.* 2016, **26**:990–999.
34. Min X, Chen N, Chen T, Jiang R: **DeepEnhancer: Predicting enhancers by convolutional neural networks.** 2016, doi:10.1109/BIBM.2016.7822593.
35. Wolkenhauer O, Mesarovic M: **Feedback dynamics and cell function: Why systems biology is called Systems Biology.** *Mol. Biosyst.* 2005, **1**:14–16.
- * An Opinion article that defines Systems Biology as the combination of Control Theory and Cell Biology. Authors make special emphasis in the analysis of system dynamics and put it at the core of the definition of Systems Biology.
36. Tyson JJ: **Irreversible Transitions , Bistability and Checkpoint Controls in the Eukaryotic Cell Cycle : A Systems-Level Understanding.** 2013, doi:10.1016/B978-0-12-385944-0.00014-9.
37. Hong T, Xing J, Li L, Tyson JJ: **A Mathematical Model for the Reciprocal Differentiation of T Helper 17 Cells and Induced Regulatory T Cells.** *J. Biophotonics* 2011, **7**:8.
38. Kuepfer L, Peter M, Sauer U, Stelling J: **Ensemble modeling for analysis of cell signaling dynamics.** *Nat. Biotechnol.* 2007, **25**:1001–6.
39. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: **Universally sloppy parameter sensitivities in systems biology models.** *PLoS Comput. Biol.* 2007, **3**:1871–78.
40. Eriksson O, Andersson T, Zhou Y, Tegner J: **Decoding complex biological networks - tracing essential and modulatory parameters in complex and simplified models of the cell cycle.** *BMC Syst Biol* 2011, **5**:123.
41. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z: **Reconstructing dynamic regulatory maps.** *Mol. Syst. Biol.* 2007, **3**:74.
- * An original analytical framework to derive models from time-series. Authors make use of an input-output hidden Markov model to model regulatory networks.
42. Härdin HM, Zagaris A, Krab K, Westerhoff H V.: **Simplified yet highly accurate enzyme kinetics for cases of low substrate concentrations.** *FEBS J.* 2009, **276**:5491–5506.

43. Sumner T, Shephard E, Bogle IDL: **A methodology for global-sensitivity analysis of time-dependent outputs in systems biology modelling.** *J. R. Soc. Interface* 2012, [no volume].
44. Rand DA: **Mapping global sensitivity of cellular network dynamics: sensitivity heat maps and a global summation law.** *J. R. Soc. Interface* 2008, **5 Suppl 1**:S59-69.
45. Geris L, Gomez-Cabrero D (Eds): *Uncertainty in Biology: A Computational Modeling Approach.* Springer International Publishing; 2015.
- ** This book provides a complete (and conceptually organized) overview of current challenges in Computational Modeling in three major sections: Modeling under uncertainty, Parameter estimation (both computational and experimental) and, most importantly, Hypothesis generation. The book provides the reader an easy-to-follow introduction and examples to all methods and challenges.
46. Gomez-Cabrero D, Ardid S, Cano-Colino M, Tegner J, Compte A: **Neuroswarm: A Methodology to Explore the Constraints that Function Imposes on Simulation Parameters in Large-Scale Networks of Biological Neurons.** In *Uncertainty in Biology: A Computational Modeling Approach.* Edited by Geris L, Gomez-Cabrero D. Springer International Publishing; 2015:442–447.
47. Cedersund G: **Conclusions via unique predictions obtained despite unidentifiability – new definitions and a general method.** 2012, **279**:3513–3527.
48. Menche J, Sharma A, Kitsak M, Ghiassian S, Vidal M, Loscalzo J, Barabási A-L: **Uncovering disease-disease relationships through the human interactome.** *Science (80-)*. 2015, **347**.
- * An outstanding example of a Network Medicine application and a natural follow-up of the Network Medicine field described in the review of Network Medicine generated by Barabasi and colleagues [6].
49. Park J, Lee D-S, Christakis N a, Barabási A-L, Data S: **The impact of cellular networks on disease comorbidity.** *Mol. Syst. Biol.* 2009, **5**.
50. Cusick ME, Klitgord N, Vidal M, Hill DE: **Interactome: Gateway into systems biology.** *Hum. Mol. Genet.* 2005, **14**:171–181.
51. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, et al.: **Architecture of the human regulatory network derived from ENCODE data.** *Nature* 2012, **489**:91–100.
52. Alon U: **Network motifs: theory and experimental approaches.** *Nat. Rev. Genet.* 2007, **8**:450–61.
53. Gomez-Cabrero D, Lluch-Ariet M, Tegnér J, Cascante M, Miralles F, Roca J: **Synergy-COPD: a systems approach for understanding and managing chronic diseases.** *J. Transl. Med.* 2014, **12 Suppl 2**:S2.
54. Gomez-Cabrero D, Menche J, Cano I, Abugessaisa I, Huertas-Migueláñez M, Tenyi A, de Mas I, Kiani N a, Marabita F, Falciani F, et al.: **Systems Medicine: from molecular features and models to the clinic in COPD.** *J. Transl. Med.* 2014, **12**.

* Overview of all efforts conducted in a truly Systems Medicine project, Synergy-COPD, using Chronic Obstructive Pulmonary Disease as a case-study. Outcomes and bottlenecks are both described. The manuscript also provides an introduction to a Synergy-COPD Supplement extending each one of the facets of Systems Medicine addressed in the project.

55. Li C, Donizelli M, Rodriguez N, Dharuri H, Endler L, Chelliah V, Li L, He E, Henry A, Stefan MI, et al.: **BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models.** *BMC Syst. Biol.* 2010, **4**:92.
56. Forster AC, Church GM: **Towards synthesis of a minimal cell.** *Mol. Syst. Biol.* 2006, **2**:45.
57. Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, Przybylski D, Platt RJ, Tirosh I, Sanjana NE, et al.: **A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks.** *Cell* 2015, **162**:675–686.
58. Oldham P, Hall S, Burton G: **Synthetic biology: mapping the scientific landscape.** *PLoS One* 2012, **7**:e34368.
59. Sismour M, Benner S: **Synthetic biology.** *Nat. Rev. Genet.* 2005, **6**:533–543.

* At the time of the publication of the paper, Synthetic Biology was becoming an emergent field. The paper reviews, in a well-described and organized manner, the basic elements of Synthetic Biology and their applications in biology and medicine.

60. Hutchison CA, Chuang R-YR-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, et al.: **Design and synthesis of a minimal bacterial genome.** *Science (80-.)*. 2016, **351**:aad6253-aad6253.
61. Weiss E-M, Schmidt A, Vobis D, Garbi N, Lahl K, Mayer CT, Sparwasser T, Ludwig A, Suri-Payer E, Oberle N, et al.: **Foxp3-Mediated Suppression of CD95L Expression Confers Resistance to Activation-Induced Cell Death in Regulatory T Cells.** *J. Immunol.* 2011, **187**:1684 LP-1691.
62. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F: **Genome engineering using the CRISPR-Cas9 system.** *Nat Protoc* 2013, **8**:2281–2308.
63. Komor AC, Badran AH, Liu DR: **CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes.** *Cell* 2016, **168**:1–17.
64. Ruder WC, Lu T, Collins JJ: **Synthetic biology moving into the clinic.** *Science* 2011, **333**:1248–1252.

** In this paper authors demonstrate the value of Synthetic Biology in clinical applications.

65. Jones P a.: **Functions of DNA methylation: islands, start sites, gene bodies and beyond.** *Nat. Rev. Genet.* 2012, doi:10.1038/nrg3230.
66. Molnár a, Georgopoulos K: **The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins.** *Mol. Cell. Biol.* 1994, **14**:8292–303.
67. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John

- S, Sandstrom R, Johnson AK, et al.: **An expansive human regulatory lexicon encoded in transcription factor footprints.** *Nature* 2012, **489**:83–90.
68. Berezikov E: **Evolution of microRNA diversity and regulation in animals.** *Nat. Rev. Genet.* 2011, **12**:846–60.
69. SpiceWorks: *Tech Flashback: Evolution of IT: 2000 vs. 2015.* 2015.
70. Council NR: *Frontiers in Massive Data Analysis.* The National Academies Press; 2013.
71. The 1000 Genomes Project Consortium: **A map of human genome variation from population-scale sequencing.** *Nature* 2010, **467**:1061–73.
72. Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat. Methods* 2008, **5**:1–8.
73. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, et al.: **ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia.** *Genome Res.* 2012, **22**:1813–31.
74. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C a., Doyle F, Epstein CB, Fietze S, Harrow J, Kaul R, et al.: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57–74.
75. Gomez-cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merckenschlager M, Gisel A, Ballestar E, Bongcam-rudloff E, Conesa A, Tegnér J: **Data integration in the era of omics : current and future challenges.** *BMC Syst. Biol.* 2014, **8**:11.
76. Gole J, Gore A, Richards A, Chiu Y-J, Fung H-L, Bushman D, Chiang H-I, Chun J, Lo Y-H, Zhang K: **Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells.** *Nat. Biotechnol.* 2013, **31**:1126–32.
77. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P, Linnarsson S: **Quantitative single-cell RNA-seq with unique molecular identifiers.** *Nat. Methods* 2013, **11**.
78. Marinov GK, Williams B a, McCue K, Schroth GP, Gertz J, Myers RM, Wold BJ: **From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing.** *Genome Res.* 2013, doi:10.1101/gr.161034.113.
79. La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, Borm LE, Stott SRW, Toledo EM, Villaescusa JC, et al.: **Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells.** *Cell* 2016, **167**:566–580.e19.
80. Guo H, Zhu P, Wu X, Li X, Wen L, Tang F: **Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing.** *Genome Res.* 2013, **23**:2126–35.
81. Ramani V, Deng X, Gunderson KL, Steemers FJ, Disteche CM, Noble WS, Duan Z, Shendure J: **Massively multiplex single-cell Hi-C.** *Nat. Methods* 2016, **14**:263–268.
82. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P: **Single-cell Hi-C reveals cell-to-cell variability in chromosome structure.** *Nature* 2013, **502**:59–64.
83. Macaulay IC, Ponting CP, Voet T: **Single-Cell Multiomics: Multiple Measurements from Single Cells.** *Trends Genet.* 2017, **xx**:1–14.
84. Vidal M, Cusick ME, Barabási A-L: **Interactome networks and human disease.** *Cell* 2011, **144**:986–98.
85. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási a L: **The large-scale organization of metabolic networks.** *Nature* 2000, **407**:651–4.
86. Liu Y-Y, Slotine J-J, Barabási A-L: **Controllability of complex networks.** *Nature* 2011, **473**:167–73.

87. Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs M a, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, et al.: **Visualization of omics data for systems biology.** *Nat. Methods* 2010, **7**:S56-68.
88. Cvijovic M, Höfer T, Aćimović J, Alberghina L, Almaas E, Besozzi D, Blomberg A, Bretschneider T, Cascante M, Collin O, et al.: **Strategies for structuring interdisciplinary education in Systems Biology: an European perspective.** *npj Syst. Biol. Appl.* 2016, **2**:16011.
89. Cascante M, de Atauri P, Gomez-Cabrero D, Wagner P, Centelles JJ, Marin S, Cano I, Velickovski F, Marin de Mas I, Maier D, et al.: **Workforce preparation: the Biohealth computing model for Master and PhD students.** *J. Transl. Med.* 2014, **12 Suppl 2**:S11.
90. Clermont G, Auffray C, Moreau Y, Rocke DM, Dalevi D, Dubhashi D, Marshall DR, Raasch P, Dehne F, Provero P, et al.: **Bridging the gap between systems biology and medicine.** *Genome Med.* 2009, **1**:88.
91. Gomez-Cabrero D, Marabita F, Tarazona S, Cano I, Roca J, Conesa A, Sabatier P, Tegnér J: **On the emergence of short advanced courses in Systems Medicine and Biology: recent experiences and recommended guidelines.** *Cell Syst.* 2017, [no volume].

* Educational efforts are necessary in the context of Systems Biology and Systems Medicine. Hypothesis driven research combining dry and wet-lab requires a new generation of scientists in a fast-changing and data-oriented research environment.

Limitations

- Unknown regulatory relevant features.
- Measurements of different omics from different samples.
- Snap-shots of cell stages.



Solutions

- Profiling: DNA Methylation, Chromatin Accessibility, .
- Multi-omics in the same sample (e.g. DNA and RNA).
- Multi-omics in the same cells.
- Technologies for continuous monitoring

Biological Knowledge

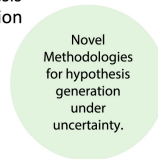
Computational Modeling (*dry*)



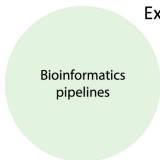
Analysis of Experiments



Hypothesis Generation



Wet-lab experiments



Limitations on 2002

- Costs for high-throughput very high.
- Amount of material required for samples.



Solutions

- Reducing cost of sequencing.
- Low-quantities and single-cell protocols.

Limitations

- Computational Resources.
- Generating System Structure.
- Profiles generated as average over cell populations.
- Limited number of samples



Solutions

- Increase in Computational power.
- Increased availability of Computational resources.
- Single-cell technologies.

Highlights

- A review in 2017 of Systems Biology through visionary Kitano's 2002 work.
- The gaps in hypothesis driven research and the existing tools to overcome them.
- Future of Systems Medicine.

ACCEPTED MANUSCRIPT