Spring 2019 – Epigenetics and Systems Biology
Lecture Outline (Systems Biology)
Michael K. Skinner – Biol 476/576
CUE 418, 10:35-11:50 am, Tuesdays & Thursdays
January 22 & 29, 2019
Weeks 3 and 4

Systems Biology (Components & Technology)

Components (DNA, Expression, Cellular, Organ, Physiology, Organism, Differentiation, Development, Phenotype, Evolution)

Technology (Genomics, Transcriptomes, Proteomics)
(Interaction, Signaling, Metabolism)

Omics (Data Processing and Resources)

Required Reading

ENCODE (2012) ENCODE Explained. Nature 489:52-55.

Tavassoly I, Goldfarb J, Iyengar R. (2018) Essays Biochem. 62(4):487-500.

Literature

Sundaram V, Wang T. Transposable Element Mediated Innovation in Gene Regulatory Landscapes of Cells: Re-Visiting the "Gene-Battery" Model. Bioessays. 2018 40(1). doi: 10.1002/bies.201700155.

Abil Z, Ellefson JW, Gollihar JD, Watkins E, Ellington AD. Compartmentalized partnered replication for the directed evolution of genetic parts and circuits. Nat Protoc. 2017 Dec;12(12):2493-2512.

Filipp FV. Crosstalk between epigenetics and metabolism-Yin and Yang of histone demethylases and methyltransferases in cancer. Brief Funct Genomics. 2017 Nov 1;16(6):320-325.

Chen Z, Li S, Subramaniam S, Shyy JY, Chien S. Epigenetic Regulation: A New Frontier for Biomedical Engineers. Annu Rev Biomed Eng. 2017 Jun 21;19:195-219.

Hollick JB. Paramutation and related phenomena in diverse species. Nat Rev Genet. 2017 Jan;18(1):5-23.

Macovei A, Pagano A, Leonetti P, Carbonera D, Balestrazzi A, Araújo SS. Systems biology and genome-wide approaches to unveil the molecular players involved in the pre-germinative metabolism: implications on seed technology traits. Plant Cell Rep. 2017 May;36(5):669-688.

Bagchi DN, Iyer VR. The Determinants of Directionality in Transcriptional Initiation. Trends Genet. 2016 Jun;32(6):322-333.

Sosa-Hernández JE, Villalba-Rodríguez AM, Romero-Castillo KD, Aguilar-Aguila-Isaías MA, García-Reyes IE, Hernández-Antonio A, Ahmed I, Sharma A, Parra-Saldívar R, Iqbal HMN. Organs-on-a-

Chip Module: A Review from the Development and Applications Perspective. Micromachines (Basel). 2018 Oct 22;9(10).
Talug B, Tokcaer-Keskin Z. Induced Pluripotent Stem Cells in Disease Modelling and Regeneration. Adv Exp Med Biol. 2018 Nov 10. doi: 10.1007/5584_2018_290. [Epub ahead of print]
Tavassoly I, Goldfarb J, Iyengar R. Systems biology primer: the basic methods and approaches. Essays Biochem. 2018 Oct 26;62(4):487-500.
Gomez-Pinilla F, Yang X. System biology approach intersecting diet and cell metabolism with pathogenesis of brain disorders. Prog Neurobiol. 2018 Oct;169:76-90.
Nawroth J, Rogal J, Weiss M, Brucker SY, Loskill P. Organ-on-a-Chip Systems for Women's Health Applications. Adv Healthc Mater. 2018 Jan;7(2). doi: 10.1002/adhm.201700550.
Choi J, Iich E, Lee JH. Organogenesis of adult lung in a dish: Differentiation, disease and therapy. Dev Biol. 2016 Dec 15;420(2):278-286.
Kurz FT, Kembro JM, Flesia AG, Armoundas AA, Cortassa S, Aon MA, Lloyd D. Network dynamics: quantitative analysis of complex behavior in metabolism, organelles, and cells, from experiments to models and back. Wiley Interdiscip Rev Syst Biol Med. 2017 Jan;9(1). doi: 10.1002/wsbm.1352. Epub 2016 Sep 7.
Frank JA, Broichhagen J, Yushchenko DA, Trauner D, Schultz C, Hodson DJ. Optical tools for understanding the complexity of β-cell signalling and insulin release. Nat Rev Endocrinol. 2018 Dec;14(12):721-737.
Zhou Y, Horowitz JC, Naba A, et al. Extracellular matrix in lung development, homeostasis and disease. Matrix Biol. 2018 Nov;73:77-104.
Yadav A, Sinha H. Gene-gene and gene-environment interactions in complex traits in yeast. Yeast. 2018 Jun;35(6):403-416.
Leulier F, MacNeil LT, Lee WJ, Rawls JF, Cani PD, Schwarzer M, Zhao L, Simpson SJ. Integrative Physiology: At the Crossroads of Nutrition, Microbiota, Animal Physiology, and Human Health. Cell Metab. 2017 Mar 7;25(3):522-534.
Chen X, Gonçalves MAFV. DNA, RNA, and Protein Tools for Editing the Genetic Information in Human Cells. iScience. 2018 Aug 31;6:247-263.
Howe DG, Blake JA, Bradford YM, Bult CJ, Calvi BR, Engel SR, Kadin JA, Kaufman TC, Kishore R, Laulederkind SJF, Lewis SE, Moxon SAT, Richardson JE, Smith C. Model organism data evolving in support of translational medicine. Lab Anim (NY). 2018 Oct;47(10):277-289.
Kiani NA, Shang MM, Zenil H, Tegner J. Predictive Systems Toxicology. Methods Mol Biol. 2018;1800:535-557.
Peng X, Zhang Q, Liao C, Han W Xu F. Epigenomic Control of Thermogenic Adipocyte Differentiation and Function. Int J Mol Sci. 2018 Jun 17;19(6).
Griffiths JA, Scialdone A, Marioni JC. Using single-cell genomics to understand developmental processes and cell fate decisions. Mol Syst Biol. 2018 Apr 16;14(4):e8046.
Werner S, Vu HT, Rink JC. Self-organization in development, regeneration and organoids. Curr Opin Cell Biol. 2017 Feb;44:102-109.
Soyer OS, O'Malley MA. Evolutionary systems biology: what it is and why it matters. Bioessays. 2013 Aug;35(8):696-705.
Kasper C, Vierbuchen M, Ernst U, Fischer S, Radersma R, Raulo A, Cunha-Saraiva F, Wu M, Mobley KB, Taborsky B. Genetics and developmental biology of cooperation. Mol Ecol. 2017 Sep;26(17):4364-4377.
Van Laere S, Dirix L, Vermeulen P. Molecular profiles to biology and pathways: a systems biology approach. Chin J Cancer. 2016 Jun 16;35(1):53.

Nam S. Databases and tools for constructing signal transduction networks in cancer. BMB Rep. 2017 Jan;50(1):12-19.
Ronan T, Qi Z, Naegle KM. Avoiding common pitfalls when clustering biological data. Sci Signal. 2016 Jun 14;9(432):re6.
Zheng H, Porebski PJ, Grabowski M Cooper DR1, Minor W. Databases, Repositories, and Other Data Resources in Structural Biology. Methods Mol Biol. 2017;1607:643-665.
Lin Y, Qian F, Shen L, Chen F, Chen J, Shen B. Computer-aided biomarker discovery for precision medicine: data resources, models and applications. Brief Bioinform. 2017 Nov 29. doi: 10.1093/bib/bbx158. [Epub ahead of print]
Köhler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res. 2017 Jan 4;45(D1):D865-D876.
Kannan L, Ramos M, Re A, El-Hachem N, et al. Public data and open source tools for multi-assay genomic investigation of disease. Brief Bioinform. 2016 Jul;17(4):603-15.
Sun YV, Hu YJ. (2016) Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. Adv Genet. 2016;93:147-90.
Horgusluoglu E, Nudelman K, Nho K, Saykin AJ. (2016) Adult neurogenesis and neurodegenerative diseases: A systems biology perspective. Am J Med Genet B Neuropsychiatr Genet. 2016 Feb 16. doi: 10.1002/ajmg.b.32429. [Epub ahead of print]
Zenil H, Kiani NA, Tegnér J. (2016) Methods of information theory and algorithmic complexity for network biology. Semin Cell Dev Biol. 51:32-43.
Ostaszewski M, Skupin A, Balling R. (2016) Neurological Diseases from a Systems Medicine Point of View. Methods Mol Biol. 2016;1386:221-50.
Nishi A, Milner DA Jr, Giovannucci EL, et al. (2016) Integration of molecular pathology, epidemiology and social science for global precision medicine. Expert Rev Mol Diagn. 2016;16(1):11-23.
Davidsen PK, Turan N, Egginton S, Falciani F. (1985) Multilevel functional genomics data integration as a tool for understanding physiology: a network biology perspective. J Appl Physiol (1985). 2016 Feb 1;120(3):297-309.
Prathipati P, Mizuguchi K. (2016) Systems Biology Approaches to a Rational Drug Discovery Paradigm. Curr Top Med Chem. 2016;16(9):1009-25.
Qin Y, Jiao X1, Simpson JL, Chen ZJ. (2015) Genetics of primary ovarian insufficiency: new developments and opportunities. Hum Reprod Update. 2015 Nov-Dec;21(6):787-808.
Parikshak NN, Gandal MJ, Geschwind DH. (2015) Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat Rev Genet. 16(8):441-58.
Xie L, Draizen EJ, Bourne PE. (2016) Harnessing Big Data for Systems Pharmacology. Annu Rev Pharmacol Toxicol. 2016 Oct 13. [Epub ahead of print]
Macovei A, Pagano A, Leonetti P, Carbonera D, Balestrazzi A, Araújo SS. (2016) Systems biology and genome-wide approaches to unveil the molecular players involved in the pre-germinative metabolism: implications on seed technology traits. Plant Cell Rep. 2016 Oct 11. [Epub ahead of print]
Tosto G, Reitz C. (2016) Use of "omics" technologies to dissect neurologic disease. Handb Clin Neurol. 2016;138:91-106.
Altaf-Ul-Amin M, Afendi FM, Kiboi SK, Kanaya S. (2014) Systems biology in the context of big data and networks. Biomed Res Int. 2014;2014:428570.
Putri SP, Nakayama Y, Matsuda F, et al. (2013) Current metabolomics: practical applications. J Biosci Bioeng. 115(6):579-89.

Manning T, Sleator RD, Walsh P. (2013) Naturally selecting solutions: the use of genetic algorithms in bioinformatics. Bioengineered. 4(5):266-78.
Shekari F, Baharvand H, Salekdeh GH. (2014) Organellar proteomics of embryonic stem cells. Adv Protein Chem Struct Biol. 95:215-30.
Wu X, Hasan MA, Chen JY. (2014) Pathway and network analysis in proteomics. J Theor Biol. 2014 Jun 6. pii: S0022-5193(14)00304-X. [Epub ahead of print]
Liu Z, Wang Y, Xue Y. (2013) Phosphoproteomics-based network medicine. FEBS J. 280(22):5696-704.
Dharuri H, Demirkan A, van Klinken JB, et al. (2014) Genetics of the human metabolome, what is next? Biochim Biophys Acta.1842(10):1923-1931.
Stumpf MP. (2014) Approximate Bayesian inference for complex ecosystems. F1000Prime Rep. 17;6:60.
Purcell O, Lu TK. (2014) Synthetic analog and digital circuits for cellular computation and memory. Curr Opin Biotechnol. 29:146-55.
Mason CE, Porter SG, Smith TM. (2014) Characterizing multi-omic data in systems biology. Adv Exp Med Biol. 799:15-38.
Sarpeshkar R. (2014) Analog synthetic biology. Philos Trans A Math Phys Eng Sci. 24;372(2012):20130110.
Rekhi R, Qutub AA. (2013) Systems approaches for synthetic biology: a pathway toward mammalian design. Front Physiol. 9;4:285.
Renda BA, Hammerling MJ, Barrick JE. (2014) Engineering reduced evolutionary potential for synthetic biology. Mol Biosyst. 10(7):1668-78.
Cronin RM, Field JR, Bradford Y, et al. (2014) Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. Front Genet. 5;5:250.
Svahn AJ, Becker TS, Graeber MB. (2014) Emergent properties of microglia. Brain Pathol. 24(6):665-70.
Caterino M, Aspesi A, Pavesi E, et al. (2014) Analysis of the interactome of ribosomal protein S19 mutants. Proteomics. 14(20):2286-96.
Singh R, Dangol S, Jwa NS. (2014) Yeast two-hybrid system for dissecting the rice MAPK interactome. Methods Mol Biol. 1171:195-216.
Petrey D, Honig B. (2014) Structural bioinformatics of the interactome. Annu Rev Biophys. 43:193-210.
Garcia B, Datta G, Cosgrove GP, Strong M. (2014) Network and matrix analysis of the respiratory disease interactome. BMC Syst Biol. 22;8:34.
Blomme J, Inzé D, Gonzalez N. (2014) The cell-cycle interactome: a source of growth regulators? J Exp Bot. 65(10):2715-30.
Stevens A, De Leonibus C, Hanson D, et al. (2014) Network analysis: a new approach to study endocrine disorders. J Mol Endocrinol. 19;52(1):R79-93.
Salvo SA, Hirsch CN, Buell CR, Kaeppler S, Kaeppler HF. (2014) Whole Transcriptome Profiling of Maize during Early Somatic Embryogenesis Reveals Altered Expression of Stress Factors and Embryogenesis-Related Genes. PLoS One. 30;9(10):e111407.
Xuan J, Yu Y, Qing T, Guo L, Shi L. (2013) Next-generation sequencing in the clinic: promises and challenges. Cancer Lett. 1;340(2):284-95.
Robinson SW, Fernandes M, Husi H. (2014) Current advances in systems and integrative biology. Comput Struct Biotechnol J. 11(18):35-46.

Sharma A, Rai A, Lal S. (2013) Workflow management systems for gene sequence analysis and evolutionary studies - A Review. Bioinformation. 17;9(13):663-72.
Street ME, Buscema M, Smerieri A, Montanini L, Grossi E. (2013) Artificial Neural Networks, and Evolutionary Algorithms as a systems biology approach to a data-base on fetal growth restriction. Prog Biophys Mol Biol. 113(3):433-8.
Caccia D, Dugo M, Callari M, Bongarzone I. (2013) Bioinformatics tools for secretome analysis. Biochim Biophys Acta. 1834(11):2442-53.
Ecker JR, et al. (2012) Genomics: ENCODE explained. Nature. 6;489(7414):52-5.
Gerstein MB, et al. (2012) Architecture of the human regulatory network derived from ENCODE data. Nature. 6;489(7414):91-100.
Thurman RE, et al. (2012) The accessible chromatin landscape of the human genome. Nature. 6;489(7414):75-82.
Sanyal A, Lajoie BR, Jain G, Dekker J. (2012) The long-range interaction landscape of gene promoters. Nature. 6;489(7414):109-13.
Afacan NJ, Fjell CD, Hancock RE. (2012) A systems biology approach to nutritional immunology - focus on innate immunity. Mol Aspects Med. 33(1):14-25.
Wilson RA. (2012) The cell biology of schistosomes: a window on the evolution of the early metazoa. Protoplasma. 249(3):503-18.
Murphy BF, Thompson MB. (2012) A review of the evolution of viviparity in squamate reptiles: the past, present and future role of molecular biology and genomics. J Comp Physiol B. 181(5):575-94.
Fritzsch FS, Dusny C, Frick O, Schmid A. (2012) Single-cell analysis in biotechnology, systems biology, and biocatalysis. Annu Rev Chem Biomol Eng. 3:129-55.
Tian Q, Price ND, Hood L. (2012) Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine. J Intern Med. 271(2):111-21.
Weckwerth W. (2011) Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. J Proteomics. 10;75(1):284-305.
St-Denis N, Gingras AC. (2012) Mass spectrometric tools for systematic analysis of protein phosphorylation. Prog Mol Biol Transl Sci. 106:3-32.
Amit I, Regev A, Hacohen N. (2011) Strategies to discover regulatory circuits of the mammalian immune system. Nat Rev Immunol. 18;11(12):873-80.
Zhao S, Iyengar R. (2012) Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. Annu Rev Pharmacol Toxicol. 10;52:505-21.
Habibi E, Masoudi-Nejad A, Abdolmaleky HM, Haggarty SJ. (2011) Emerging roles of epigenetic mechanisms in Parkinson's disease. Funct Integr Genomics. 11(4):523-37.
Liu ZP, Chen L. (2012) Proteome-wide prediction of protein-protein interactions from high-throughput data. Protein Cell. 3(7):508-20.
Markowitz VM, et al. (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res. 40(Database issue):D115-22.
Diercks A, Aderem A. (2012) Systems Approaches to Dissecting Immunity. Curr Top Microbiol Immunol. 2012 Aug 11. [Epub ahead of print]
Scholz B, Marschalek R. (2012) Epigenetics and blood disorders. Br J Haematol. 158(3):307-22.
Borenstein E. (2012) Computational systems biology and in silico modeling of the human microbiome. Brief Bioinform. 13(6):769-80.
Gupta RK, Rosen ED, Spiegelman BM. (2011) Identifying novel transcriptional components controlling energy metabolism. Cell Metab. 7;14(6):739-45.

Galliot B, Quiquand M. (2011) A two-step process in the emergence of neurogenesis. Eur J Neurosci. 34(6):847-62.
Rodin AS, Gogoshin G, Boerwinkle E. (2011) Systems biology data analysis methodology in pharmacogenomics. Pharmacogenomics. 12(9):1349-60.
Sobie EA, Lee YS, Jenkins SL, Iyengar R. (2011) Systems biology--biomedical modeling. Sci Signal. 6;4(190):tr2
Habibi E, Masoudi-Nejad A, Abdolmaleky HM, Haggarty SJ. (2011) Emerging roles of epigenetic mechanisms in Parkinson's disease. Funct Integr Genomics. 11(4):523-37.
Day JJ, Sweatt JD. (2012) Epigenetic treatments for cognitive impairments. Neuropsychopharmacology. 37(1):247-60.
Prezioso C, Orlando V. (2011) Polycomb proteins in mammalian cell differentiation and plasticity. FEBS Lett. 7;585(13):2067-77.
Zhang X, Yap Y, Wei D, Chen G, Chen F. Novel omics technologies in nutrition research. Biotechnol Adv. 2008 Mar-Apr;26(2):169-76.

FORUM: Genomics

ENCODE explained

The Encyclopedia of DNA Elements (ENCODE) project dishes up a hearty banquet of data that illuminate the roles of the functional elements of the human genome. Here, five scientists describe the project and discuss how the data are influencing research directions across many fields. SEE ARTICLES P.57, P.75, P.83, P.91, P.101 & LETTER P.109

Serving up a genome feast

JOSEPH R. ECKER

Starting with a list of simple ingredients and blending them in the precise amounts needed to prepare a gourmet meal is a challenging task. In many respects, this task is analogous to the goal of the ENCODE project[1], the recent progress of which is described in this issue[2–7]. The project aims to fully describe the list of common ingredients (functional elements) that make up the human genome (Fig. 1). When mixed in the right proportions, these ingredients constitute the information needed to build all the types of cells, body organs and, ultimately, an entire person from a single genome.

The ENCODE pilot project[8] focused on just 1% of the genome — a mere appetizer — and its results hinted that the list of human genes was incomplete. Although there was scepticism about the feasibility of scaling up the project to the entire genome and to many hundreds of cell types, recent advances in low-cost, rapid DNA-sequencing technology radically changed that view[9]. Now the ENCODE consortium presents a menu of 1,640 genome-wide data sets prepared from 147 cell types, providing a six-course serving of papers in *Nature*, along with many companion publications in other journals.

One of the more remarkable findings described in the consortium's 'entrée' paper (page 57)[2] is that 80% of the genome contains elements linked to biochemical functions, dispatching the widely held view that the human genome is mostly 'junk DNA'. The authors report that the space between genes is filled with enhancers (regulatory DNA elements), promoters (the sites at which DNA's transcription into RNA is initiated) and numerous previously overlooked regions that encode RNA transcripts that are not translated into proteins but might have regulatory roles. Of note, these results show that many DNA variants previously correlated with certain diseases lie within or very near non-coding functional DNA elements, providing new leads for linking genetic variation and disease.

The five companion articles[3–7] dish up diverse sets of genome-wide data regarding the mapping of transcribed regions, DNA binding of regulatory proteins (transcription factors) and the structure and modifications of chromatin (the association of DNA and proteins that makes up chromosomes), among other delicacies.

Djebali and colleagues[3] (page 101) describe ultra-deep sequencing of RNAs prepared from many different cell lines and from specific compartments within the cells. They conclude that about 75% of the genome is transcribed at some point in some cells, and that genes are highly interlaced with overlapping transcripts that are synthesized from both DNA strands. These findings force a rethink of the definition of a gene and of the minimum unit of heredity.

Moving on to the second and third courses, Thurman *et al.*[4] and Neph *et al.*[5] (pages 75 and 83) have prepared two tasty chromatin-related treats. Both studies are based on the DNase I hypersensitivity assay, which detects genomic regions at which enzyme access to, and subsequent cleavage of, DNA is unobstructed by chromatin proteins. The authors identified cell-specific patterns of DNase I hypersensitive sites that show remarkable concordance with experimentally determined and computationally predicted binding sites of transcription factors. Moreover, they have doubled the number of known recognition sequences for DNA-binding proteins in the human genome, and have revealed a 50-base-pair 'footprint' that is present in thousands of promoters[5].

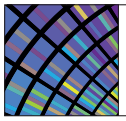
The next course, provided by Gerstein and colleagues[6] (page 91) examines the principles behind the wiring of transcription-factor networks. In addition to assigning relatively simple functions to genome elements (such as 'protein X binds to DNA element Y'), this study attempts to clarify the hierarchies of transcription factors and how the intertwined networks arise.

Beyond the linear organization of genes and transcripts on chromosomes lies a more complex (and still poorly understood) network of chromosome loops and twists through which promoters and more distal elements, such as enhancers, can communicate their regulatory information to each other. In the final course of the ENCODE genome feast, Sanyal and colleagues[7] (page 109) map more than 1,000 of these long-range signals in each cell type. Their findings begin to overturn the long-held (and probably oversimplified) prediction that the regulation of a gene is dominated by its proximity to the closest regulatory elements.

> ***"These findings force a rethink of the definition of a gene and of the minimum unit of heredity."***

One of the major future challenges for ENCODE (and similarly ambitious projects) will be to capture the dynamic aspects of gene regulation. Most assays provide a single snapshot of cellular regulatory events, whereas a time series capturing how such processes change is preferable. Additionally, the examination of large batches of cells — as required for the current assays — may present too simplified a view of the underlying regulatory complexity, because individual cells in a batch (despite being genetically identical) can sometimes behave in different ways. The development of new technologies aimed at the simultaneous capture of multiple data types, along with their regulatory dynamics in single cells, would help to tackle these issues.

A further challenge is identifying how the genomic ingredients are combined to assemble the gene networks and biochemical pathways that carry out complex functions, such as cell-to-cell communication, which enable organs and tissues to develop. An even greater challenge will be to use the rapidly growing body



ENCODE
Encyclopedia of DNA Elements
nature.com/encode

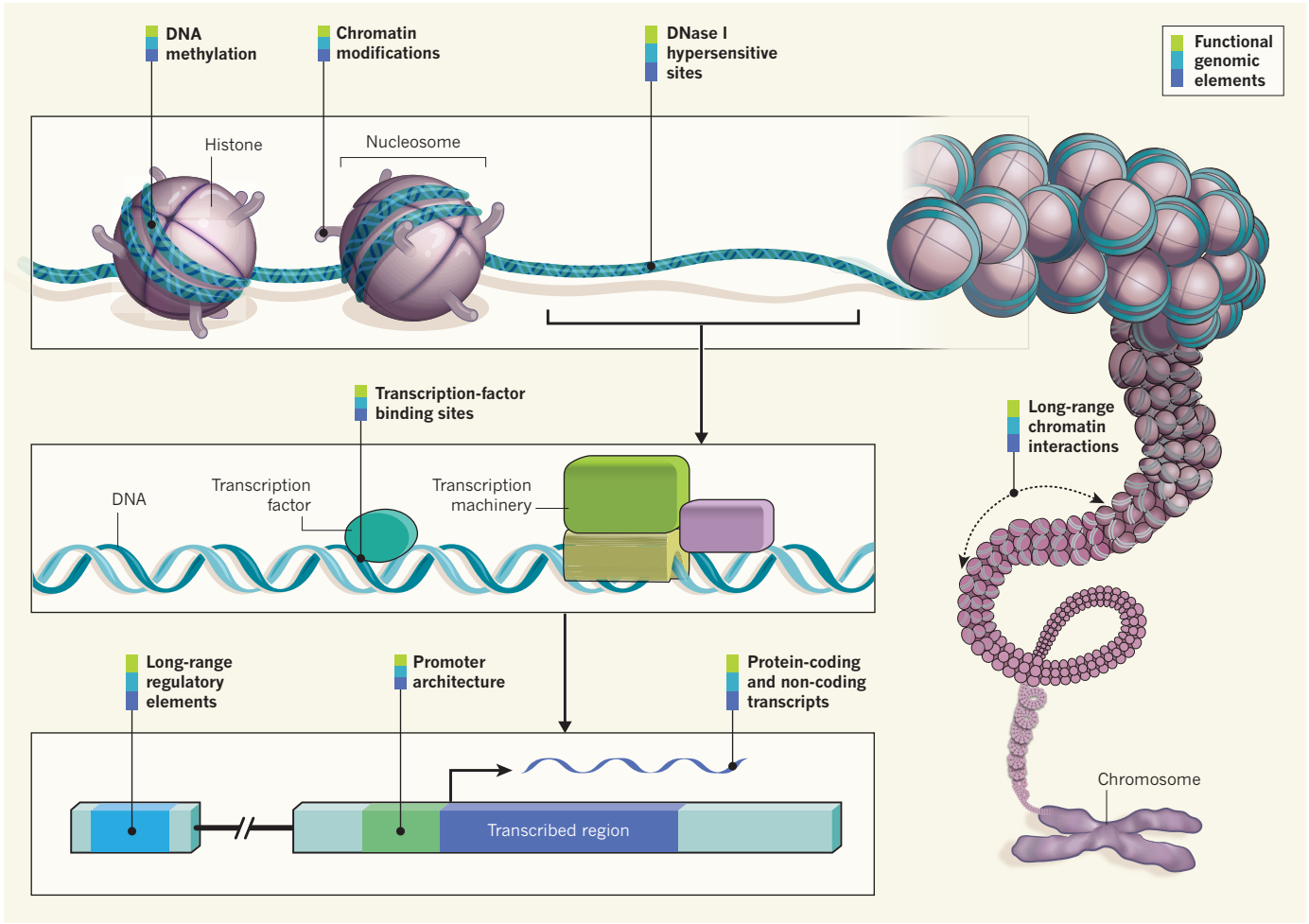


Figure 1 | Beyond the sequence. The ENCODE project[2–7] provides information on the human genome far beyond that contained within the DNA sequence — it describes the functional genomic elements that orchestrate the development and function of a human. The project contains data about the degree of DNA methylation and chemical modifications to histones that can influence the rate of transcription of DNA into RNA molecules (histones are the proteins around which DNA is wound to form chromatin). ENCODE also examines long-range chromatin interactions, such as looping, that alter the relative proximities of different chromosomal regions in three dimensions and also affect transcription. Furthermore, the project describes the binding activity of transcription-factor proteins and the architecture (location and sequence) of gene-regulatory DNA elements, which include the promoter region upstream of the point at which transcription of an RNA molecule begins, and more distant (long-range) regulatory elements. Another section of the project was devoted to testing the accessibility of the genome to the DNA-cleavage protein DNase I. These accessible regions, called DNase I hypersensitive sites, are thought to indicate specific sequences at which the binding of transcription factors and transcription-machinery proteins has caused nucleosome displacement. In addition, ENCODE catalogues the sequences and quantities of RNA transcripts, from both non-coding and protein-coding regions.

of data from genome-sequencing projects to understand the range of human phenotypes (traits), from normal developmental processes, such as ageing, to disorders such as Alzheimer's disease[10].

Achieving these ambitious goals may require a parallel investment of functional studies using simpler organisms — for example, of the type that might be found scampering around the floor, snatching up crumbs in the chefs' kitchen. All in all, however, the ENCODE project has served up an all-you-can-eat feast of genomic data that we will be digesting for some time. Bon appétit!

Joseph R. Ecker *is at the Howard Hughes Medical Institute and the Salk Institute for Biological Studies, La Jolla, California 92037, USA.*
e-mail: ecker@salk.edu

Expression control

WENDY A. BICKMORE

Once the human genome had been sequenced, it became apparent that an encyclopaedic knowledge of chromatin organization would be needed if we were to understand how gene expression is regulated. The ENCODE project goes a long way to achieving this goal and highlights the pivotal role of transcription factors in sculpting the chromatin landscape.

Although some of the analyses largely confirm conclusions from previous smaller-scale studies, this treasure trove of genome-wide data provides fresh insight into regulatory pathways and identifies prodigious numbers of regulatory elements. This is particularly so for Thurman and colleagues' data[4] regarding DNase I hypersensitive sites (DHSs) and for Gerstein and colleagues' results[6] concerning DNA binding of transcription factors. DHSs are genomic regions that are accessible to enzymatic cleavage as a result of the displacement of nucleosomes (the basic units of chromatin) by DNA-binding proteins (Fig. 1). They are the hallmark of cell-type-specific enhancers, which are often located far away from promoters.

The ENCODE papers expose the profusion of DHSs — more than 200,000 per cell type, far outstripping the number of promoters — and their variability between cell types. Through the simultaneous presence in the same cell type of a DHS and a nearby active promoter, the researchers paired half a million enhancers with their probable target genes. But this leaves

11 Years Ago

The draft human genome

OUR GENOME UNVEILED

Unless the human genome contains a lot of genes that are opaque to our computers, it is clear that we do not gain our undoubted complexity over worms and plants by using many more genes. Understanding what does give us our complexity — our enormous behavioural repertoire, ability to produce conscious action, remarkable physical coordination (shared with other vertebrates), precisely tuned alterations in response to external variations of the environment, learning, memory … need I go on? — remains a challenge for the future.

David Baltimore

From *Nature* 15 February 2001

GENOME SPEAK

With the draft in hand, researchers have a new tool for studying the regulatory regions and networks of genes. Comparisons with other genomes should reveal common regulatory elements, and the environments of genes shared with other species may offer insight into function and regulation beyond the level of individual genes. The draft is also a starting point for studies of the three-dimensional packing of the genome into a cell's nucleus. Such packing is likely to influence gene regulation … The human genome lies before us, ready for interpretation.

Peer Bork and Richard Copley

From *Nature* 15 February 2001

more than 2 million putative enhancers without known targets, revealing the enormous expanse of the regulatory genome landscape that is yet to be explored. Chromosome-conformation-capture methods that detect long-range physical associations between distant DNA regions are attempting to bridge this gap. Indeed, Sanyal and colleagues[7] applied these techniques to survey such associations across 1% of the genome.

The ENCODE data start to paint a picture of the logic and architecture of transcriptional networks, in which DNA binding of a few high-affinity transcription factors displaces nucleosomes and creates a DHS, which in turn facilitates the binding of further, lower-affinity factors. The results also support the idea that transcription-factor binding can block DNA methylation (a chemical modification of DNA that affects gene expression), rather than the other way around — which is highly relevant to the interpretation of disease-associated sites of altered DNA methylation[11].

The exquisite cell-type specificity of regulatory elements revealed by the ENCODE studies emphasizes the importance of having appropriate biological material on which to test hypotheses. The researchers have focused their efforts on a set of well-established cell lines, with selected assays extended to some freshly isolated cells. Challenges for the future include following the dynamic changes in the regulatory landscape during specific developmental pathways, and understanding chromatin structure in tissues containing heterogeneous cell populations.

Wendy A. Bickmore *is in the Medical Research Council Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.*
e-mail: wendy.bickmore@igmm.ed.ac.uk

Non-coding but functional

INÊS BARROSO

The vast majority of the human genome does not code for proteins and, until now, did not seem to contain defined gene-regulatory elements. Why evolution would maintain large amounts of 'useless' DNA had remained a mystery, and seemed wasteful. It turns out, however, that there are good reasons to keep this DNA. Results from the ENCODE project[2–8] show that most of these stretches of DNA harbour regions that bind proteins and RNA molecules, bringing these into positions from which they cooperate with each other to regulate the function and level of expression of protein-coding genes. In addition, it seems that widespread transcription from non-coding DNA potentially acts as a reservoir for the creation of new functional molecules, such as regulatory RNAs.

What are the implications of these results for genetic studies of complex human traits and disease? Genome-wide association studies (GWAS), which link variations in DNA sequence with specific traits and diseases, have in recent years become the workhorse of the field, and have identified thousands of DNA variants associated with hundreds of complex traits (such as height) and diseases (such as diabetes). But association is not causality, and identifying those variants that are causally linked to a given disease or trait, and understanding how they exert such influence, has been difficult. Furthermore, most of these associated variants lie in non-coding regions, so their functional effects have remained undefined.

"The results imply that sequencing studies focusing on protein-coding sequences risk missing crucial parts of the genome."

The ENCODE project provides a detailed map of additional functional non-coding units in the human genome, including some that have cell-type-specific activity. In fact, the catalogue contains many more functional non-coding regions than genes. These data show that results of GWAS are typically enriched for variants that lie within such non-coding functional units, sometimes in a cell-type-specific manner that is consistent with certain traits, suggesting that many of these regions could be causally linked to disease. Thus, the project demonstrates that non-coding regions must be considered when interpreting GWAS results, and it provides a strong motivation for reinterpreting previous GWAS findings. Furthermore, these results imply that sequencing studies focusing on protein-coding sequences (the 'exome') risk missing crucial parts of the genome and the ability to identify true causal variants.

However, although the ENCODE catalogues represent a remarkable tour de force, they contain only an initial exploration of the depths of our genome, because many more cell types must yet be investigated. Some of the remaining challenges for scientists searching for causal disease variants lie in: accessing data derived from cell types and tissues relevant to the disease under study; understanding how these functional units affect genes that may be distantly located[7]; and the ability to generalize such results to the entire organism.

Inês Barroso *is at the Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK, and at the University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Cambridge, UK.*
e-mail: ib1@sanger.ac.uk

Evolution and the code

JONATHAN K. PRITCHARD & YOAV GILAD

One of the great challenges in evolutionary biology is to understand how differences in DNA sequence between species determine differences in their phenotypes. Evolutionary change may occur both through changes in protein-coding sequences and through sequence changes that alter gene regulation.

There is growing recognition of the importance of this regulatory evolution, on the basis of numerous specific examples as well as on theoretical grounds. It has been argued that potentially adaptive changes to protein-coding sequences may often be prevented by natural selection because, even if they are beneficial in one cell type or tissue, they may be detrimental elsewhere in the organism. By contrast, because gene-regulatory sequences are frequently associated with temporally and spatially specific gene-expression patterns, changes in these regions may modify the function of only certain cell types at specific times, making it more likely that they will confer an evolutionary advantage[12].

However, until now there has been little information about which genomic regions have regulatory activity. The ENCODE project has provided a first draft of a 'parts list' of these regulatory elements, in a wide range of cell types, and moves us considerably closer to one of the key goals of genomics: understanding the functional roles (if any) of every position in the human genome.

Nonetheless, it will take a great deal of work to identify the critical sequence changes in the newly identified regulatory elements that drive functional differences between humans and other species. There are some precedents for identifying key regulatory differences (see, for example, ref. 13), but ENCODE's improved identification of regulatory elements should greatly accelerate progress in this area. The data may also allow researchers to begin to identify sequence alterations occurring simultaneously in multiple genomic regions, which, when added together, drive phenotypic change — a process called polygenic adaptation[14].

However, despite the progress brought by the ENCODE consortium and other research groups, it remains difficult to discern with confidence which variants in putative regulatory regions will drive functional changes, and what these changes will be. We also still have an incomplete understanding of how regulatory sequences are linked to target genes. Furthermore, the ENCODE project focused mainly on the control of transcription, but many aspects of post-transcriptional regulation, which may also drive evolutionary changes, are yet to be fully explored.

Nonetheless, these are exciting times for studies of the evolution of gene regulation. With such new resources in hand, we can expect to see many more descriptions of adaptive regulatory evolution, and how this has contributed to human evolution.

Jonathan K. Pritchard *and* **Yoav Gilad** *are in the Department of Human Genetics, University of Chicago, Chicago 60637 Illinois, USA.* **J.K.P.** *is also at the Howard Hughes Medical Institute, University of Chicago.*
e-mails: pritch@uchicago.edu; gilad@uchicago.edu

From catalogue to function

ERAN SEGAL

Projects that produce unprecedented amounts of data, such as the human genome project[15] or the ENCODE project, present new computational and data-analysis challenges and have been a major force driving the development of computational methods in genomics. The human genome project produced one bit of information per DNA base pair, and led to advances in algorithms for sequence matching and alignment. By contrast, in its 1,640 genome-wide data sets, ENCODE provides a profile of the accessibility, methylation, transcriptional status, chromatin structure and bound molecules for every base pair. Processing the project's raw data to obtain this functional information has been an immense effort.

For each of the molecular-profiling methods used, the ENCODE researchers devised novel processing algorithms designed to remove outliers and protocol-specific biases, and to ensure the reliability of the derived functional information. These processing pipelines and quality-control measures have been adapted by the research community as the standard for the analysis of such data. The high quality of the functional information they produce is evident from the exquisite detail and accuracy achieved, such as the ability to observe the crystallographic topography of protein–DNA interfaces in DNase I footprints[5], and the observation of more than one-million-fold variation in dynamic range in the concentrations of different RNA transcripts[3].

"The high quality of the functional information produced is evident from the exquisite detail and accuracy achieved."

But beyond these individual methods for data processing, the profound biological insights of ENCODE undoubtedly come from computational approaches that integrated multiple data types. For example, by combining data on DNA methylation, DNA accessibility and transcription-factor expression. Thurman *et al.*[4] provide fascinating insight into the causal role of DNA methylation in gene silencing. They find that transcription-factor binding sites are, on average, less frequently methylated in cell types that express those transcription factors, suggesting that binding-site methylation often results from a passive mechanism that methylates sites not bound by transcription factors.

Despite the extensive functional information provided by ENCODE, we are still far from the ultimate goal of understanding the function of the genome in every cell of every person, and across time within the same person. Even if the throughput rate of the ENCODE profiling methods increases dramatically, it is clear that brute-force measurement of this vast space is not feasible. Rather, we must move on from descriptive and correlative computational analyses, and work towards deriving quantitative models that integrate the relevant protein, RNA and chromatin components. We must then describe how these components interact with each other, how they bind the genome and how these binding events regulate transcription.

If successful, such models will be able to predict the genome's function at times and in settings that have not been directly measured. By allowing us to determine which assumptions regarding the physical interactions of the system lead to models that better explain measured patterns, the ENCODE data provide an invaluable opportunity to address this next immense computational challenge. ■

Eran Segal *is in the Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel.*
e-mail: eran.segal@weizmann.ac.il

1. The ENCODE Project Consortium *Science* **306,** 636–640 (2004).
2. The ENCODE Project Consortium *Nature* **489,** 57–74 (2012).
3. Djebali, S. *et al. Nature* **489,** 101–108 (2012).
4. Thurman, R. E. *et al. Nature* **489,** 75–82 (2012).
5. Neph, S. *et al. Nature* **489,** 83–90 (2012).
6. Gerstein, M. B. *et al. Nature* **489,** 91–100 (2012).
7. Sanyal, A., Lajoie, B., Jain, G. & Dekker, J. *Nature* **489,** 109–113 (2012).
8. Birney, E. *et al. Nature* **447,** 799–816 (2007).
9. Mardis, E. R. *Nature* **470,** 198–203 (2011).
10. Gonzaga-Jauregui, C., Lupski, J. R. & Gibbs, R. A. *Annu. Rev. Med.* **63,** 35–61 (2012).
11. Sproul, D. *et al. Proc. Natl Acad. Sci. USA* **108,** 4364–4369 (2011).
12. Carroll, S. B. *Cell* **134,** 25–36 (2008).
13. Prabhakar, S. *et al. Science* **321,** 1346–1350 (2008).
14. Pritchard, J. K., Pickrell, J. K. & Coop, G. *Curr. Biol.* **20,** R208–R215 (2010).
15. Lander, E. S. *et al. Nature* **409,** 860–921 (2001).

Essays in Biochemistry (2018) **62** 487–500
https://doi.org/10.1042/EBC20180003

PORTLAND PRESS

Review Article

Systems biology primer: the basic methods and approaches

Iman Tavassoly, Joseph Goldfarb and Ravi Iyengar

Department of Pharmacological Sciences and Systems Biology Center New York, Icahn School of Medicine at Mount Sinai, New York, NY 10029, U.S.A.

Correspondence: Ravi Iyengar (ravi.iyengar@mssm.edu)

Systems biology is an integrative discipline connecting the molecular components within a single biological scale and also among different scales (e.g. cells, tissues and organ systems) to physiological functions and organismal phenotypes through quantitative reasoning, computational models and high-throughput experimental technologies. Systems biology uses a wide range of quantitative experimental and computational methodologies to decode information flow from genes, proteins and other subcellular components of signaling, regulatory and functional pathways to control cell, tissue, organ and organismal level functions. The computational methods used in systems biology provide systems-level insights to understand interactions and dynamics at various scales, within cells, tissues, organs and organisms. In recent years, the systems biology framework has enabled research in quantitative and systems pharmacology and precision medicine for complex diseases. Here, we present a brief overview of current experimental and computational methods used in systems biology.

Introduction

In recent decades, our knowledge of the foundation of living organisms in terms of various components of cells, tissues and organ systems has been greatly expanded due to advances in technologies for high-throughput measurements such as genomics, transcriptomics, proteomics and metabolomics. In genetics and genomics, entire genomes of many organisms have been sequenced and the gene expression profiles comprehensively mapped. In biochemistry, mass spectrometry-based protein surveys have provided extensive lists of proteins and protein complexes, while molecular and cell biology have provided information on how proteins are organized to orchestrate the functions of subcellular systems such as cell organelles and cellular machinery components. Physiology has shed light on the complex functions of cells, tissues and organ systems. This enormous amount of information at different scales of organization can be used to obtain a new perspective that starts from genes and proteins, moves through subcellular interactions and pathways and ends in the physiology of cells, tissues and organ systems [1-4]. The availability of such multiscale information has catalyzed the formation of systems biology as a discipline in biomedical sciences. Systems biology is the study of molecular interactions at different levels, enabling the identification and description of the subcellular machinery that makes functional operational units in cells, tissues and organ systems resulting in physiological behaviors [5,6].

Historically, systems biology started by looking at cells, tissues and organ systems as complex biological systems [7]. The rapid development of genomics and sequencing technologies led to the uncovering of big datasets of basic components forming these complex systems [8,9]. Later, it was shown how interactions among molecular components of cells could give rise to functional behaviors that single components by themselves cannot [10-12]. One way to think of systems biology is that it provides a new and broader perspective of physiology. While physiology provides a description of functions in cells, tissues and organ systems using largely phenomenological approaches, systems biology integrates molecular biology and

Received: 19 July 2018
Revised: 22 August 2018
Accepted: 24 August 2018

Version of Record published:
04 October 2018

Essays in Biochemistry (2018) **62** 487–500
https://doi.org/10.1042/EBC20180003

biochemistry of molecular components and their interactions and dynamics to understand how physiological functions arise and are controlled [1,13,14]. Systems biology integrates not only the molecular entities at a specific scale but also the connections among these molecular components at different scales. Integration of data is the core value in systems biology, in which the interactions of multiple components are treated as a single system. This integration can be applied at a single scale (e.g. the cellular level) to provide new systems-level insight, but also can be used to decode complex phenotypes at different scales. For example, systems biology is used to study the evolution of a cancer cell from a normal cell. This involves interactions among molecular components at the cell level. At the same time, systems biology can be used to integrate the interactions among cancer cells and the evolution of tumors. It is also capable of describing the interaction of different tissues such as blood vessels, tumors and the immune system to shed light on complex phenomena of cancer at the organ level [15-20].

Biological systems are multiscale, with multiple levels of organization and with multiple states at different times, and hence, systems-level analyses are particularly useful. Differences in scale of biological systems can be studied from molecular components to subcellular machinery (such as transcriptional and translational control machinery and cell motility machinery) and to cells, tissues, organ systems and whole organisms. In this systems-level view, as the organizational level of a system increases, it leads to new characteristics and capabilities [1,20]. Multiscale systems can be studied in two major ways: bottom-up and top-down. Both approaches have their advantages and disadvantages.

In a bottom-up approach, cellular and molecular components are studied as parts of a system that includes their interactions and dynamics leading to physiological functions. This approach has the ability to provide mechanistic insights into how different units work together to form a system. In this approach, however, as the system becomes bigger, the details may obscure the overall capabilities of the system [21]. In contrast, in the top-down approach, the system as a whole is studied, and the characteristics and potential capabilities of the system are discovered. This gives a big picture of the system, which can be comprehensive and integrative. In this approach, interactions among different units are often defined by correlation and the complexity of the biological systems often does not always allow one to make causal inferences [21]. The different experimental methods and computational approaches are summarized in Figure 1.

Genomic-wide analyses of single nucleotide polymorphisms, comprehensive transcriptomic profiling and deep proteomics that provide an extensive characterization of cellular proteins are all examples of top-down surveys that correlate molecular components with cellular, tissue or organismal level phenotypes. Although such relationships are often correlative, they can provide useful bookends for more mechanistic systems-level characterizations. In both bottom-up and top-down approaches, there are two main sets of tools: experimental tools and computational tools. Experimental studies in systems biology often start with omics, high-throughput technologies including genomics, transcriptomics, epigenomics, proteomics and metabolomics [2]. Such large datasets are analyzed by use of statistical models as well as graph theory-based models. In bottom-up approaches, low-throughput, but high fidelity experiments can provide a foundation for verification of predictions from computational models both qualitatively and quantitatively [22,23] (Figure 1).

One can also use a middle-out approach in systems biology, studying a higher level function by selecting only a limited number of lower level interactions deemed to be relevant to a specific phenotype. This approach considers modularity in systems biology and uses an approach like engineering methods that use only selected functionally vital components to build and understand a processing circuit or machine [24,25].

Systems level experimental analysis of cells

The systems-level analysis of cells requires information on all of the individual entities at different levels of cell function. Omics technologies provide such information and, in the process, yield vast amounts of data from genes, mRNAs, proteins and metabolites. These high-throughput methods measure many individual subcellular components that act as a system to control cell function. Genomics, which utilizes sequencing technologies and microarrays, can determine the sequence of genomes and characterize genomic determinants including single nucleotide polymorphisms (SNPs), indels and epigenetic regulatory sites (such as DNA methylation sites), affecting a specific phenotype or function in cells or organisms [26]. Transcriptomics measures the transcriptome of cells or tissues that consists of all RNA transcripts [27]. Epigenomics describes all epigenetic modifications such as DNA methylation and histone modifications in cells [28,29]. Proteomics, which often uses mass-spectrometry technologies, measures and catalogs proteins and post-translational modifications at a large scale [30]. Metabolomics is the large-scale study of metabolites in cells and tissues and uses liquid chromatography, mass-spectrometry and NMR technologies [31]. The information gained by such systems-wide surveys needs to be processed and organized to turn data into knowledge. The organizing and analyzing of large datasets are called Bioinformatics. Currently, there are many databases that store,

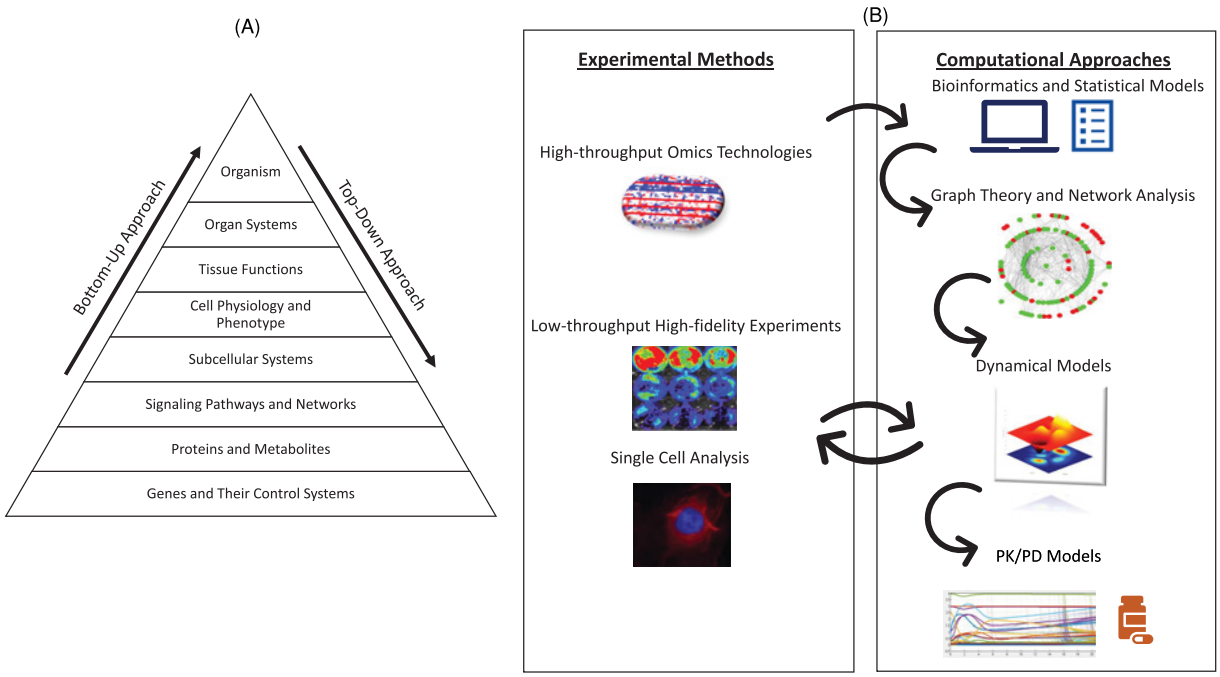


Figure 1. Systems biology approaches and methodologies

(**A**) Systems biology methodologies can be applied either in a bottom-up approach that puts small functional units together to make a system or in a top-down approach that starts from the global view of the system and then tries to study smaller subsystems. (**B**) Systems biology utilizes both experimental and computational frameworks to answer biological questions. Omics technology provides a platform to extract knowledge using bioinformatics, statistical methods and network analysis. The dynamical models of certain components in these networks must be verified by low-throughput, high-fidelity and single cell experiments that provide new strategies to improve and optimize the dynamical models. Dynamical models can be merged with PK/PD models to analyze therapeutic efficacies and design precision drug treatments.

and computational tools to analyze, these data, such as genomic characteristics including SNP profiles for diseases, mRNA profiles, protein networks etc. Systems biology integrates experiments and computational models to understand how systems function. Computation is a key feature that characterizes systems biology compared with classic biological disciplines such as biochemistry and cell biology. A good example of the use of large molecular datasets is Genome-wide Association Studies (GWAS), which is the process of finding variations in DNA sequence, usually SNPs, associated with increased risk of a specific disease or physiological state. GWAS is a useful map by which genomic data can be correlated with pathophysiological states. It can also contribute to understanding drug action and the discovery of new drug targets by evaluating genetic variations in response to drugs, and to progression of disease [32,33].

Qualitative methods include most of the omics technologies that produce large-scale, often comprehensive, lists of molecular components. Transcriptomics focuses on identifying all the mRNAs on a genome wide basis. As the cost of sequencing has come down dramatically in the past few years, transcriptomics measurements have moved from the use of microarray chips to sequencing methods [34]. Proteomics focuses on identifying proteins and their post-translational modifications using mass spectrometry [35]. Advances in computational identification of proteins from mass spectrometry data now allow for the identification of ∼10,000 proteins per cell type [36]. Metabolomics uses mass spectrometry as well as NMR technologies to identify metabolites and track metabolic pathways [37]. Each of these omic technologies has advanced detailed experimental methods as well as specific informatics tools for transcriptomics [38], proteomics [39] and metabolomics [40]. The informatics tools are needed to analyze the large datasets to produce ranked lists of molecular entities that can be cast as pathways and networks to infer function.

From molecules to pathways and networks

Experimental omics studies produce large molecular datasets. Statistical methods are required to generate ranked lists of those molecular components (genes, mRNA, proteins etc.) involved in specific physiological or pathophysiological

Essays in Biochemistry (2018) **62** 487–500
https://doi.org/10.1042/EBC20180003

states. Gene Set Enrichment Analysis (GSEA) is a statistical method to find potential molecular components responsible for phenotypes and functions based on those entities that are under- or over-represented in biological samples. The differentially expressed molecular entities (or, in general, differentially expressed biomarkers) are enriched using a specific ontology. An ontology is a set of structured terms with specific relationships that work like a classifier with hierarchical structure [41,42]. The ontology is a tool to find biological knowledge by association of data (genes or gene products) with biological processes, molecular functions and cellular components [41,42]. Several ontologies have been developed and used in systems biology including Gene Ontology (GO) and Molecular Biology of the Cell Ontology (MBCO). In addition, there are other bioinformatics tools such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), Wikipathways, Reactome Pathway, Progeny Signatures and Broad Signatures to transform data into biological knowledge [42-48]. The results from GSEA yield knowledge about the pathways, including signaling pathways regulating the specific phenotype being studied.

Signaling pathways are the main systems that process information in cells. Signaling pathways receive signals from outside the cell and control cellular physiology in response to these signals. These pathways have many components each of which receives, transmits and transduces information to other components [13,14,23,49]. The flow of information, in the form of cellular signals, occurs in time and space and can be studied mathematically using dynamical systems theory and differential equations [14,16]. Receptors, which receive signals from outside the cell, and other intracellular signaling components, enable connectivity between signaling pathways within a network. The intracellular signaling components are information processing units, signal integrators and effectors that function as output devices that represent the cellular responses to extracellular signals [11,16].

In addition to linear pathways, GESA enables the construction and analysis of functional molecular networks. Networks are formed by interactions between molecular entities. These entities are called 'nodes' and the interactions between the entities are called 'edges'. Such interactions include direct binding leading to activation or inhibition of the downstream target and enzymatic activities [50-53].

Analysis of biological networks

A network is a set of nodes connected to each other via edges and mathematically defined as a graph. The structure and function of networks are studied by graph theory. Networks can be studied as computational units and systems, which provide insights into both their organization and functions [50,52,53]. In systems biology, the network nodes are cellular components and edges are reactions or interactions among these nodes. Viewing cell systems as networks is a helpful and practical way of understanding the functional organization of cells by analyzing network topology [10,18,50,52]. In cellular networks, there are cases when the relationships among nodes are conditional rather than fixed. Those networks where edges are defined in a probabilistic manner are called Bayesian networks [54]. Bayesian networks allow us to discover probabilistic relationships among molecular components and define the conditions that increase or decrease the probability of the relationships [55,56]. Networks can be represented as directed or undirected graphs. Undirected graphs represent the relationship among nodes without specifying hierarchy and are usually constructed from high-throughput large datasets Directed graphs represent not only the relationship among nodes but also the direction of signal propagation and hierarchy such as an upstream node regulating a downstream node. For example, in a directed graph of protein networks, inhibition or activation of a protein by another protein can be shown. There are many software packages and tools that enable the visualization of networks [57]. Visualization and analysis of cellular networks give a perspective on global organization of cell systems and help in identifying the key nodes in terms of connectivity. One of the properties of each network is the degree distribution, which is the probability distribution of all degrees of nodes within a network. The degree of a node is the number of edges via which it is connected to other nodes. A node with a degree much higher than average for the network is called a hub [58]. Hubs are not observed in random networks. Networks of real systems, such as cellular signaling systems, are organized differently from random networks. Real networks have a degree distribution that follows the power law and are called scale-free networks [59]. The robustness of a network and its sensitivity to perturbations are other properties of molecular networks that affect the functions of the system [50,60]. Perturbation, in terms of removing some of the nodes and measuring the resistance of the network to change, can be used to evaluate the robustness of a network. Scale-free networks are highly robust to random removal of nodes as there are few highly connected nodes [59]. These networks, however, are fragile to the specific removal of hubs [50,59,61].

Network dynamics

Decoding signal propagation and processing in molecular networks requires consideration of the temporal aspects of signal processing [13]. Network-based models have limited capabilities to capture temporal dynamics of the system,

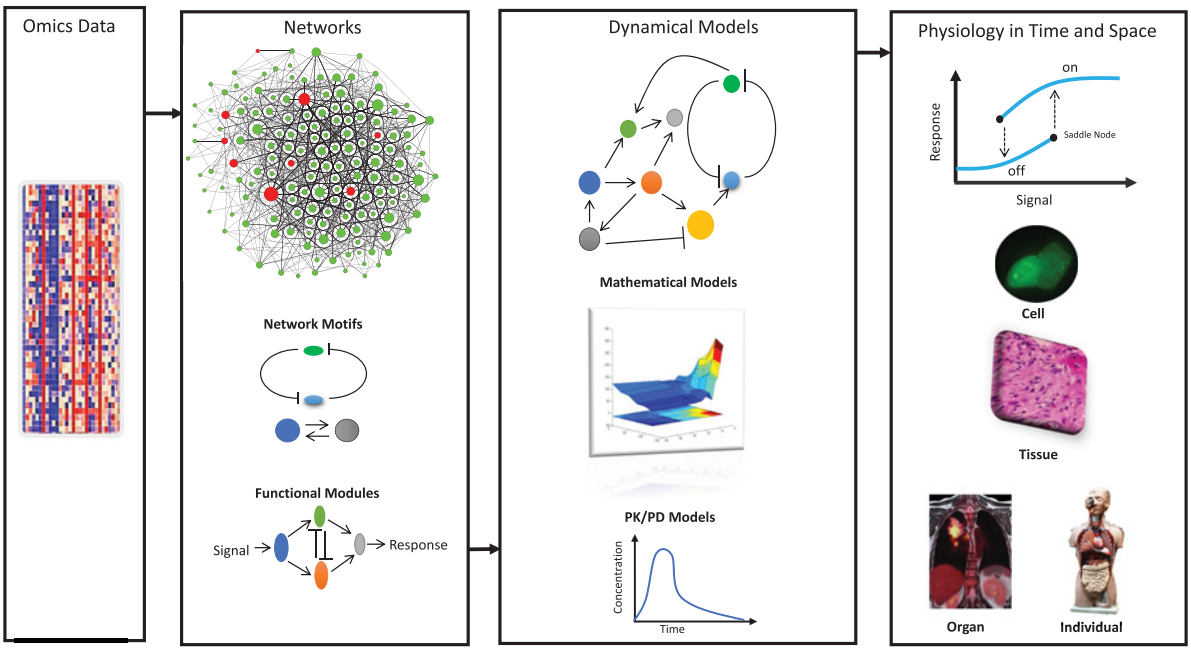


Figure 2. Computational methods in systems biology

Omics data are organized and analyzed using bioinformatics tools, and the resulting datasets are used to build networks. Within these networks of molecular interactions, the topological features of cellular wiring systems can be deduced. Network motifs and functional modules that are smaller sets of nodes and edges are commonly found in these networks and represent certain dynamical signal processing properties and carry out specific functional tasks. From these networks, sets of regulatory pathways (which include motifs, modules and feedback loops) are extracted to build dynamical models. These dynamical models are used for simulations to understand and predict the emergent behavior of the system in space and time. They also can be merged with PK/PD models to study drug action.

but temporal dynamics is essential for understanding systems behaviors at the cell and tissue level. Hence, network analysis needs to be combined with dynamic quantitative mathematical models. Dynamical models present a more accurate description of how a system progresses temporally and spatially. In fact, the networks within cells not only underlie structural and organizational aspects of cellular components but also can show emergent temporal properties defining cellular functions. Each network includes some functional modules that have a limited number of components that interact to receive a signal, to process it and then to transduce the signal to other modules. A network has a dynamic that can be studied by translating its components (nodes and edges) into a set of ordinary differential equations (ODEs) [29,30] (Figure 2). A simple way of translating a network to a mathematical formalism to study its dynamics is by using Boolean logic, which assigns a state of being 'on' (1) or 'off' (0) to each node. In addition to Boolean logic models and differential equation-based models, there are hybrid models, which use a combination of Boolean functions and differential equations, and fuzzy logic-based models that, in contrast with Boolean logic, represent nodal activity values between 0 and 1 [29,44-47].

Network motifs and functional modules

A motif is a set of a limited number of components that represent a certain dynamical behavior such as bistability or oscillations [49,62]. For instance, one of the motifs common in signaling pathways is mutual inhibition between two proteins. The structure of such a motif shows the emergent ability of this system in signal processing to make an on/off memory switch [10,62,63]. Network motifs, such as feedback and feedforward loops and bifan motifs, are recurrent and commonly found as subgraphs in biological networks [62,64,65]. A functional module consists of one or more such motifs with a particular function, such as signal integration or switch control in cells [49,62]. By extracting network motifs in molecular networks and surveying the dynamical behavior of functional modules, it is possible to decode temporal characteristics of signal transduction in cell systems. Methods from dynamical systems are used to

Essays in Biochemistry (2018) **62** 487–500
https://doi.org/10.1042/EBC20180003

analyze ODE-based mathematical models of motifs, modules and networks [11]. The dynamical systems methods also provide a roadmap to design experiments for verifying the dynamical models, as the rate of molecular events in cells follow rules of dynamical systems [66] (Figure 2).

Dynamical models

Dynamical models are built by converting a network of interactions such as a gene regulatory network or a protein–protein interaction network to ODEs. Solving and analyzing these ODEs show the qualitative and time-course changes in the network as a dynamical system [67]. Often dynamical models do not have a unique solution, as defined by a single set of parameters. Such dynamical models are considered robust with more than one set of parameters and have a spectrum of parameter sensitivity [49,68]. Exploring these parameter spaces provides new information on the biological redundancies built into the system [49]. When the signal processing is done in different cellular compartments, compartmental dynamical models are built in which biochemical reactions within a compartment are represented as groups of ODEs [69]. These dynamical models can provide information regarding the state of the system. Changes between system states can provide knowledge about different types of activities a system is capable of. Such states can be at the cellular or tissue/organ or organismal levels. Bifurcation theory is a tool used to study states of dynamical systems that undergo qualitative or topological changes. For example, dynamical patterns such as bistability (switching between two stable states) and oscillations can be studied using saddle–node bifurcation or Hopf bifurcation [16,70]. Bifurcation analysis is a mathematically and computationally challenging task when the systems of ODEs become complex and is often used to study functional modules such as feedback loops [62].

When modeling cellular processes in time and space to understand the spatial organization of time-dependent cellular functions, partial differential equations (PDEs) are used [71,72]. PDEs can compute transitions in concentration and change in location of reactants and products. Solving PDEs is more challenging than ODEs because adding spatial parameters increases the complexity of the equations, and in PDEs one deals with multivariable functions in contrast with ODEs where the functions of a single variable are considered [73].

Solving ODEs in dynamical modeling can be done analytically or, more commonly, numerically. An analytical solution is expressed as a mathematical formalism that can readily be used to simulate time-courses of different components. Numerical solutions are based on obtaining numerical approximations for ODEs of the systems being studied. ODEs representing cellular and biological systems are usually very complex and cannot be solved analytically. They are most often solved numerically using different software packages and tools such as MATLAB, COPASI, Virtual Cell etc. [74-76]. PDE are typically solved numerically. Both MATLAB and Virtual Cell have PDE solvers.

If a system's temporal evolution is fully determined by specific initial conditions and reaction rates, then it can be modeled by a deterministic ODE or PDE model. However, many important cellular processes, such as gene expression, are stochastic, and modeling them requires stochastic modeling in contrast with deterministic ODE or PDE models. Heterogeneity is a main characteristic at all levels of biological systems. One way to include the heterogeneity of these systems in terms of probability distributions of intrinsic and extrinsic noise is stochastic modeling [77]. A stochastic dynamical model describes systems or functions in which the temporal evolution of the system is computed both by specific predictable reactions and some random variables and parameters. A common methodology for stochastic systems is the Gillespie algorithm [78]. In stochastic models, a master equation is implemented to control the evolution of the system such that a probabilistic function defines the next state of the system. The master equation basically defines the probabilistic distribution of all possible states that the system can have over time. The Gillespie algorithm makes it possible to simulate each bimolecular reaction while time or space intervals between reactions adhere to a probability distribution defined by the master equation [77-79].

Another aspect of dynamical models in systems biology is linking a dynamical model built for a single cell to the behavior of a population of cells, such as within a tumor [80]. In these cases, each cell can have a distinct parameter space with some parameters following probabilistic distributions. In such cases each cell may be simulated separately, and the behavior of the population computed from the average behavior of individual cells. [49,78,80,81]. Depending on the biological questions one wants to answer, the type of mathematical model chosen is deterministic or stochastic. The parameters in dynamical systems of cellular processes and signaling pathways need to be measured directly from experiments or estimated based on experiments. Although there are toy dynamical models that are built using arbitrary parameters, which are helpful to gain mechanistic insights into the system, the most common dynamical models in systems biology are plausible models in which parameters are measured or estimated by experiments. Identifiable dynamical models are made to explain the experimental data, and variables and parameters are specific to a certain system and fitted to experimental data from that system. These models are very common in quantitative systems pharmacology (QSP) and studies that involve drug actions [20,63,67,82]. In all dynamical models of cell

systems, thermodynamic constraints must be fulfilled [13,83]. These dynamical models allow one to study and predict physiological responses in space and time (Figure 2).

Pharmacokinetic/pharmacodynamic (PK/PD) models are commonly used in the study of drug action. Pharmacokinetic models are focused on drug disposition and availability whereas pharmacodynamics focuses on mechanisms of drug action. Combining PK/PD models with dynamical models of cellular regulatory systems can be used for predicting both therapeutic and adverse effects of drugs [20].

Strengths and limitations of different types of models

The different modeling approaches in systems biology have their own applications and limitations. They are chosen based on the system under study and the complexity of the problem being addressed, and the use of multiple models may be necessary to predict system behavior. When using high-throughput and quantitative experimental approaches, model types used include statistical models, networks and dynamical models. Statistical models, which are the first layer in top-down systems biology, deal with defining molecular datasets assigned to given phenotypes and functions. These models can deal with probabilistic relationships built upon correlations. This makes statistical models useful for clinical decision making because for most complex diseases, pathological phenotypes are associated with molecular markers like genes in a probabilistic manner. These models, however, do not enable the understanding of mechanisms underlying the development of phenotypes because they do not consider the nature and direction of interactions among components [84,85]. They cannot decode information flow from pathways or the dynamics of networks within the cells. Statistical models have a static view of biological functions, and systems evolving in time and space are not fully described. For example, statistical models are inadequate to describe the time-course of initiation of a disease phenotype or acquisition of treatment resistance. Mechanistic models are required to describe an integrative view of the pathological process [20,86]. Network-based models serve as representations of whole-cell interactions and their topologies. These topologies are a vital first step to understand the dynamics of cell systems in a flexible multiscale fashion. They represent all cellular components and their relationships as a global map for information transmission in cells, tissues and organs. Inside these networks, it is possible to search for functional modules by identifying hubs and network motifs. To truly understand computation within cells, we require both network models and dynamical models. However, lack of sufficient kinetic data often prevents us from building dynamical models at the level of large networks. We usually need to select the most important components, including functional modules and computational units, to make insightful and realistic dynamical models [16,51,85,87]. In addition, the assumptions and estimated parameters needed for the construction of dynamical models require that predictions from model simulations be experimentally verified.

Quantitative experimental methods for systems biology

Quantitative methods encompass a wide range of experiments that measure the quantity of cellular components such as protein concentrations and their temporal changes in different time scales. These include standard molecular biology and cell biology experiments as well as high-throughput experiments. These experiments can be based on a single cell or cell populations. Single-cell experiments are helpful to verify and explore parameter spaces of models designed at the cell level. Specifically, when a cell population is heterogeneous (for example cancer cells), each cell may have a different parameter space and responses to signals [22,23,81]. Over the past few years single-cell transcriptomics [88-90] has been developed to provide mRNA profiles in single cells. This approach has been very useful in mapping subtypes of classes of cells within tissues and organs. Conventional molecular biology experimental methods, such as Western blots for measuring protein concentrations, provide an average result from many cells in an often heterogeneous cell population [49]. Although both single cell and population experiments can be used, the ergodic nature of cellular events favors measuring single cell dynamics from a cell population [90].

Often it is not possible to measure the concentrations and kinetic parameters of all components of a system. Thus, some component parameters used in models are estimated based on data from other components. Finding kinetic parameters is often difficult due to limitations imposed by experimental design. Quantitative measurements, such as time-course experiments, involve many components with different kinetic parameters making it difficult to explicitly measure the kinetic parameters associated with individual molecular components. Quantitative characterization of molecular components, both with respect to kinetic parameters and concentrations within different cell types is an underdeveloped area of study.

One type of experiment helpful for building precise networks and models is using omics technologies at the single-cell level. Conventional omics methods provide a list of entities from a heterogeneous cell population. However, in single-cell transcriptomics, the mRNA concentrations of expressed genes are measured in each cell in a population

Essays in Biochemistry (2018) **62** 487–500
https://doi.org/10.1042/EBC20180003

of cells. Although the number of genes identified by this method is ∼1000 per cell, the method of measurement, using 3′ unique molecular identifiers, counts each molecule of RNA and hence provides quantitative estimates of the different RNA species in each cell. These single-cell omics data are useful in describing the heterogeneity of cells in tissues and organs. Heterogeneity is an important consideration for building predictive models for complex tissues and diseases because the phenotypes are dependent on cells with different identities [91-93]. An example is the systems biology of cancer, in which both statistical and dynamical models are built to design therapeutic regimens for tumors and cell lines that contain many individual cells with heterogeneous expressions of genes and proteins [88]. Molecular information from single cells can be used to build models of cell populations by considering single cells with different identities as components, with each cell considered as a system of the biochemical and molecular network. Such an approach captures the diversity of cell subtypes in a tissue or organ system.

Artificial intelligence in systems biology

One of the main challenges in systems biology is to convert big data at different scales into actionable knowledge. This knowledge is vital to improve methodologies to study biological systems, to understand and diagnose diseases at various stages precisely and to design new therapeutic modalities focused on the individual. Mechanistic models, such as dynamical models that depend on the causality of relationships among components, can combine biological data from hypothesis-based experiments with mathematical modeling to produce predictive models. Often, such models also provide insight into mechanisms. The amount of information in biology and medicine is rapidly surpassing the current capability of building large-scale mechanistic models. An alternative way to generate predictive models from big data is through statistical models based on correlation. This process can benefit from artificial intelligence (AI) that uses statistical reasoning to detect unseen correlation, co-occurrence and dependencies in large-scale datasets [94].

In computer science, AI is a way of developing machine-based expert systems that can analyze data and predict new outcomes. Machine learning, deep learning and artificial neural networks are different approaches used in AI. Artificial neural networks were inspired by real brain neural networks and are capable of learning specific task-oriented classifications when trained by a training set [95]. Machine learning refers to a group of methods that analyze big datasets and, based on them, make predictions. Machine learning can be used in a supervised, unsupervised or semi-supervised manner. In supervised machine learning, training datasets in the form of labeled input/output relations are provided and a function is inferred that can be used to analyze new examples—to predict the output based on input data and classification. Unsupervised machine learning is when the data are not labeled, and the aim is to detect underlying patterns with no guide. Semi-supervised learning is a modality between supervised and unsupervised learning when there is limited labeled data [96-98]. Deep learning is a machine learning method that uses multilayer computational processing units for data representation and detection of intrinsic patterns in big data [99,100].

AI is a powerful tool for developing models and optimizing them. In AI, an algorithm and a dataset are used. The dataset usually has two properties, one is a large set of measurements (e.g. molecular signatures such as genes, proteins or metabolites) and the other is the resultant prediction (e.g. the resulting phenotype). The underlying algorithm, usually a statistical model, is trained using the dataset and then a test set is used to evaluate the predictions (Figure 3) [96,97,101]. This training and testing procedure can be optimized as a loop by using new data as training datasets. The algorithm adjusts itself to make better predictions, and thus the AI system learns as it is being used. Network building and analyses can benefit from AI methods to extract data from omics data in terms of finding interactions and relationships among molecular entities in given phenotypes, finding network motifs and functional modules, and decoding main pathways involving selected functions (Figure 3) [54,102]. Exploring the parameter spaces of dynamical models and sensitivity analysis also use machine learning approaches because they deal with big numerical datasets [103]. The predictive models arising from AI and machine learning are potentially powerful tools in precision medicine as they can extract genomic signatures related to drug treatment and therapeutic responses. In such cases, both large clinical and biomolecular datasets are used as training sets that are assigned to specific responses. The predictions of these models are tested and valid predictions are used as new data for making the training set bigger [101]. AI, for example, has been successfully used in predicting cancer outcomes based on molecular biomarkers and pathology [104]. The use of artificial intelligence in studies of basic biological systems, as well as in clinical data, are schematically shown in Figure 3.

Systems pharmacology and systems biomedicine

Systems-level insights serve as building blocks to advance medicine to a higher level of personalized and precision care. Currently, systems biology approaches are being implemented in both drug discovery and the practice of

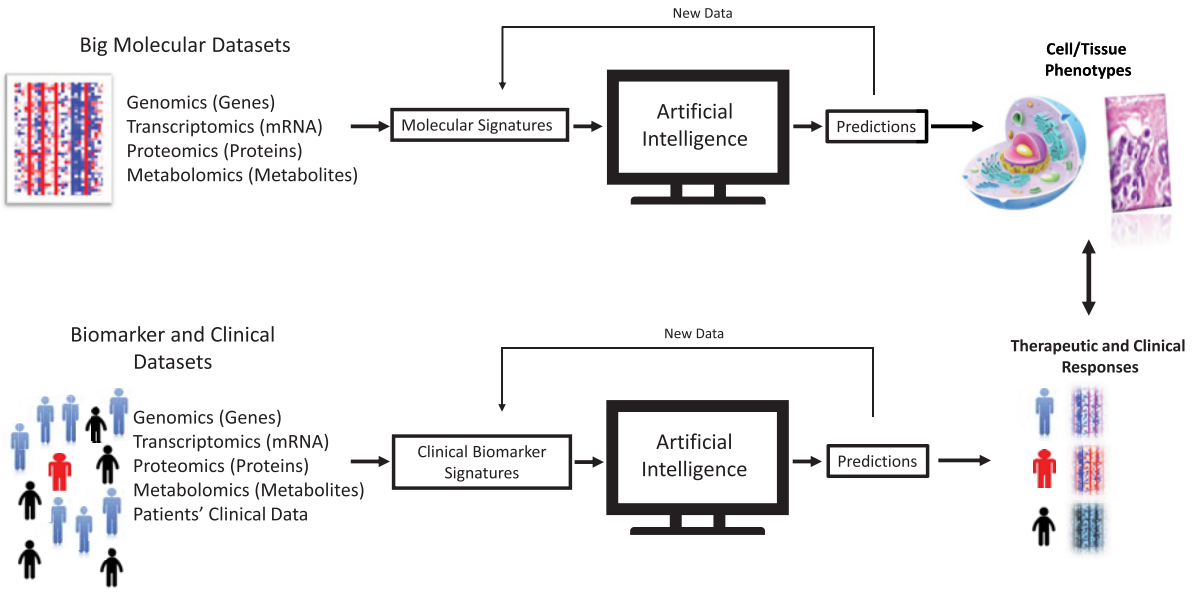


Figure 3. AI in systems biology and precision medicine

AI and machine learning approaches are helpful tools for finding molecular signatures from high-throughput measurement datasets and assigning them to higher level phenotypes and cellular functions. They also can be used to tailor treatment based on an individual patient's molecular markers for precision medicine. Clinical data and omics data from patients are used to extract clinical biomarkers for training sets and to build predictive models of the course of the disease and patient responses to treatment. Verified predictions are used as additional training sets to make the predictive models progressively more accurate.

medicine. Systems biology has enabled pharmacology to become a systems science. Systems pharmacology, of which pharmacogenomics is a part, has been shown to be useful for drug discovery and predicting therapeutic responses and drug adverse effects. While pharmacogenomics uses genomic data of drug metabolizing enzymes for prediction of drug responses and effects, QSP utilizes network and dynamical models integrated with pharmacodynamics and pharmacokinetics to find optimized therapeutics for specific patients with a given disease [20,85,105,106]. These advances enable adjustment of drug regimens and drug doses for individual patients based on their molecular markers [85,107].

With advances in biosensors that are able to collect time-course data from patients, liquid biopsies and biomarker discovery, the practice of precision medicine based on systems biology approaches seems feasible. These data collection tools provide the basic materials for predictive models using systems biology approaches. Biosensors can collect real-time data on the concentration of different components in the blood of a patient or record quantitative data on physiological signals such as heart rate and electrical activities of brain and heart [108,109]. Liquid biopsy collects cancer cells or other tissue components in fluids such as blood, saliva and urine that can be used for omics data and biomarker detection [110]. Advances in high-content image analysis that requires quantitative analysis of vast numbers of images such as pathology slides can, with the help of machine learning, provide an accurate diagnostic tool for predictive models of disease states and progression [111,112].

Genomic signatures in systems therapeutics

One of the recent advances in the field of precision medicine is the discovery of genomic signatures related to pathogenesis and therapeutic responses in different diseases [107,113-115]. Drug treatments change the gene expression profiles of cells, and measuring these changes before and after treatment *in vitro*, *in vivo* in animal models, and in patients, can bring new insights about genomic determinants of drug responses and drug adverse effects. Genomic signatures also can be used as prognostic markers for patients suffering from chronic diseases such as cancer and help in selecting individuals for specific treatment plans. For example, in the case of cancer immunotherapy, there has been

Essays in Biochemistry (2018) **62** 487–500
https://doi.org/10.1042/EBC20180003

a major effort to detect the genomic determinates of responses to immunotherapy agents such as PD-L1 inhibitors. Currently, the main practice is based on the expression of PD-L1 protein in tumor tissue, but several genomic and clinical markers, both tumor genomic profiles and patients' immune system characteristics, have been found useful in guiding immunotherapy [116].

Perspective

By investigating qualitative as well as quantitative properties, both temporal and spatial, and emerging functions of molecular interactions in biological systems, we are able to understand many phenomena in cells, tissues/organs, and at the level of whole organisms. The transmission of information from genes to organismal behaviors, and complex phenotypes arising from molecular and cellular networks, can be explored using systems biology methodologies. Statistical, network and dynamical models are essential tools in systems biology leading to discoveries at various scales of biological organization. These discoveries are basic building blocks for future advances in medicine, leading to precision and individualization of treatment. Advances in computational and experimental methods, including faster and more accurate technologies, will enable systems biology to provide basic understanding of cells, tissues and organs, as well as future medical advances.

Summary

- Systems biology studies cells, tissues and organ systems as systems of interacting components.
- Omics technologies are the main sources of information on individual molecular entities in cells.
- Bioinformatics organizes the big data obtained from systems-wide surveys.
- Statistical methods enable analyses of big datasets based on high-throughput technologies that can then be used to decode pathways and molecular networks.
- Dynamical models of networks of cellular components help to explain the emergent properties of cell and tissue physiology in time and space.
- Systems biology methods can be used for drug discovery and development of systems pharmacology approaches.
- Artificial intelligence and machine learning approaches are used as tools in systems biology to link molecular datasets to phenotypes and physiological behaviors at the organismal level.
- Insights from systems biology studies are useful for the design of precision and individualized medicine protocols.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

Work leading to this review was funded by Systems Biology Center grant [GM-071558] from the National Institute of General Medical Sciences (NIGMS).

Abbreviations

AI, artificial intelligence; GO, Gene Ontology; GSEA, Gene Set Enrichment Analysis; GWAS, Genome-Wide Association Studies; KEGG, Kyoto Encyclopedia of Genes and Genomes; MBCO, Molecular Biology of the Cell Ontology; ODE, ordinary differential equation; PDE, partial differential equation; QSP, quantitative systems pharmacology; SNP, single nucleotide polymorphisms.

References

1 Kitano, H. (2002) Systems biology: a brief overview. *Science* **295**, 1662–1664, https://doi.org/10.1126/science.1069492
2 Hood, L., Heath, J.R., Phelps, M.E. and Lin, B. (2004) Systems biology and new technologies enable predictive and preventative medicine. *Science* **306**, 640–643, https://doi.org/10.1126/science.1104635
3 Wist, A.D., Berger, S.I. and Iyengar, R. (2009) Systems pharmacology and genome medicine: a future perspective. *Genome Med.* **1**, 11, https://doi.org/10.1186/gm11
4 Weng, G., Bhalla, U.S. and Iyengar, R. (1999) Complexity in biological signaling systems. *Science* **284**, 92–96, https://doi.org/10.1126/science.284.5411.92
5 Ideker, T., Galitski, T. and Hood, L. (2001) A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372, https://doi.org/10.1146/annurev.genom.2.1.343
6 Kirschner, M.W. (2005) The meaning of systems biology. *Cell* **121**, 503–504, https://doi.org/10.1016/j.cell.2005.05.005
7 Jensen, H.J. (1998) Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems. Cambridge University Press
8 Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K. et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934, https://doi.org/10.1126/science.292.5518.929
9 Consortium IHGS (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860, https://doi.org/10.1038/35057062
10 Bhalla, U.S. and Iyengar, R. (1999) Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387, https://doi.org/10.1126/science.283.5400.381
11 Tyson, J.J., Chen, K.C. and Novak, B. (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231, https://doi.org/10.1016/S0955-0674(03)00017-6
12 Alon, U., Surette, M.G., Barkai, N. and Leibler, S. (1999) Robustness in bacterial chemotaxis. *Nature* **397**, 168, https://doi.org/10.1038/16483
13 Tyson, J.J., Chen, K. and Novak, B. (2001) Network dynamics and cell physiology. *Nat. Rev. Mol. Cell Biol.* **2**, 908, https://doi.org/10.1038/35103078
14 Kholodenko, B.N. (2006) Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **7**, 165, https://doi.org/10.1038/nrm1838
15 Clarke, R., Shajahan, A.N., Wang, Y., Tyson, J.J., Riggins, R.B., Weiner, L.M. et al. (2011) Endoplasmic reticulum stress, the unfolded protein response, and gene network modeling in antiestrogen resistant breast cancer. *Hormone Mol. Biol. Clin. Invest.* **5**, 35–44, https://doi.org/10.1515/HMBCI.2010.073
16 Tyson, J.J., Baumann, W.T., Chen, C., Verdugo, A., Tavassoly, I., Wang, Y. et al. (2011) Dynamic modelling of oestrogen signalling and cell fate in breast cancer cells. *Nat. Rev. Cancer* **11**, 523–532, Epub 2011/06/16. eng, https://doi.org/10.1038/nrc3081
17 Clarke, R., Cook, K.L., Hu, R., Facey, C.O., Tavassoly, I., Schwartz, J.L. et al. (2012) Endoplasmic reticulum stress, the unfolded protein response, autophagy, and the integrated regulation of breast cancer cell fate. *Cancer Res.* **72**, 1321–1331
18 Kreeger, P.K. and Lauffenburger, D.A. (2009) Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**, 2–8, https://doi.org/10.1093/carcin/bgp261
19 Anderson, A.R. and Chaplain, M. (1998) Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bull. Math. Biol.* **60**, 857–899, https://doi.org/10.1006/bulm.1998.0042
20 Iyengar, R., Zhao, S., Chung, S.-W., Mager, D.E. and Gallo, J.M. (2012) Merging systems biology with pharmacodynamics. *Sci. Transl. Med.* **4**, P126ps7, https://doi.org/10.1126/scitranslmed.3003563
21 Bruggeman, F.J. and Westerhoff, H.V. (2007) The nature of systems biology. *Trends Microbiol.* **15**, 45–50, https://doi.org/10.1016/j.tim.2006.11.003
22 Batchelor, E., Loewer, A. and Lahav, G. (2009) The ups and downs of p53: understanding protein dynamics in single cells. *Nat. Rev. Cancer* **9**, 371, https://doi.org/10.1038/nrc2604
23 Purvis, J.E. and Lahav, G. (2013) Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945–956, https://doi.org/10.1016/j.cell.2013.02.005
24 Noble, D. (2008) *The Music of Life: Biology Beyond Genes*, Oxford University Press
25 Walker, D.C. and Southgate, J. (2009) The virtual cell—a candidate co-ordinator for 'middle-out'modelling of biological systems. *Brief. Bioinform.* **10**, 450–461, https://doi.org/10.1093/bib/bbp010
26 Morozova, O. and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255–264, https://doi.org/10.1016/j.ygeno.2008.07.001
27 Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57, https://doi.org/10.1038/nrg2484
28 Epigenomics, E.V. (2006) Mapping the methylome. *Cell Cycle* **5**, 155–158, https://doi.org/10.4161/cc.5.2.2367
29 Jones, P.A. and Baylin, S.B. (2007) The epigenomics of cancer. *Cell* **128**, 683–692, https://doi.org/10.1016/j.cell.2007.01.029
30 Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N. et al. (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173, https://doi.org/10.1038/nature04209
31 German, J.B., Hammock, B.D. and Watkins, S.M. (2005) Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* **1**, 3–9, https://doi.org/10.1007/s11306-005-1102-8
32 Bush, W.S. and Moore, J.H. (2012) Genome-wide association studies. *PLoS Comput. Biol.* **8**, e1002822, https://doi.org/10.1371/journal.pcbi.1002822
33 Hirschhorn, J.N. and Daly, M.J. (2005) Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**, 95, https://doi.org/10.1038/nrg1521
34 Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K. et al. (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5**, 613, https://doi.org/10.1038/nmeth.1223

35 Aebersold, R. and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347, https://doi.org/10.1038/nature19949
36 Doll, S., Dreßen, M., Geyer, P.E., Itzhak, D.N., Braun, C., Doppler, S.A. et al. (2017) Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* **8**, 1469, https://doi.org/10.1038/s41467-017-01747-2
37 Higashi, R.M., Fan, T.W-.M., Lorkiewicz, P.K., Moseley, H.N.B and Lane, A.N. (2014) Stable isotope-labeled tracers for metabolic pathway elucidation. *Mass Spectrometry in Metabolomics. Methods in Molecular Biology (Methods and Protocols)*, vol. **1198**, Humana Press, New York, https://doi.org/10.1007/978-1-4939-1258-2˙11
38 Garber, M., Grabherr, M.G., Guttman, M. and Trapnell, C. (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* **8**, 469, https://doi.org/10.1038/nmeth.1613
39 Tyanova, S., Temu, T. and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301, https://doi.org/10.1038/nprot.2016.136
40 Kirpich, A.S., Ibarra, M., Moskalenko, O., Fear, J.M., Gerken, J., Mi, X. et al. (2018) SECIMTools: a suite of metabolomics data analysis tools. *BMC Bioinformatics* **19**, 151, https://doi.org/10.1186/s12859-018-2134-1
41 Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. et al. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25, https://doi.org/10.1038/75556
42 Hansen, J., Meretzky, D., Woldesenbet, S., Stolovitzky, G. and Iyengar, R. (2017) A flexible ontology for inference of emergent whole cell function from relationships between subcellular processes. *Sci. Rep.* **7**, 17689, https://doi.org/10.1038/s41598-017-16627-4
43 Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30, https://doi.org/10.1093/nar/28.1.27
44 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550, https://doi.org/10.1073/pnas.0506580102
45 Gundersen, G.W., Jones, M.R., Rouillard, A.D., Kou, Y., Monteiro, C.D., Feldmann, A.S. et al. (2015) GEO2Enrichr: browser extension and server app to extract gene sets from GEO and analyze them for biological functions. *Bioinformatics* **31**, 3060–3062, https://doi.org/10.1093/bioinformatics/btv297
46 Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z. et al. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97, https://doi.org/10.1093/nar/gkw377
47 Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S. et al. (2018) Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9**, 20, https://doi.org/10.1038/s41467-017-02391-6
48 Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. and Mesirov, J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740, https://doi.org/10.1093/bioinformatics/btr260
49 Tavassoly, I., Parmar, J., Shajahan-Haq, A.N., Clarke, R., Baumann, W.T. and Tyson, J.J. (2015) Dynamic modeling of the interaction between autophagy and apoptosis in mammalian cells. *CPT Pharmacometrics Syst. Pharmacol.* **4**, 263–272, Epub 2015/04/17.eng, https://doi.org/10.1002/psp4.29
50 Barabasi, A.-L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101, https://doi.org/10.1038/nrg1272
51 Berger, S.I. and Iyengar, R. (2009) Network analyses in systems pharmacology. *Bioinformatics* **25**, 2466–2472, https://doi.org/10.1093/bioinformatics/btp465
52 Ideker, T. and Krogan, N.J. (2012) Differential network biology. *Mol. Syst. Biol.* **8**, 565, https://doi.org/10.1038/msb.2011.99
53 Zhao, S. and Iyengar, R. (2012) Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. *Annu. Rev. Pharmacol. Toxicol.* **52**, 505–521, https://doi.org/10.1146/annurev-pharmtox-010611-134520
54 Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A. and Nolan, G.P. (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**, 523–529, https://doi.org/10.1126/science.1105809
55 Needham, C.J., Bradford, J.R., Bulpitt, A.J. and Westhead, D.R. (2007) A primer on learning in Bayesian networks for computational biology. *PLoS Comput. Biol.* **3**, e129, https://doi.org/10.1371/journal.pcbi.0030129
56 Sachs, K., Gifford, D., Jaakkola, T., Sorger, P. and Lauffenburger, D.A. (2002) Bayesian network approach to cell signaling pathway modeling. *Sci STKE* **2002**, pe38–pe
57 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504, https://doi.org/10.1101/gr.1239303
58 Han, J.-D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V. et al. (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88, https://doi.org/10.1038/nature02555
59 Albert, R. (2005) Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957, https://doi.org/10.1242/jcs.02714
60 Hornung, G. and Barkai, N. (2008) Noise propagation and signaling sensitivity in biological networks: a role for positive feedback. *PLoS Comput. Biol.* **4**, e8, https://doi.org/10.1371/journal.pcbi.0040008
61 Barabási, A.-L. (2009) Scale-free networks: a decade and beyond. *Science* **325**, 412–413, https://doi.org/10.1126/science.1173299
62 Tyson, J.J. and Novák, B. (2010) Functional motifs in biochemical reaction networks. *Annu. Rev. Phys. Chem.* **61**, 219–240, https://doi.org/10.1146/annurev.physchem.012809.103457
63 Azeloglu, E.U. and Iyengar, R. (2015) Signaling networks: information flow, computation, and decision making. *Cold Spring Harbor Perspect. Biol.* **7**, a005934, https://doi.org/10.1101/cshperspect.a005934
64 Alon, U. (2007) Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450, https://doi.org/10.1038/nrg2102

65 Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002) Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827, https://doi.org/10.1126/science.298.5594.824
66 Kafri, R., Levy, J., Ginzberg, M.B., Oh, S., Lahav, G. and Kirschner, M.W. (2013) Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. *Nature* **494**, 480, https://doi.org/10.1038/nature11897
67 Neves, S.R. and Iyengar, R. (2002) Modeling of signaling networks. *Bioessays* **24**, 1110–1117, https://doi.org/10.1002/bies.1154
68 Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R. and Sethna, J.P. (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189, https://doi.org/10.1371/journal.pcbi.0030189
69 Yadaw, A.S., Siddiq, M.M., Rabinovich, V., Tolentino, R., Iyengar, R. and Hansen, J. (2018) Dynamic balance between vesicle transport and microtubule growth enables neurite growth. *bioRxiv* 153569, https://doi.org/10.1101/153569
70 Novák, B. and Tyson, J.J. (2008) Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9**, 981, https://doi.org/10.1038/nrm2530
71 Neves, S.R. and Iyengar, R. (2009) Models of spatially restricted biochemical reaction systems. *J. Biol. Chem.* **284**, 5445–5449, https://doi.org/10.1074/jbc.R800058200
72 Rangamani, P., Lipshtat, A., Azeloglu, E.U., Calizo, R.C., Hu, M., Ghassemi, S. et al. (2013) Decoding information in cell shape. *Cell* **154**, 1356–1369, https://doi.org/10.1016/j.cell.2013.08.026
73 Leung, A.W. (2013) *Systems of Nonlinear Partial Differential Equations: Applications to Biology and Engineering*, Springer Science & Business Media
74 Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N. et al. (2006) COPASI—a complex pathway simulator. *Bioinformatics* **22**, 3067–3074, https://doi.org/10.1093/bioinformatics/btl485
75 Schmidt, H. and Jirstrand, M. (2005) Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics* **22**, 514–515, https://doi.org/10.1093/bioinformatics/bti799
76 Loew, L.M. and Schaff, J.C. (2001) The Virtual Cell: a software environment for computational cell biology. *Trends Biotechnol.* **19**, 401–406, https://doi.org/10.1016/S0167-7799(01)01740-1
77 Szekely, T. and Burrage, K. (2014) Stochastic simulation in systems biology. *Comput. Struct. Biotechnol. J.* **12**, 14–25, https://doi.org/10.1016/j.csbj.2014.10.003
78 Gillespie, D.T. (2007) Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55, https://doi.org/10.1146/annurev.physchem.58.032806.104637
79 Wilkinson, D.J. (2009) Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Genet.* **10**, 122, https://doi.org/10.1038/nrg2509
80 Tavassoly, I. (2015) *Dynamics of Cell Fate Decision Mediated by the Interplay of Autophagy and Apoptosis in Cancer Cells: Mathematical Modeling and Experimental Observations*, Springer
81 Loewer, A. and Lahav, G. (2011) We are all individuals: causes and consequences of non-genetic heterogeneity in mammalian cells. *Current Opin. Genet. Dev.* **21**, 753–758, https://doi.org/10.1016/j.gde.2011.09.010
82 Azeloglu, E.U. and Iyengar, R. (2015) Good practices for building dynamical models in systems biology. *Sci. Signal* **8**, fs8-fs, https://doi.org/10.1126/scisignal.aab0880
83 Morohashi, M., Winn, A.E., Borisuk, M.T., Bolouri, H., Doyle, J. and Kitano, H. (2002) Robustness as a measure of plausibility in models of biochemical networks. *J. Theor. Biol.* **216**, 19–30, https://doi.org/10.1006/jtbi.2002.2537
84 Boran, A.D. and Iyengar, R. (2010) Systems approaches to polypharmacology and drug discovery. *Current Opin. Drug Discov. Dev.* **13**, 297
85 Hansen, J., Zhao, S. and Iyengar, R. (2011) Systems pharmacology of complex diseases. *Ann. N. Y. Acad. Sci.* 1245
86 Iyengar, R., Altman, R.B., Troyanskya, O. and FitzGerald, G.A. (2015) Personalization in practice. *Science* **350**, 282–283, https://doi.org/10.1126/science.aad5204
87 Eungdamrong, N.J. and Iyengar, R. (2004) Computational approaches for modeling regulatory cellular networks. *Trends Cell Biol.* **14**, 661–669, https://doi.org/10.1016/j.tcb.2004.10.007
88 Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H. et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401
89 Saliba, A.-E., Westermann, A.J., Gorski, S.A. and Vogel, J. (2014) Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* **42**, 8845–8860, https://doi.org/10.1093/nar/gku555
90 Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D. et al. (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363, https://doi.org/10.1038/nature13437
91 Sandberg, R. (2014) Entering the era of single-cell transcriptomics in biology and medicine. *Nat. Methods* **11**, 22, https://doi.org/10.1038/nmeth.2764
92 Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R. et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236, https://doi.org/10.1038/nature12172
93 Wang, D. and Bodovitz, S. (2010) Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol.* **28**, 281–290, https://doi.org/10.1016/j.tibtech.2010.03.002
94 Baker, R.E., Peña, J.-M., Jayamohan, J. and Jérusalem, A. (2018) Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biol. Lett.* **14**, 20170660, https://doi.org/10.1098/rsbl.2017.0660
95 Baxt, W.G. (1995) Application of artificial neural networks to clinical medicine. *Lancet North Am. Ed.* **346**, 1135–1138, https://doi.org/10.1016/S0140-6736(95)91804-3
96 Deo, R.C. (2015) Machine learning in medicine. *Circulation* **132**, 1920–1930, https://doi.org/10.1161/CIRCULATIONAHA.115.001593
97 Leung, M.K., Delong, A., Alipanahi, B. and Frey, B.J. (2016) Machine learning in genomic medicine: a review of computational problems and data sets. *Proc. IEEE* **104**, 176–197, https://doi.org/10.1109/JPROC.2015.2494198

Essays in Biochemistry (2018) **62** 487–500
https://doi.org/10.1042/EBC20180003

98 Sommer, C. and Gerlich, D.W. (2013) Machine learning in cell biology–teaching computers to recognize phenotypes. *J. Cell Sci.*, jcs. 123604, https://doi.org/10.1242/jcs.123604
99 Janowczyk, A. and Madabhushi, A. (2016) Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J. Pathol. Informatics* **7**, https://doi.org/10.4103/2153-3539.186902
100 LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep learning. *Nature* **521**, 436, https://doi.org/10.1038/nature14539
101 Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C. and Collins, J.J. (2018) Next-generation machine learning for biological networks. *Cell* **173**, 1581–1592, https://doi.org/10.1016/j.cell.2018.05.015
102 Kell, D.B. (2006) Metabolomics, modelling and machine learning in systems biology–towards an understanding of the languages of cells. *FEBS J.* **273**, 873–894, https://doi.org/10.1111/j.1742-4658.2006.05136.x
103 Moles, C.G., Mendes, P. and Banga, J.R. (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* **13**, 2467–2474, https://doi.org/10.1101/gr.1262503
104 Catto, J.W., Linkens, D.A., Abbod, M.F., Chen, M., Burton, J.L., Feeley, K.M. et al. (2003) Artificial intelligence in predicting bladder cancer outcome: a comparison of neuro-fuzzy modeling and artificial neural networks. *Clin. Cancer Res.* **9**, 4172–4177
105 Berger, S.I., Ma'ayan, A. and Iyengar, R. (2010) Systems pharmacology of arrhythmias. *Sci. Signal.* **3**, ra30, https://doi.org/10.1126/scisignal.2000723
106 van der Graaf, P.H. and Benson, N. (2011) Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development. *Pharm. Res.* **28**, 1460–1464, https://doi.org/10.1007/s11095-011-0467-9
107 Zhao, S., Nishimura, T., Chen, Y., Azeloglu, E.U., Gottesman, O., Giannarelli, C. et al. (2013) Systems pharmacology of adverse event mitigation by drug combinations. *Sci. Transl. Med.* **5**, 206ra140–206ra140, https://doi.org/10.1126/scitranslmed.3006548
108 Asada, H.H., Shaltis, P., Reisner, A., Rhee, S. and Hutchinson, R.C. (2003) Mobile monitoring with wearable photoplethysmographic biosensors. *IEEE Eng. Med. Biol. Mag.* **22**, 28–40, https://doi.org/10.1109/MEMB.2003.1213624
109 Turner, A.P. (2000) Biosensor–sense and sensitivity. *Science* **290**, 1315–1317, https://doi.org/10.1126/science.290.5495.1315
110 Crowley, E., Di Nicolantonio, F., Loupakis, F. and Bardelli, A. (2013) Liquid biopsy: monitoring cancer-genetics in the blood. *Nat. Rev. Clin. Oncol.* **10**, 472, https://doi.org/10.1038/nrclinonc.2013.110
111 Pantanowitz, L. (2010) Digital images and the future of digital pathology. *J. Pathol. Informatics* **1**, https://doi.org/10.4103/2153-3539.68332
112 Jelinek, H.F., Rocha, A., Carvalho, T., Goldenstein, S. and Wainer, J. (2011) Machine learning and pattern classification in identification of indigenous retinal pathology. Engineering in Medicine and Biology Society, EMBC, 2011. *Annual Int. Conf. IEEE*
113 van Hasselt, J.C. and Iyengar, R. (2017) Systems pharmacology-based identification of pharmacogenomic determinants of adverse drug reactions using human iPSC-derived cell lines. *Curr. Opin. Syst Biol.* **4**, 9–15, https://doi.org/10.1016/j.coisb.2017.05.006
114 Tavassoly, I. and Iyengar, R. (2018) Analysis of sensitivity of genomic signatures of therapeutic responses of non-small cell lung cancer in patient-derived xenograft models. *AACR*, https://doi.org/10.1158/1538-7445
115 Tavassoly, I., Hu, Y., Zhao, S., Mariottini, C., Boran, A., Chen, Y. et al. (2018) Systems therapeutics analyses identify genomic signatures defining responsiveness to allopurinol and combination therapy for lung cancer. *bioRxiv.*, https://doi.org/10.1101/396697
116 Voong, K.R., Feliciano, J., Becker, D. and Levy, B. (2017) Beyond PD-L1 testing-emerging biomarkers for immunotherapy in non-small cell lung cancer. *Ann. Transl. Med.* **5**, https://doi.org/10.21037/atm.2017.06.48

Spring 2019 – Epigenetics and Systems Biology
Lecture Outline (Systems Biology)
Michael K. Skinner – Biol 476/576
CUE 418, 10:35-11:50 am, Tuesdays & Thursdays
January 22 & 29, 2019
Weeks 3 and 4

Systems Biology (Components & Technology)

Components (DNA, Expression, Cellular, Organ, Physiology, Organism, Differentiation, Development, Phenotype, Evolution)

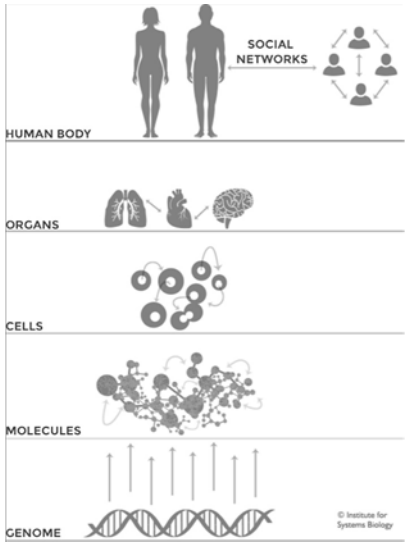
Technology (Genomics, Transcriptomes, Proteomics) (Interaction, Signaling, Metabolism)

Omics (Data Processing and Resources)

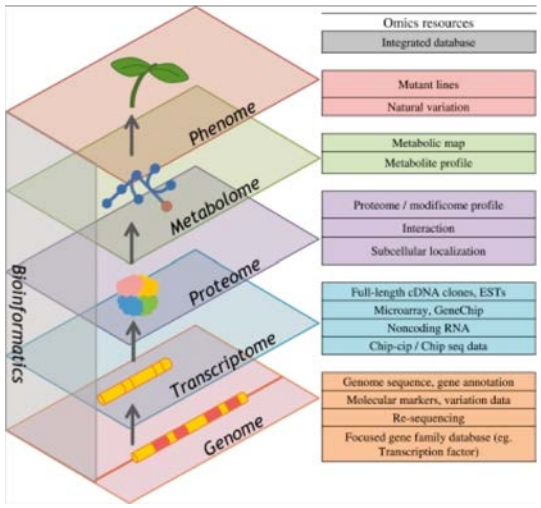
Required Reading

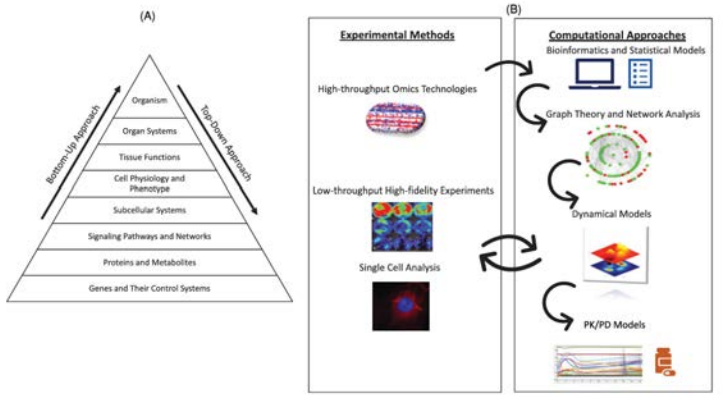
ENCODE (2012) ENCODE Explained. Nature 489:52-55.

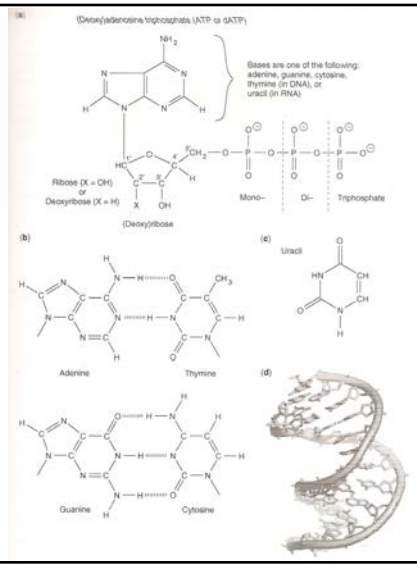
Tavassoly I, Goldfarb J, Iyengar R. (2018) Essays Biochem. 62(4):487-500.

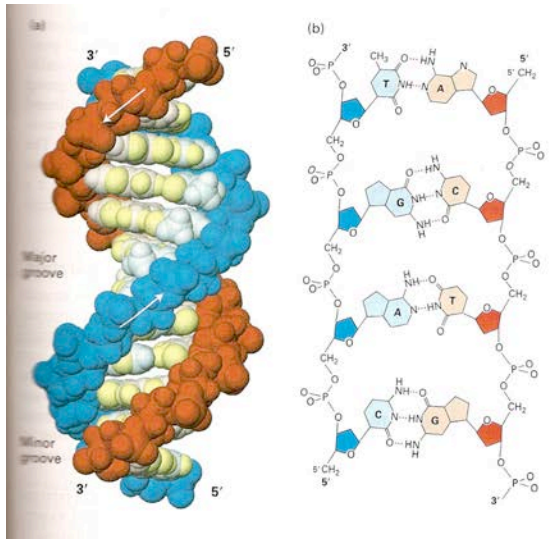


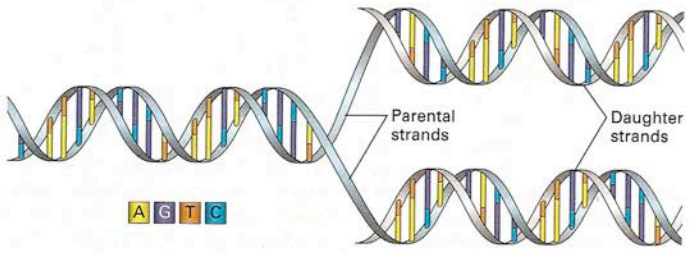
The human being as a dynamic network of networks. In systems medicine the human organism is envisioned as a system of systems or network of networks. At every scale of biological organization (molecular, cellular, organ, individual and social/environmental) systems are portrayed as giving rise to and embedding each other. At all levels the network of networks is seen as a dynamic or four-dimensional process (as opposed to a static thing) (Copyright: The Institute for Systems Biology, used with permission)

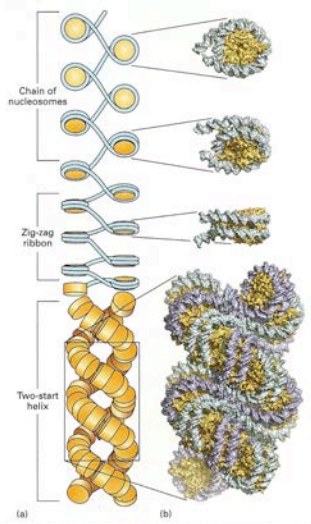





(Deoxy)adenosine triphosphate (ATP or dATP)
Bases are one of the following: adenine, guanine, cytosine, thymine (in DNA), or uracil (in RNA)
Ribose (X = OH) or Deoxyribose (X = H)
(Deoxy)ribose
Mono-
Di-
Triphosphate
(b)
(c)
Uracil
Adenine
Thymine
(d)
Guanine
Cytosine


(a)
(b)
3′
5′
Major groove
Minor groove
3′
5′
5′
3′


Parental strands
Daughter strands
A
G
T
C


Chain of nucleosomes
Zig-zag ribbon
Two-start helix
(a)
(b)

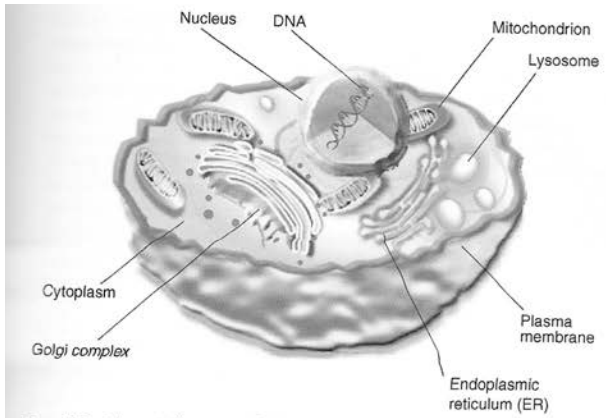
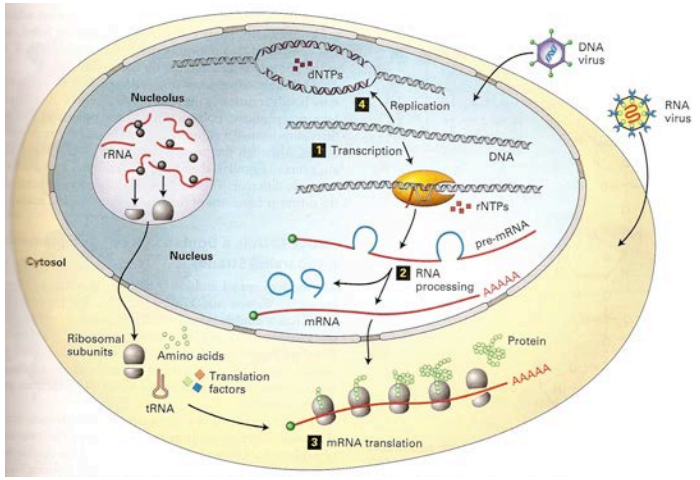
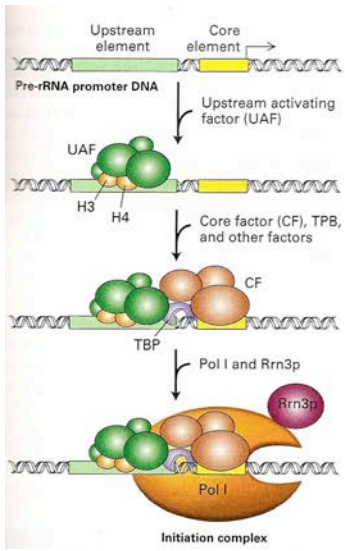
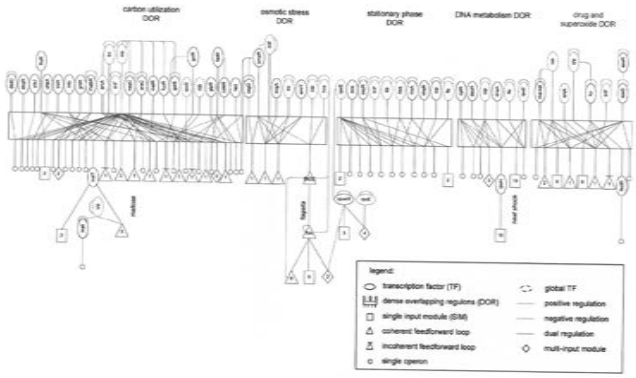
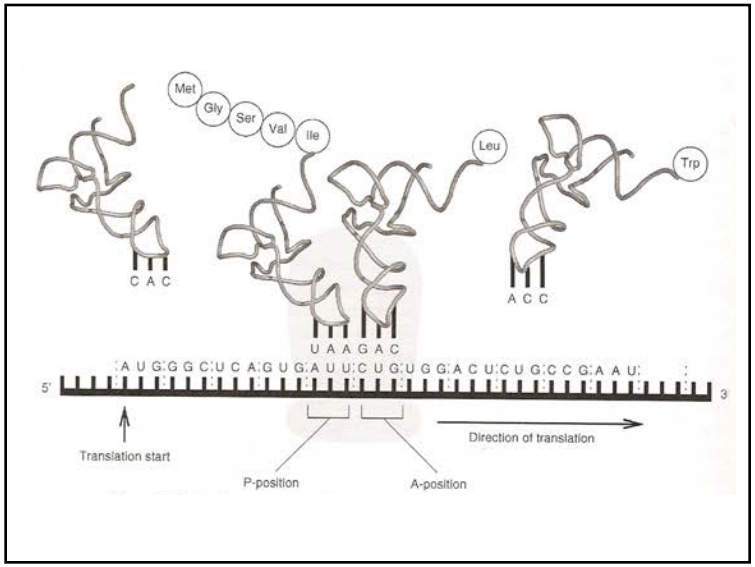


Figure 10.8 Schematic illustration of an animal cell with its major organelles.









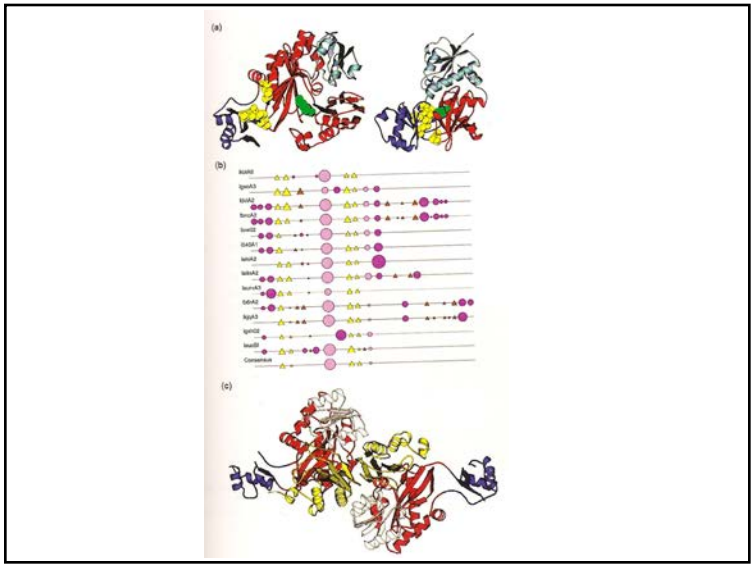
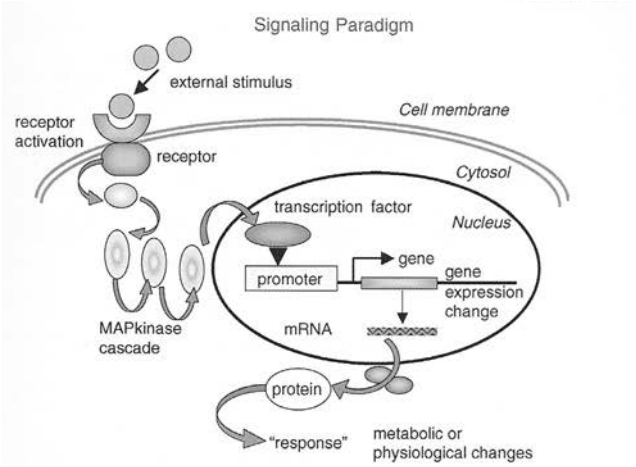
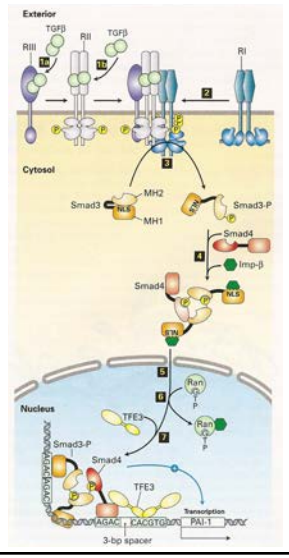
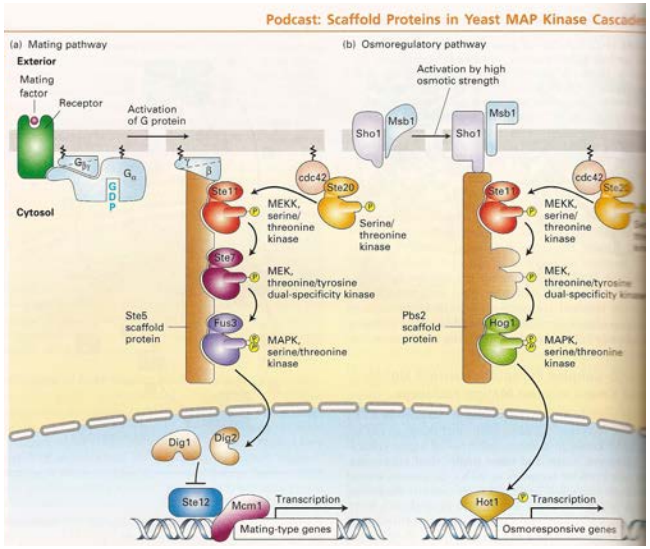


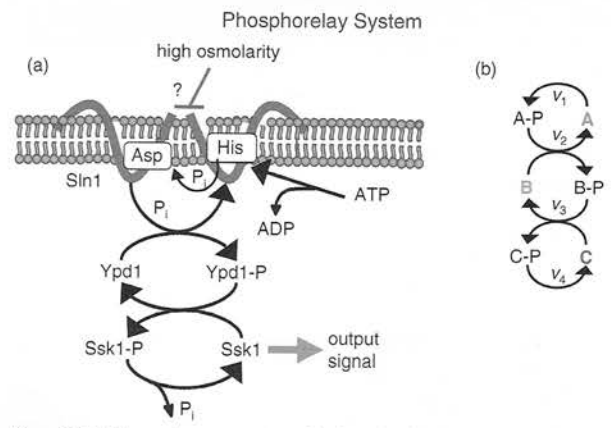
TABLE 2.2 Timescales for the Reactions in the Transcription Network of the Bacterium *E. coli* (Order of Magnitude)

Binding of a small molecule (a signal) to a transcription factor, causing a change in transcription factor activity	~1 msec
Binding of active transcription factor to its DNA site	~1 sec
Transcription + translation of the gene	~5 min
Timescale for 50% change in concentration of the translated protein (stable proteins)	~1 h (one cell generation)









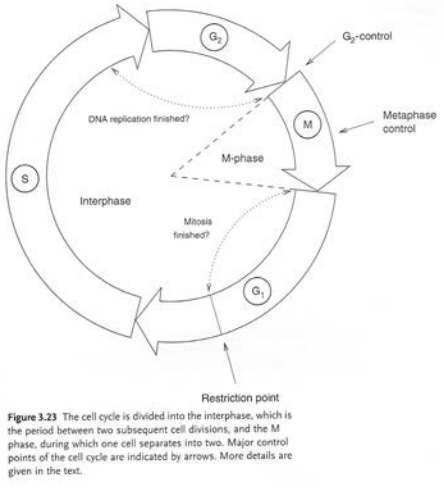


Figure 3.23 The cell cycle is divided into the interphase, which is the period between two subsequent cell divisions, and the M phase, during which one cell separates into two. Major control points of the cell cycle are indicated by arrows. More details are given in the text.

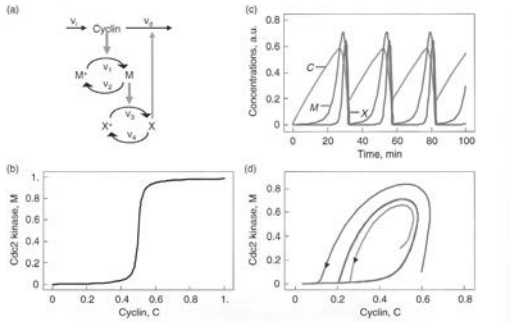
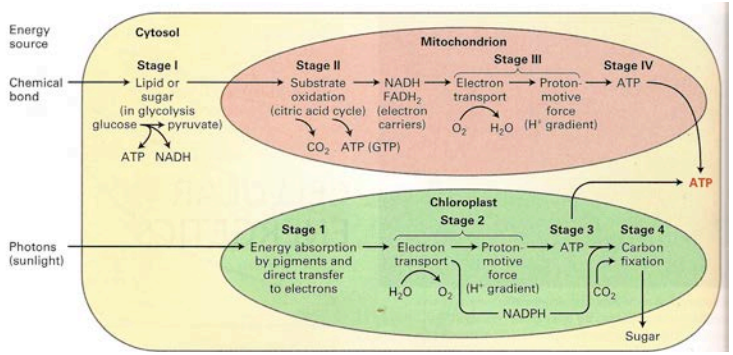
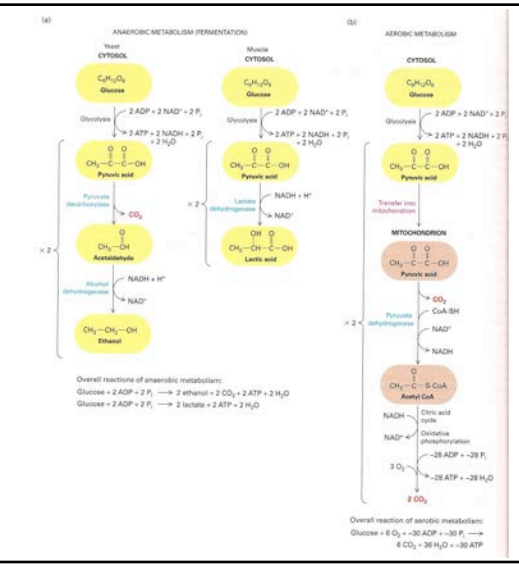
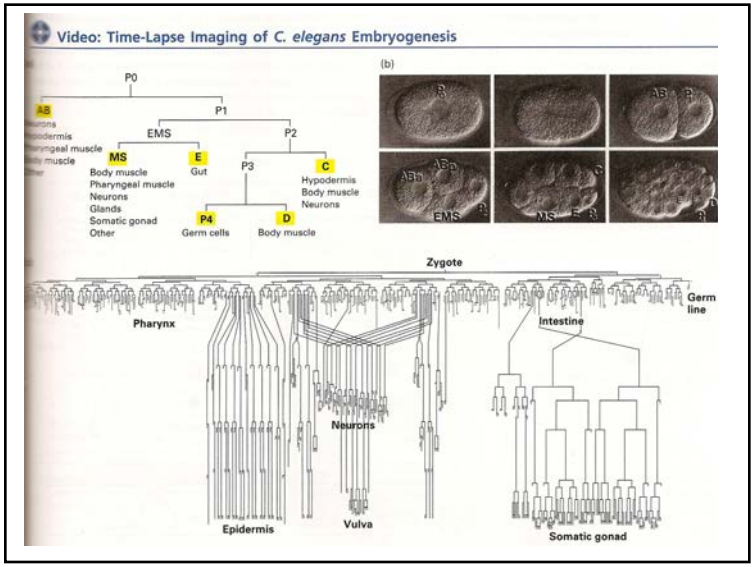
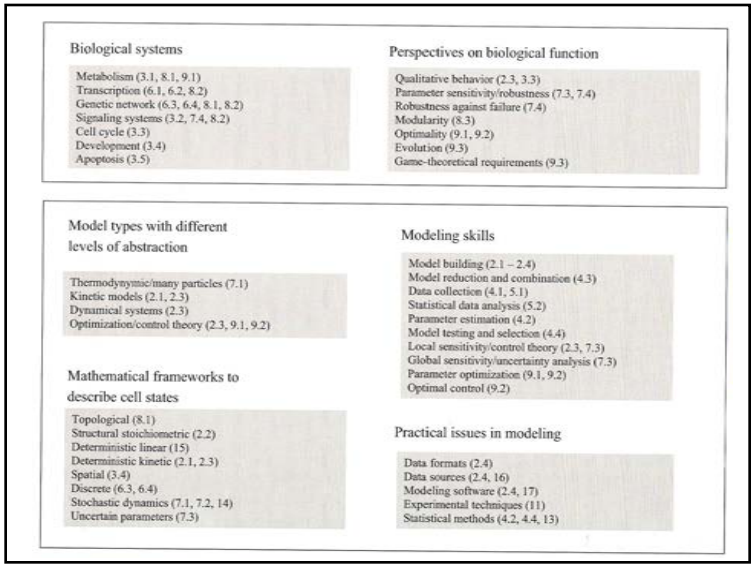


Figure 3.24 Goldbeter's minimal model of the mitotic oscillator. (a) Illustration of the model comprising cyclin production and degradation, phosphorylation and dephosphorylation of Cdc2 kinase, and phosphorylation and dephosphorylation of the cyclin protease (see text). (b) Threshold-type dependence of the fractional concentration of active Cdc2 kinase on the cyclin concentration. (c) Time courses of cyclin (C), active Cdc2 kinase (M), and active cyclin protease (X) exhibiting oscillations according to Equation system in Eq. (3.1). (d) Limit cycle behavior, represented for the variables C and M. Parameter values: $K_{mi} = 0.05$ ($i = 1, \ldots, 4$), $K_{mr} = 0.5$, $k_d = 0.01$, $\nu_i = 0.025$, $\nu_d = 0.25$, $V_{M1} = 3$, $V_{M2} = 1.5$, $V_{m3} = 1$, $V_{m4} = 0.5$. Initial conditions in (b) are $C(0) = M(0) = X(0) = 0.01$, and in (c) are $X(0) = 0.01$. Units: μM and min^{-1}.

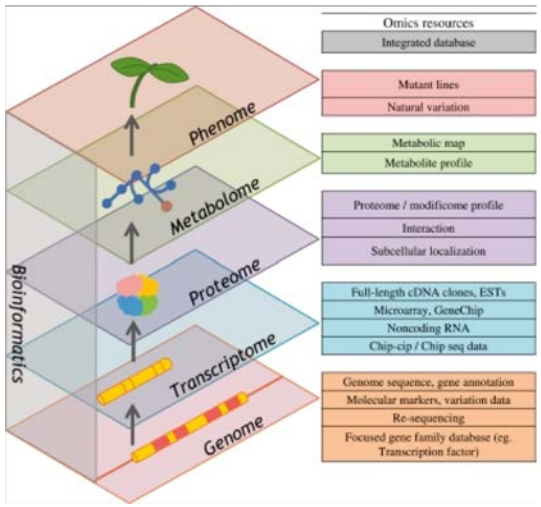


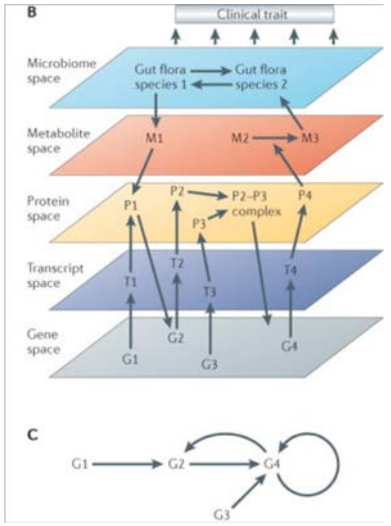


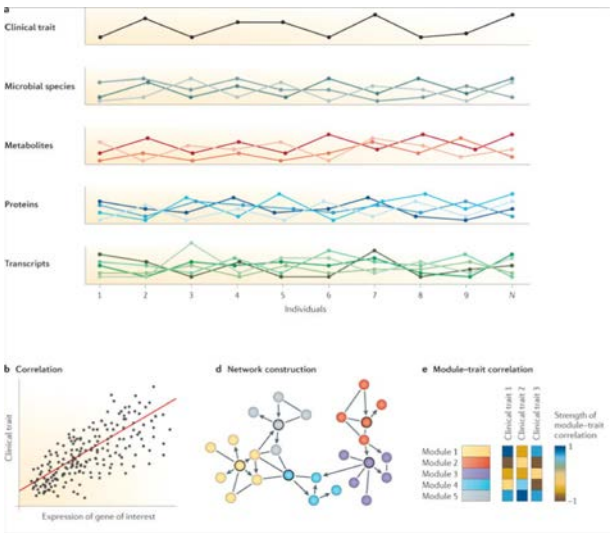


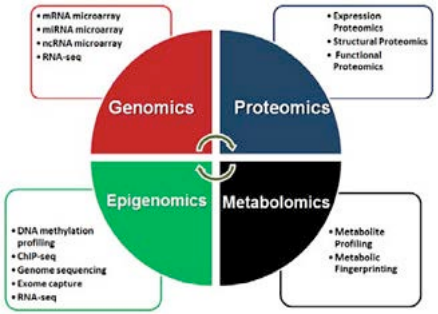


Omics Technology

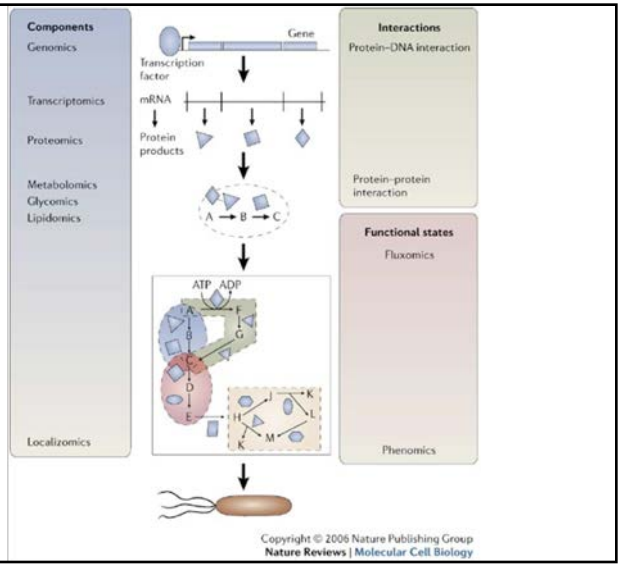




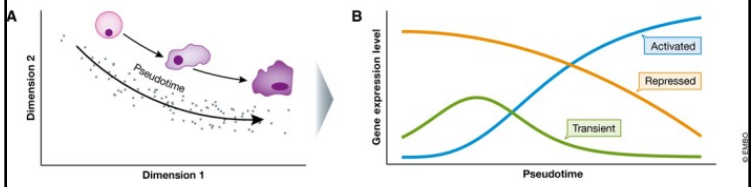




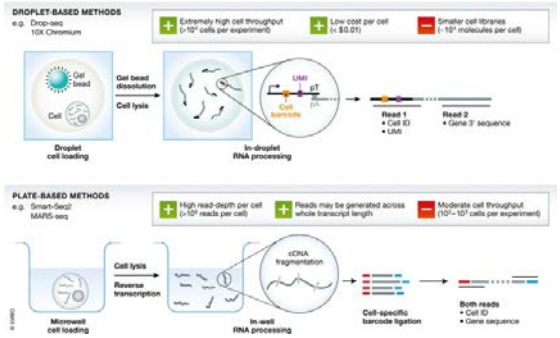
Genomics	Transcriptomics	Proteomics	Metabolomics	Protein–DNA interactions	Protein–protein interactions	Fluxomics	Phenomics
Genomics (sequence annotation)	• ORF validation • Regulatory element identification[74]	• SNP effect on protein activity or abundance	• Enzyme annotation	• Binding-site identification[75]	• Functional annotation[76]	• Functional annotation	• Functional annotation[5,103] • Biomarkers[115]
	Transcriptomics (microarray, SAGE)	• Protein: transcript correlation[20]	• Enzyme annotation[104]	• Gene-regulatory networks[78]	• Functional annotation[79] • Protein complex identification[82]		• Functional annotation[102]
		Proteomics (abundance, post-translational modification)	• Enzyme annotation[105]	• Regulatory complex identification	• Differential complex formation	• Enzyme capacity	• Functional annotation
			Metabolomics (metabolite abundance)	• Metabolic–transcriptional response		• Metabolic pathway bottlenecks	• Metabolic flexibility • Metabolic engineering[106]
				Protein–DNA interactions (ChIP–chip)	• Signalling cascades[86,102]		• Dynamic network responses[87]
					Protein–protein interactions (yeast 2H, coAP–MS)		• Pathway identification activity[88]
						Fluxomics (isotopic tracing)	• Metabolic engineering
							Phenomics (phenotype arrays, RNAi screens, synthetic lethals)



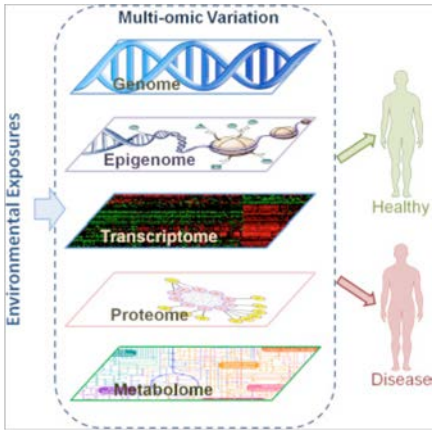
Using single-cell genomics to understand developmental processes and cell fate decisions.
Mol Syst Biol. 2018 Apr 16;14(4):e8046.
Griffiths JA, Scialdone A, Marioni JC.

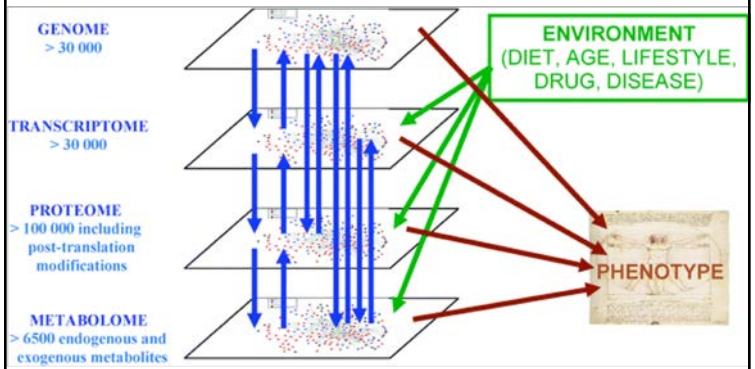


Pseudotime recapitulates developmental trajectories (A) By observing similarities between the expression profiles of cells, it is possible to order cells along an axis of pseudotime that recapitulates developmental processes. (B) Having established this ordering, genes that show significant changes in expression along the developmental pathway may be identified.

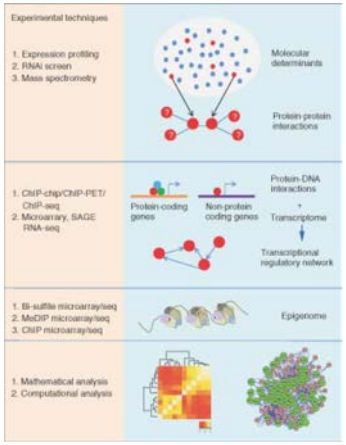


Single-cell library preparation summary There are two primary methods for generating single-cell transcriptomics data: plate-based and droplet-based methods, shown above. In summary, droplet-based approaches offer high cell throughput, while plate-based approaches provide higher resolution in each individual cell. Note that different implementations of these methods provide slightly different outputs and that some steps are excluded for clarity (e.g. cDNA amplification).



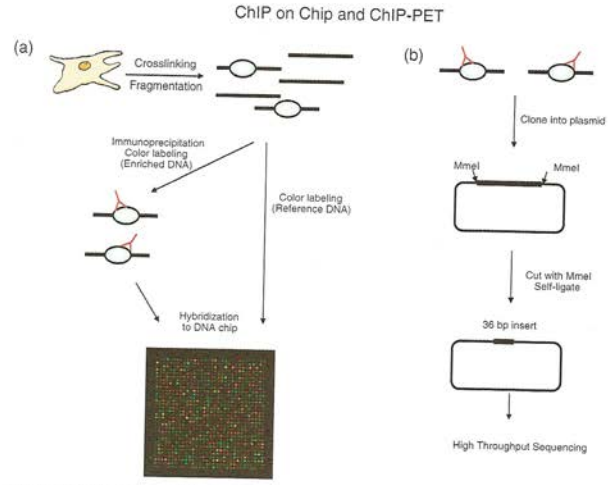


Genomics



Tiling array

A high-density microarray that contains evenly spaced, or ‘tiled’, sets of probes that span the genome or chromosome, and can be used in many experimental applications such as transcriptome characterization, gene discovery, alternative-splicing analysis, ChIP-chip, DNA-methylation analysis, DNA-polymorphism analysis, comparative genome analysis and genome resequencing.



Complete strategy for TFBSs focused ChIP-Seq and RNA-Seq data analysis.

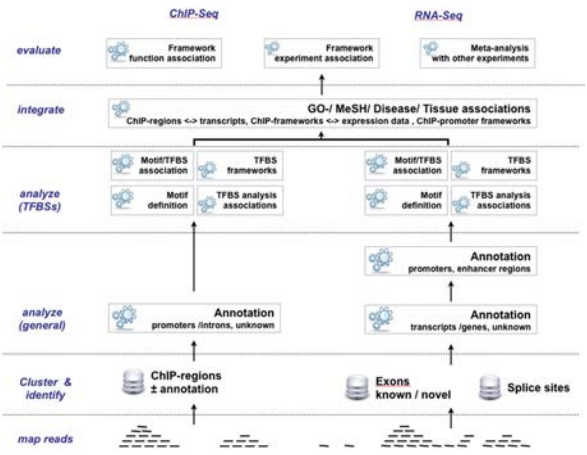


Table 2.1 Types of high-throughput sequencing technologies

Optical sequencing					
Platform	Instrument	Template preparation	Chemistry	Average length	Longest read
Illumina	HiSeq2500	BridgePCR/cluster	Rev. term., SBS	100	150
Illumina	HiSeq2000	BridgePCR/cluster	Rev. term., SBS	100	150
Illumina	MiSeq	BridgePCR/cluster	Rev. term., SBS	250	300
GnuBio	GnuBio	emPCR	Hyb-assist sequencing	1,000[a]	64,000[a]
Life Technologies	SOLiD 5500	emPCR	Seq. by Lig.	75	100
LaserGen	LaserGen	emPCR	Rev. term., SBS	25[a]	100[a]
Pacific biosciences	RS	Polymerase binding	Real-time	1,800	15,000
454	Titanium	emPCR	PyroSequencing	650	1,100
454	Junior	emPCR	PyroSequencing	400	650
Helicos	Heliscope	Adaptor ligation	Rev. term., SBS	35	57
Intelligent BioSystems	MAX-Seq	Rolony amplification	Two-step SBS (label/unlabell)	2 × 100	300
Intelligent BioSystems	MINI-20	Rolony amplification	Two-step SBS (label/unlabell)	2 × 100	300
ZS Genetics	N/A	Atomic lableing	Electron microscope	N/A	N/A
Halcyon Molecular	N/A	N/A	Direct observation of DNA	N/A	N/A
Electical sequencing					
Platform	Instrument	Template preparation	Chemistry	Average length	Longest read
IBM DNA transistor	N/A	None	Microchip nanopore	N/A	N/A
NABsys	N/A	None	Nanochannel	N/A	N/A
Bionanogenomics	N/A	Anneal 7mers	Nanochannel	N/A	N/A
Life Technologies	PGM	emPCR	Semi-conductor	150	300
Life Technologies	Proton	emPCR	Semi-conductor	120	240
Life Technologies	Proton 2	emPCR	Semi-conductor	400[a]	800[a]
Genia	N/A	None	Protein nanopore (a-hemalysin)	N/A	N/A
Oxford nanopore	MinION	None	Protein nanopore	10,000	10,000[a]
Oxford nanopore	GridION 2K	None	Protein nanopore	10,000	500,000[a]
Oxford nanopore	GridION 8K	None	Protein nanopore	10,000	500,000[a]

[a]Values are estimates from companies that have not yet released actual data

Table 1. A comparison of representative third-generation DNA sequencing companies

	Helicos	Pacific Biosciences	Oxford Nanopore	Complete Genomics	Ion Torrent
Key technology	Amplification-free sequencing	Zero-mode waveguide nanostructure arrays	Protein nanopores	Self-assembling DNA nanoarrays	Chemical sensitive field-effect transistor arrays
Single-molecule detection	Yes	Yes	Yes	No	Undisclosed
Commercialization	Launched in 2008	Launched in March 2010	Undisclosed estimated launch	Sequencing services	Launched in March 2010
Funding to date (millions USD)[a]	$115	$266	$64	$45	$23
Refs	[43]	[44]	[46]	[47]	[66]

[a]Company website data.

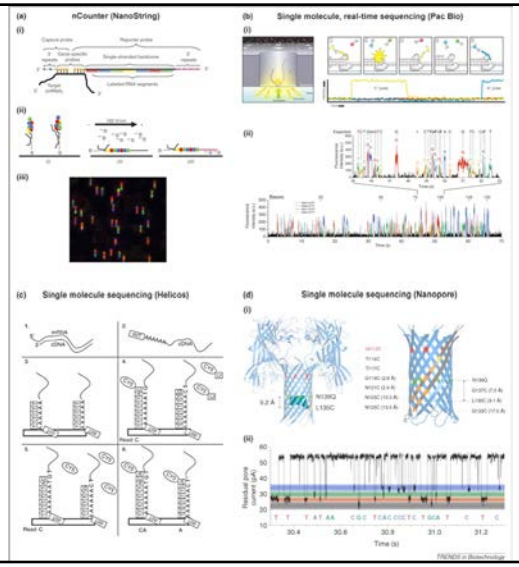
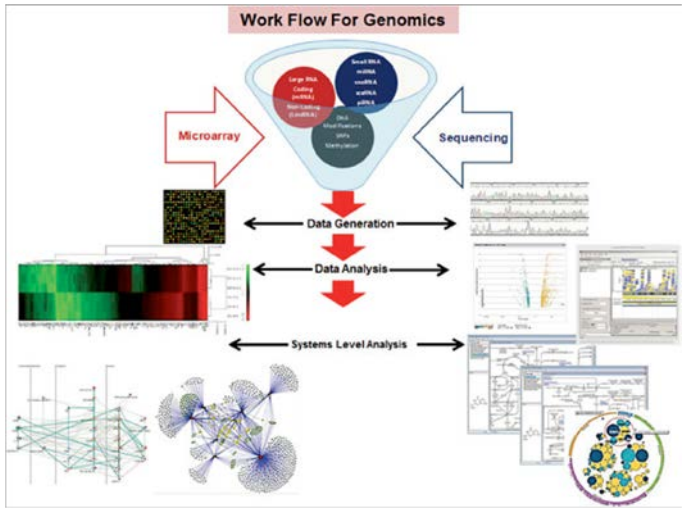


Table 2
Cancer driver mutations discovered by large-scale next generation sequencing.

Gene	Aberration type	Tumor type	Biological function	Tumor effect	Sequencing method	Number of samples	Sample type	Reference
EBF1-PDGFRB, BCR-JAK2, NUP214-ABL1	Fusion	ALL	Kinase signaling	Activating	Whole-genome	15	Acute lymphoblastic leukemia	22
IL7R, SH2B3	Mutation	ALL	Cytokine signaling	Activating	Whole-genome	15	Acute lymphoblastic leukemia	22
TP53	Mutation	Cell Carcinoma	Cell cycle regulation	Inactivating	Whole-genome	457	Peripheral blood	23
VTI1A-TCF7L2	Fusion	Colon	Transcription factor	Activating	Whole-genome	9	Colorectal adenocarcinomas	24
ARID1A, ARID1B, ARID2, MLL, MLL3	Mutation	Liver	Chromatin regulation	Inactivating	Whole-genome	27	Hepatocellular carcinoma	25
PREX2	Mutation	Melanoma	Rac exchange factor	Inactivating	Whole-genome	25	Melanomas	26
ATRX	Mutation	Neuroblastoma	Telomere maintenance	Inactivating	Whole-genome	40	Neuroblastomas	27
BRIP1	Mutation	Ovary	DNA repair	Inactivating	Whole-genome	457	Peripheral blood	28
DNMT3A	Mutation	AML	DNA methylation	Inactivating	Exome	112	Acute monocytic leukemias	54
CBFB	Mutation	Breast	Transcription factor	Inactivating	Exome	103	Breast cancers	55
MAGI3-AKT3	Fusion	Breast	Cell signaling	Activating	Exome	103	Breast cancers	55
NOTCH1	Mutation	Cell carcinoma	Cell signaling	Inactivating	Exome	32	Head and neck squamous cell carcinomas	56
SF3B1	Mutation	CML	mRNA splicing	Inactivating	Exome	105	Chronic lymphocytic leukemias	57
MXRA5	Mutation	Lung	Matrix remodeling	Activating	Exome	14	Non-small cell lung carcinomas	58
CSMD3	Mutation	Lung	Unknown	Inactivating	Exome	31	Non-small cell lung carcinomas	59
RAC1	Mutation	Melanoma	Cell signaling	Activating	Exome	147	Melanomas	60
GRIN2A	Mutation	Melanoma	Glutamate receptor	Unknown	Exome	14	Melanomas	61
SPOP, FOXA1, MED 12	Mutation	Prostate	Transcription regulation	Unknown	Exome	112	Prostate tumors	62
FAT4	Mutation	Stomach	Cell adhesion	Inactivating	Exome	15	Gastric adenocarcinomas	63
ARID1A	Mutation	Stomach	Chromatin remodeling	Inactivating	Exome	15	Gastric adenocarcinomas	63



Transcriptome

Table 1

Comparison of Approaches to Identify Transcriptional Components

Identifying TFs Through Expression Profiling		
Method	**Pros**	**Cons**
Gene expression arrays	Technology and analysis tools very widespread; nonTF target genes can be assessed in concert with TF genes	Small dynamic range for quantitative analysis of gene expression; low sensitivity; limited to known transcripts
qPCR (Quanttrx)	Real-Time PCR technology available to most molecular labs; data easy to analyze; very sensitive; quantitative over wide range of expression; highly reproducible; direct analysis of TFs	Limited to pre-defined list of TFs and isoforms
Sequencing (RNA-seq)	Very Sensitive; Quantitative over wide dynamic range; Identifies all known and unknown transcripts; nonTF target genes can be assessed in concert with TF genes	Technology and analysis methods not yet widespread; large datasets require bioinformatics expertise
Nanostring	Very sensitive; quantitative over wide range of expression; high-throughput; direct analysis of TFs	Technology not currently widespread; requires up front investment in specialized equipment; limited to pre-defined list of TFs and isoforms
DNase-Seq	Increased sequencing depth can reveal 'footprinted' motifs; requires less starting material than ChIP-Seq	Does not distinguish between enhancers, promoters, or other regulatory elements; optimization can be troublesome; cost of deep sequencing is high (but declining); does not immediately reveal relevant TF involved; requirement for significant downstream analysis at genome-wide level
FAIRE-Seq	Technically simple; requires less starting material than ChIP-Seq	Does not distinguish between enhancers, promoters, or other regulatory elements; cost of deep sequencing is high (but declining); does not immediately reveal relevant TF involved; requirement for significant downstream analysis at genome-wide level
ChIP-seq of modified histone marks	Can identify enhancers specifically, and can distinguish between poised and active enhancers.	Can require significant amounts of starting material ($>10^6$ cells per epitope); need for high quality antibody; cost of deep sequencing is high (but declining); does not immediately reveal relevant TF involved; requirement for significant downstream analysis at genome-wide level

5.1
High-Throughput Experiments

Summary

The analysis of transcriptome data has become increasingly popular over the last decades due to the advent of new high-throughput technologies in genome research. Often, these data build the basis for defining the essential molecular read-outs for a particular disease, developmental state, or drug response being subject to computational modeling. In particular, DNA arrays have become the most prominent experimental technique to analyze gene expression data. A DNA array consists of a solid support that carries DNA sequences representing genes – the probes. In hybridization experiments with the target sample of labeled mRNAs and through subsequent data capture a numerical value, the signal intensity, is assigned to each probe. It is assumed that this signal intensity is proportional to the amount of molecules of the respective gene in the target sample. Changes in signal intensities are interpreted as concentration changes. Several experimental platforms are available that enable the genome-wide analysis of gene expression. Another recently emerging high-throughput technology is next generation sequencing. These new sequencing techniques provide in many cases flexible alternatives to DNA array techniques in identifying the abundance of specific sequences, providing information on transcript abundance or RNA processing events.

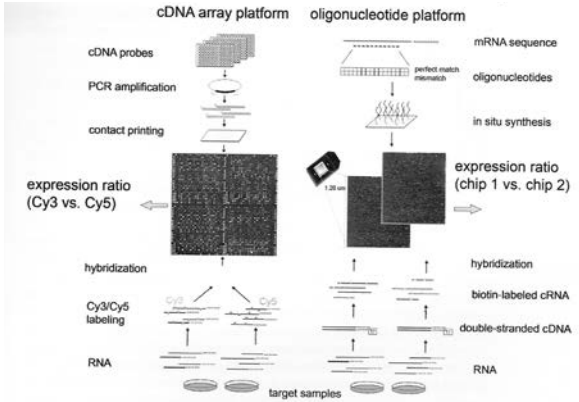


Figure 5.1 Scheme of DNA array analysis using cDNAs (left) and oligonucleotides (right, here Affymetrix technology) as probes. Probe design and construction of the array is shown in the upper part of the figure and the different preparation steps of the target samples are shown in the lower part. In the center of the figure, the strategy for case-control studies is shown using a two-color approach (left panel) and a single-color approach (right panel). Array images were taken from [69] and from http://www.affymetrix.com.

5.2
Analysis of Gene Expression Data

Summary

The analysis of genome-wide gene expression data involves basic concepts from multivariate statistics. Most applications belong to two groups: the first group consists of case-control studies comparing a certain transcriptome state of the biological system (e.g., disease state, perturbed state) to the control situation; the second group of applications consist of multiple case studies involving different states (e.g., drug response time series, groups of patients, etc.). The analysis of case-control studies involves testing of statistical hypotheses. Here, expression changes are observed that deviate from a predefined hypothesis and this deviation is judged for significance. The basic methods for multicase studies are clustering and classification. Here, groups of coexpressed genes serve to identify functionally related groups of genes or experiments. These types of analysis result in the identification of marker genes and their related interactions, which are the basis for further network studies.

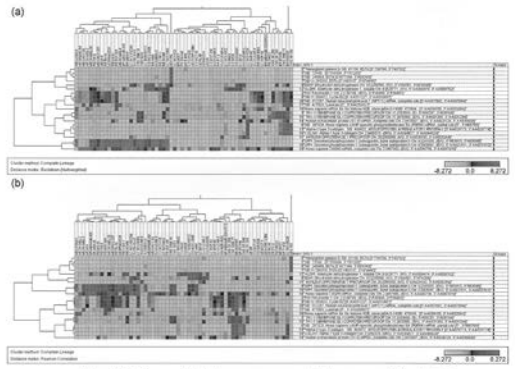
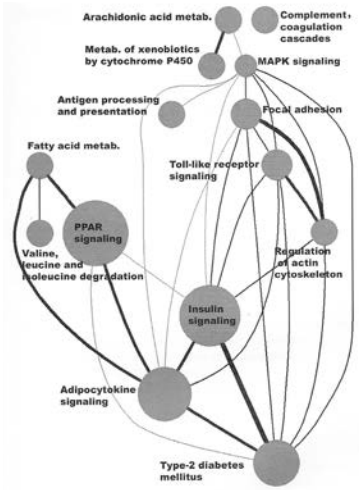
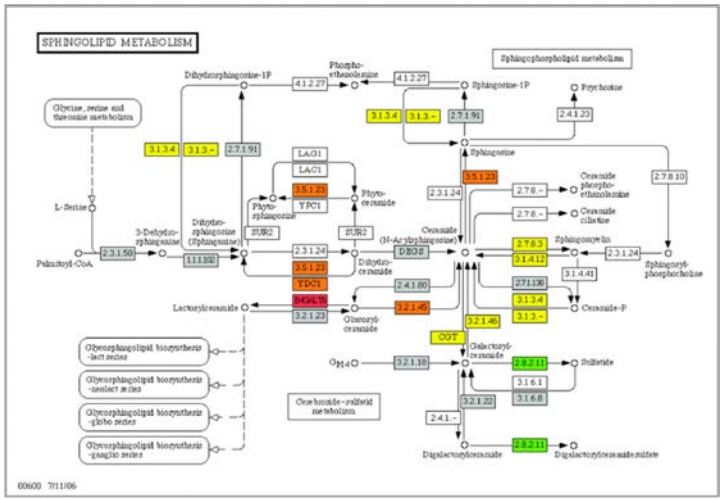


Figure 5.5 Influence of similarity measure on clustering. Two dendrograms of a subgroup of genes using the microarray expression data of Ross *et al.* [70] were generated using hierarchical clustering with Euclidean distance (a) and Pearson correlation (b) as pairwise similarity measure. Although all other parameters are kept constant, results show differences in both gene and cancer cell line groupings. Clustering was performed with the *J-Express Pro* software package (Molmine, Bergen Norway).





6.3
Dynamic Models of Gene Regulation

Summary

In order to comprehend the functioning of organisms at the molecular level, we wish to know which genes are expressed, to what level, where, and when. A network of interactions between DNA, mRNA, proteins, and other molecules realizes the regulation of gene expression. This network comprises many components. According to the central dogma of molecular biology formulated by Francis Crick [71], there is a forward flow of information from gene to mRNA to protein. Moreover, positive and negative feedback loops and information exchange with signaling pathways and energy metabolism ensure the appropriate regulation of expression according to the current state of the cell and the environment.

Modeling of gene expression is used here as an example to apply different modeling techniques. The dynamics and the regulatory patterns of gene expression will be mathematically described with various graphs, Boolean networks, Bayesian networks, ordinary and partial differential equation systems, stochastic processes, and with rule-based formalisms.

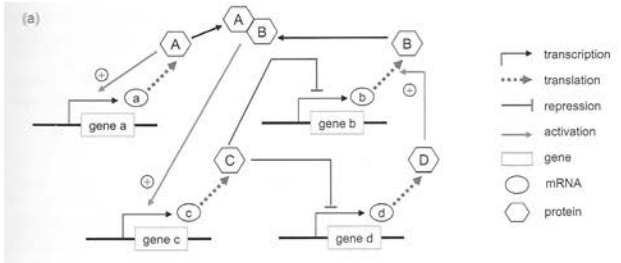
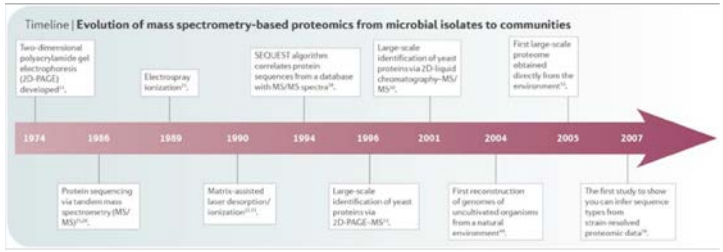


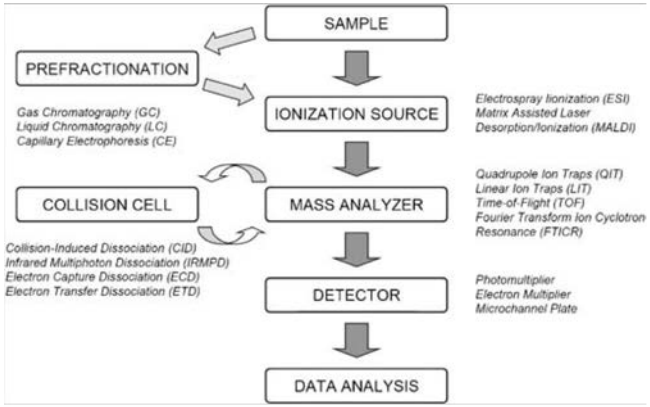
Figure 6.9 Gene regulatory network with four genes and mutual regulation.

Proteome

Mass spectrometry
An analysis technique that identifies biochemical molecules (such as proteins, metabolites or fatty acids) on the basis of their mass and charge.



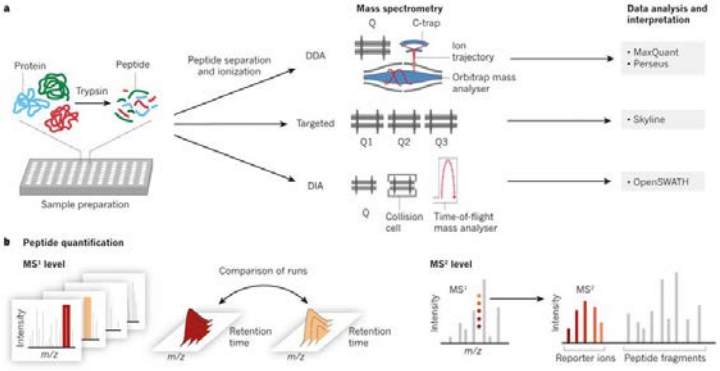
Tandem mass spectrometry
This combines two mass spectrometers: one (MS1) for the detection and selection of precursor ions, which is followed by a second (MS2) for the analysis of fragment ion spectra generated from selected precursor ions after collision-induced fragmentation. The information from the fragment ion spectra is used for peptide identification.

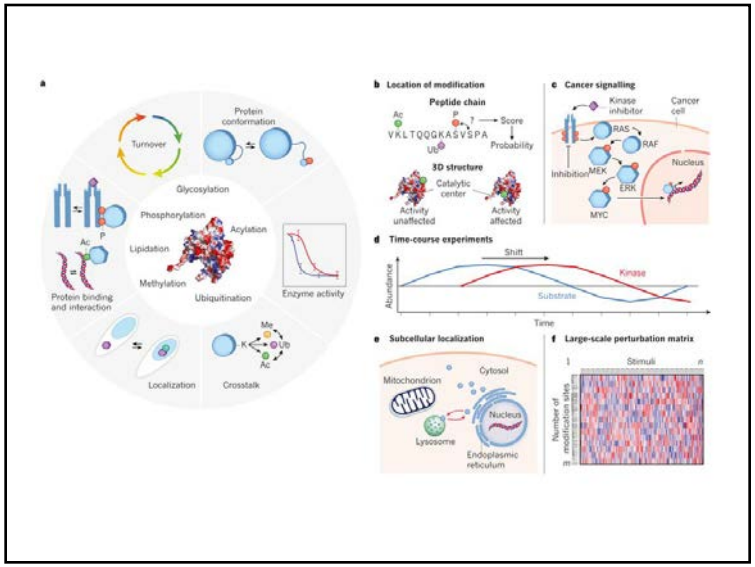


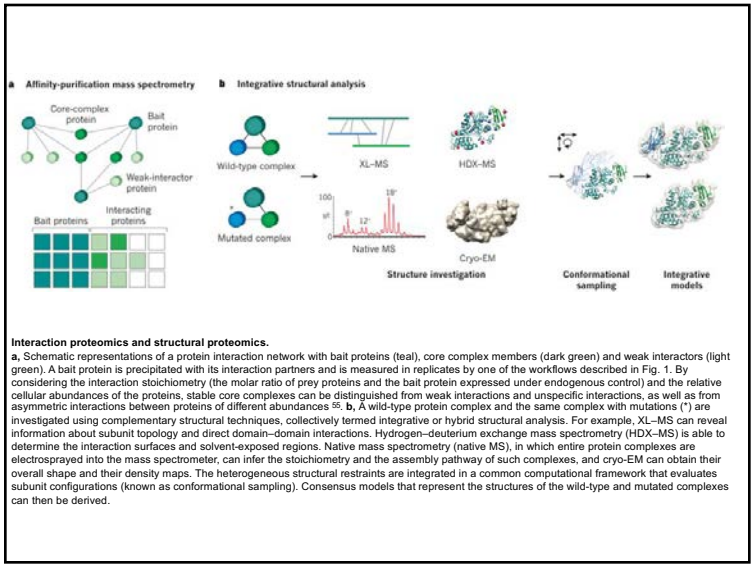
Liquid chromatography–tandem mass spectrometry

Liquid chromatography is used in MS-based proteomics to separate peptides in complex mixtures primarily on the basis of their charge or hydrophobicity using strong cation exchange or reversed-phase chromatography columns.

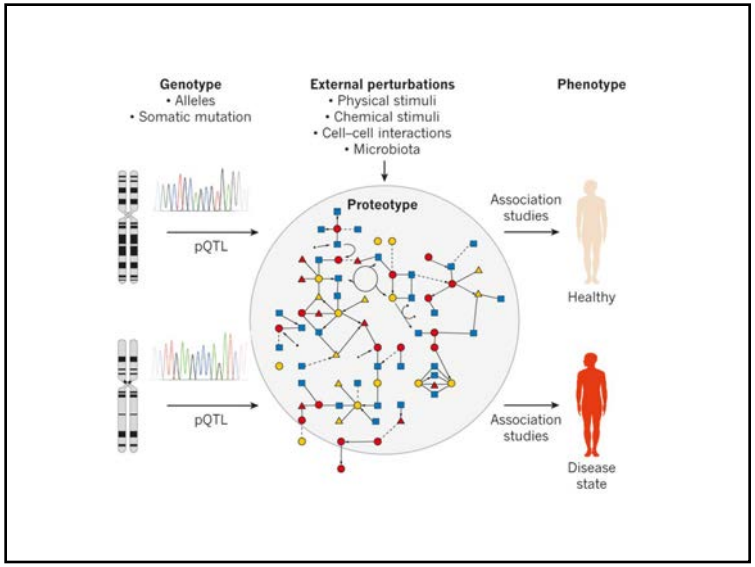
Mass-spectrometric exploration of proteome structure and function.
Aebersold R, Mann M.
Nature. 2016 Sep 15;537(7620):347-55.







Interaction proteomics and structural proteomics.
a, Schematic representations of a protein interaction network with bait proteins (teal), core complex members (dark green) and weak interactors (light green). A bait protein is precipitated with its interaction partners and is measured in replicates by one of the workflows described in Fig. 1. By considering the interaction stoichiometry (the molar ratio of prey proteins and the bait protein expressed under endogenous control) and the relative cellular abundances of the proteins, stable core complexes can be distinguished from weak interactions and unspecific interactions, as well as from asymmetric interactions between proteins of different abundances [55]. **b,** A wild-type protein complex and the same complex with mutations (*) are investigated using complementary structural techniques, collectively termed integrative or hybrid structural analysis. For example, XL–MS can reveal information about subunit topology and direct domain–domain interactions. Hydrogen–deuterium exchange mass spectrometry (HDX–MS) is able to determine the interaction surfaces and solvent-exposed regions. Native mass spectrometry (native MS), in which entire protein complexes are electrosprayed into the mass spectrometer, can infer the stoichiometry and the assembly pathway of such complexes, and cryo-EM can obtain their overall shape and their density maps. The heterogeneous structural restraints are integrated in a common computational framework that evaluates subunit configurations (known as conformational sampling). Consensus models that represent the structures of the wild-type and mutated complexes can then be derived.



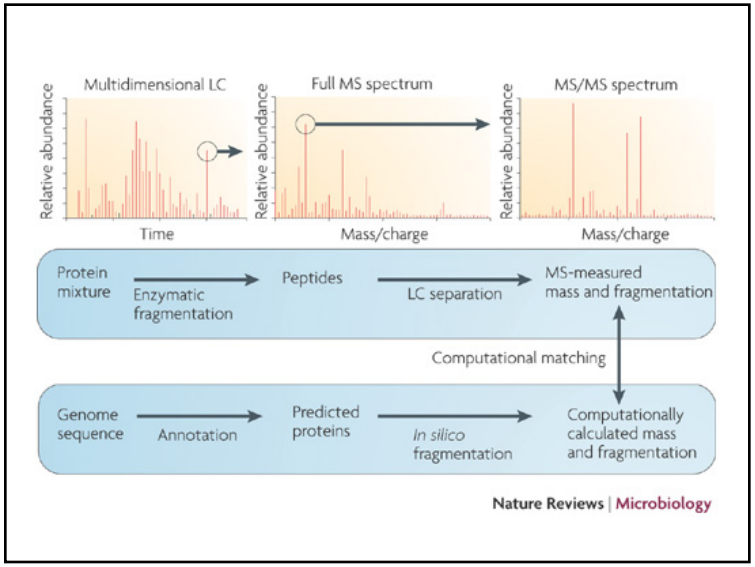
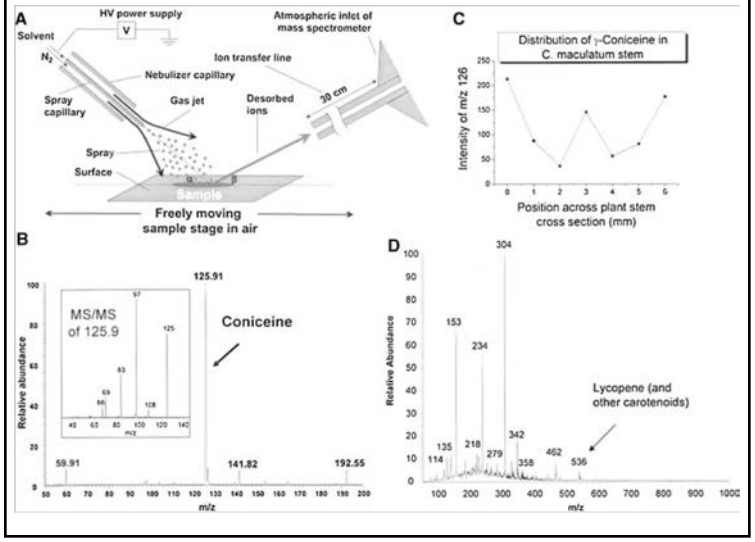
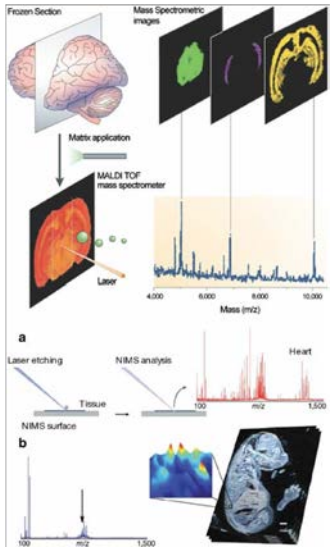
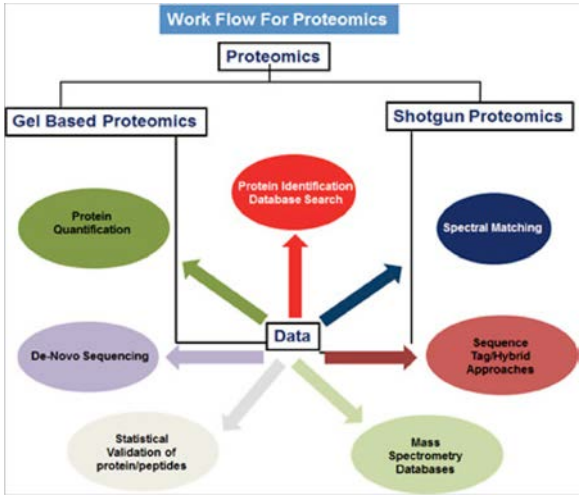


TABLE 1. Potential Cancer Biomarkers Identified by Mass Spectrometry-Based 'Omics' Technologies

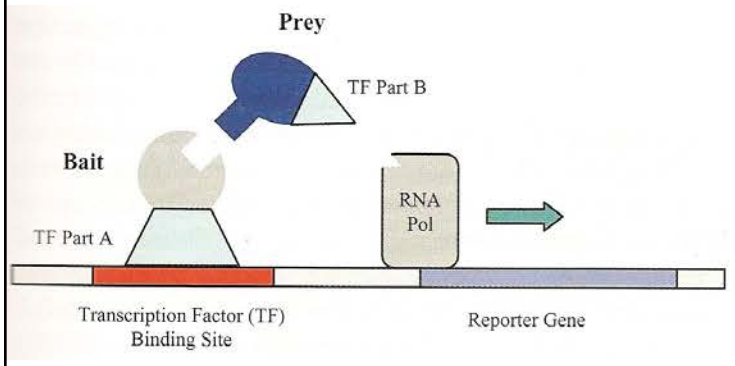
Biomarkers	"omics" platforms	MS methods	Sample source	Cancer type	References
Apolipoprotein A1, Inter-α-trypsin inhibitor Haptoglobin-a-subunit Transthyretin	Proteomics	SELDI-TOF	Serum	Ovarian	Ye et al., 2003; Zhang et al., 2004
Vitamin D-binding protein	Proteomics	SELDI-TOF	Serum	Prostate	Hlavaty et al., 2003
Stathmin (Op18), GRP 78 14-3-3 isoforms, Transthyretin Protein disulfide Isomerase	Proteomics	ESI-MS	Tissue	Lung	Chen et al., 2003
Peroxiredoxin, Enolase Protein disulfide Isomerase	Proteomics	MALDI-TOF, LC-MS	Tissue	Breast	Somiari et al., 2003
HSP 70, α -1-antitrypsin HSP 27	Proteomics	MALDI-TOF	Serum	Liver	Feng et al., 2005
Annexin I, Cofilin, GST Superoxide dismutase Peroxiredoxin, Enolase Protein disulfide Isomerase	Proteomics	MALDI-TOF, ESI-MS, Q-TOF	Tissue	Colon	Seike et al., 2003; Stierum et al., 2003
Neutrophil peptides 1-3	Proteomics	SELDI-TOF	Nipple aspirate fluid	Breast	Li et al., 2005b
PCa-24	Proteomics	MALDI-TOF	Tissue	Prostate	Zheng et al., 2003
Alkanes, Benzenes	Metabonomics	GC-MS	Breath	Lung	Phillips et al., 1999
Decanes, Heptanes	Metabonomics	GC-MS	Breath	Breast	Phillips et al., 2003
Hexanal, Heptanal	Metabonomics	LC-MS	Serum	Lung	Deng et al., 2004
Pseu, m1A, m1I	Metabonomics	HPLC, LC-MS	Urine	Liver	Yang et al., 2005b

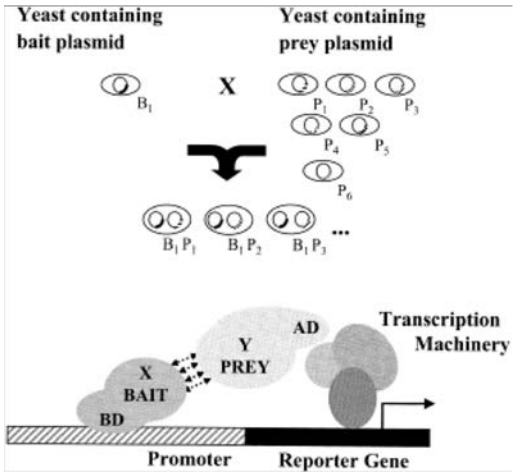


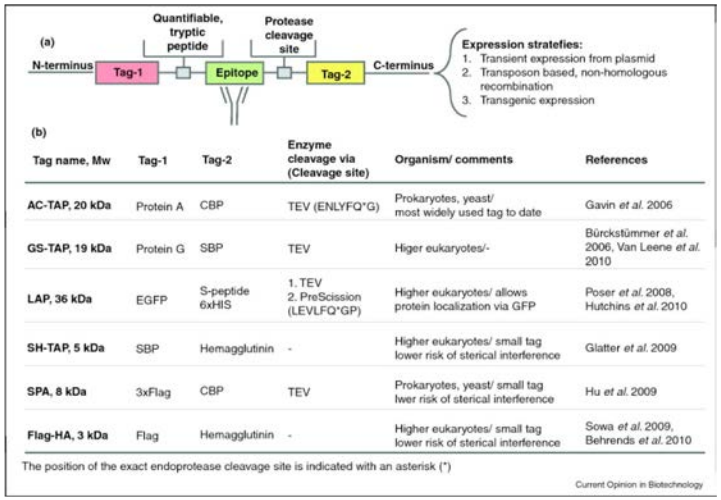




Protein Interactome



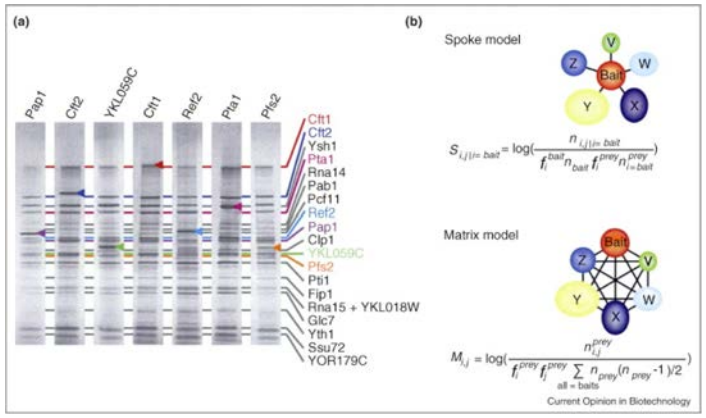




Tag name, Mw	Tag-1	Tag-2	Enzyme cleavage via (Cleavage site)	Organism/ comments	References
AC-TAP, 20 kDa	Protein A	CBP	TEV (ENLYFQ*G)	Prokaryotes, yeast/ most widely used tag to date	Gavin *et al.* 2006
GS-TAP, 19 kDa	Protein G	SBP	TEV	Higer eukaryotes/-	Bürckstümmer *et al.* 2006, Van Leene *et al.* 2010
LAP, 36 kDa	EGFP	S-peptide 6xHIS	1. TEV 2. PreScission (LEVLFQ*GP)	Higher eukaryotes/ allows protein localization via GFP	Poser *et al.* 2008, Hutchins *et al.* 2010
SH-TAP, 5 kDa	SBP	Hemagglutinin	-	Higher eukaryotes/ small tag lower risk of sterical interference	Glatter *et al.* 2009
SPA, 8 kDa	3xFlag	CBP	TEV	Prokaryotes, yeast/ small tag lwer risk of sterical interference	Hu *et al.* 2009
Flag-HA, 3 kDa	Flag	Hemagglutinin	-	Higher eukaryotes/ small tag lower risk of sterical interference	Sowa *et al.* 2009, Behrends *et al.* 2010

The position of the exact endoprotease cleavage site is indicated with an asterisk (*)

Current Opinion in Biotechnology



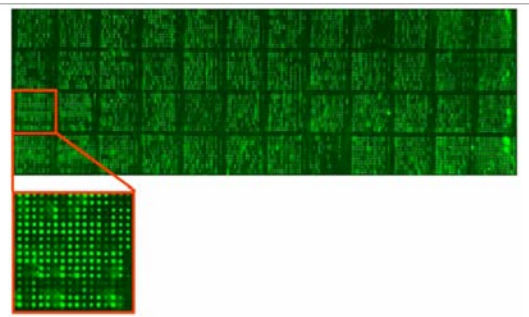
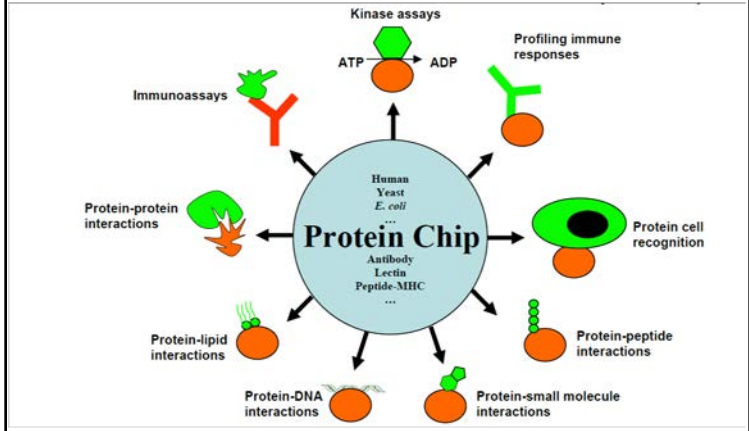
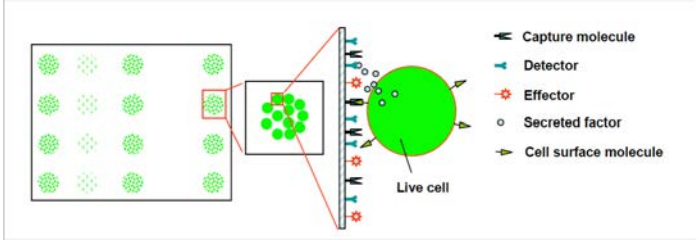


Fig. (1). Image of a typical protein microarray. A yeast protein microarray has been probed with an anti-GST antibody, followed by a Cy3-conjugated secondary antibody. An enlarged image of one of the 48 blocks is shown below the protein chip.





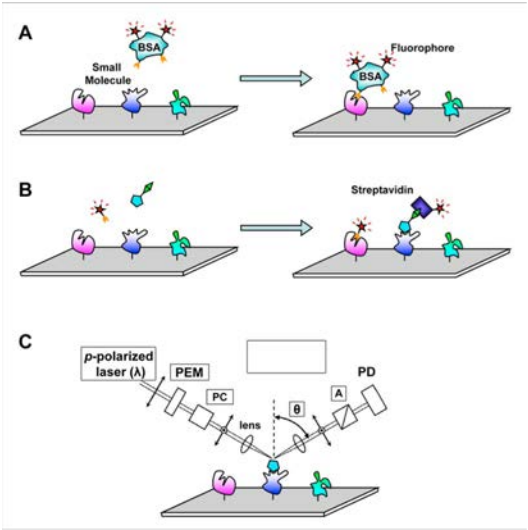
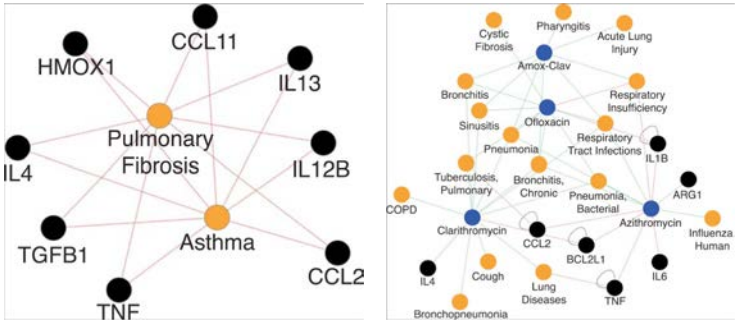


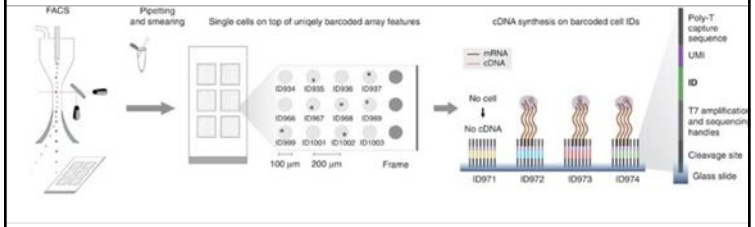
Table 1
Selected pathway/network analysis resource that can benefit Proteomics data analysis.

Name	Description	Link	Reference	Functional info using	Topological info using
GoMiner	Gene ontology (GO) analysis for omic data	http://discover.nci.nih.gov/gominer/	Zeeberg et al. (2003)	Single molecule	Non
KEGG	Kyoto encyclopedia of genes and genomes	http://www.genome.jp/kegg/	Kanehisa and Goto (2000)	Molecular pathway	Non
DAVID	The database for annotation, visualization and integrated discovery	http://david.abcc.ncifcrf.gov/	Dennis et al. (2003)	Molecular pathway	Small-scale
PID	Pathway interaction database	http://pid.nci.nih.gov/	Schaefer et al. (2009)	Cellular pathway	Non
HPD	Human pathway database	http://bio.informatics.iupui.edu/HPD	Chowbina et al. (2009)	Cellular pathway	Small-scale
GESA	Gene set enrichment analysis	http://www.broadinstitute.org/gsea/	Subramanian et al. (2005)	Cellular pathway	Small-scale
IPA	Ingenuity pathway analysis	http://www.ingenuity.com/	N/A	Molecular pathway	Small-scale
MetaCore	Thomson Reuters pathway analysis and knowledge mining	http://thomsonreuters.com/metacore/	N/A	Cellular pathway	Small-scale
Pathway-Express	A systems biology approach for pathway level impact analysis	http://vortex.cs.wayne.edu/projects.htm	Draghici et al. (2007)	Molecular pathway	Mid-Scale
SPIA	Signaling pathway impact analysis	http://www.bioconductor.org/packages/2.12/bioc/html/SPIA.html	Tarca et al. (2009)	Molecular pathway	Mid-Scale
PAGED	An integrated pathway and gene enrichment database	http://bio.informatics.iupui.edu/PAGED	Huang et al. (2012)	System pathway	Mid-Scale
HAPPI	Human annotated and predicted protein interaction database	http://bio.informatics.iupui.edu/HAPPI	Chen et al. (2009)	Single molecule	Large-scale
STRING	Search tool for the retrieval of interacting genes/proteins	http://string.embl.de/	Franceschini et al. (2013)	Single molecule	Large-scale
CytoScape	An open source platform for complex network analysis and visualization	http://www.cytoscape.org/	Shannon et al. (2003)	Molecular pathway	Large-scale
ACOR	Ant colony optimization reordering	N/A	Wu et al. (2009), (2009b), (2009c), (2012)	Molecular pathway	Large-scale
Gene-Terrain	Terrain-based visual analysis for complex networks	N/A	Kim et al. (2001), You et al. (2010)	Network module	Large-scale

Network and matrix analysis of the respiratory disease interactome.
Garcia B, Datta G, Cosgrove GP, Strong M.
BMC Syst Biol. 2014 Mar 22;8:34.



Massive and parallel expression profiling using microarrayed single-cell sequencing.
Vickovic S, Ståhl PL, Salmén F, et al.
Nat Commun. 2016 Oct 14;7:13182.



MASC-Seq overview.
A FACS machine sorts single cells onto a barcoded microarray, printed with six replicates on an activated glass slide. The throughput of the method and microarray design as a 33 × 35 ID matrix is illustrated. An alternative is to pipette and smear cells which then distribute randomly onto the array. Positions of the cells and IDs are noted in a high-resolution image and cDNA is only transcribed when an individual cell lands on top of the barcoded oligo-dTVN primer (ID).

Spring 2015 – Epigenetics and Systems Biology
Lecture Outline (Systems Biology)
Michael K. Skinner – Biol 476/576
Weeks 3 and 4 (January 27/ February 3)

Systems Biology (Components & Technology)

Components (DNA, Expression, Cellular, Organ, Physiology, Organism, Differentiation, Development, Phenotype, Evolution)

Technology (Genomics, Transcriptomes, Proteomics) (Interaction, Signaling, Metabolism)

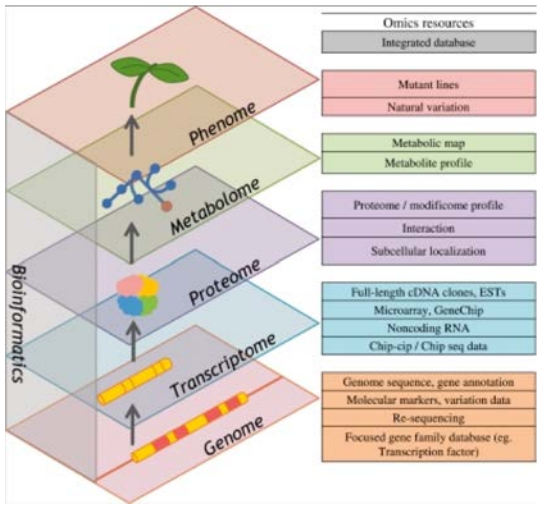
Omics (Data Processing and Resources)

Required Reading

Joyce AR, Palsson BØ. (2006) The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol. Mar;7(3):198-210. Review.

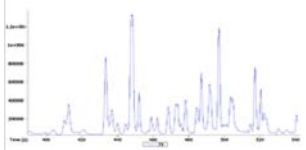
ENCODE (2012) ENCODE Explained. Nature 489:52-55.

Street ME, et al. (2013) Artificial Neural Networks, and Evolutionary Algorithms as a systems biology approach to a data-base on fetal growth restriction. Prog Biophys Mol Biol. 113(3):433-8.



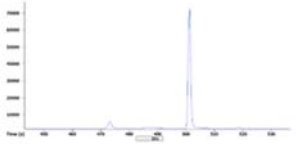
Metabolome

METABOLIC PROFILING

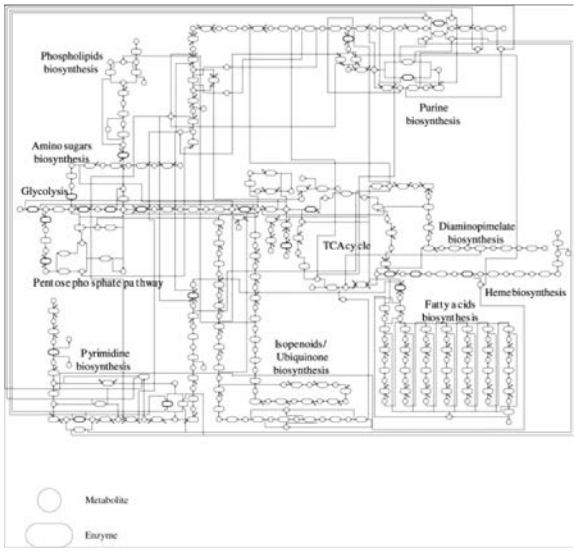


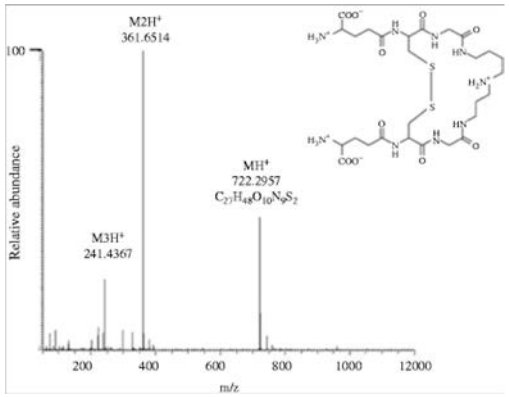
- (semi)-quantitative (global) detection of a wide range of metabolites
- data acquisition without *a priori* knowledge of biologically interesting metabolites
- search for the proverbial 'needles in the metabolic haystack'
- **inductive/hypothesis generating** experiment
- appropriate experimental design and data analysis is essential
- **MIDDLE-IN strategy**

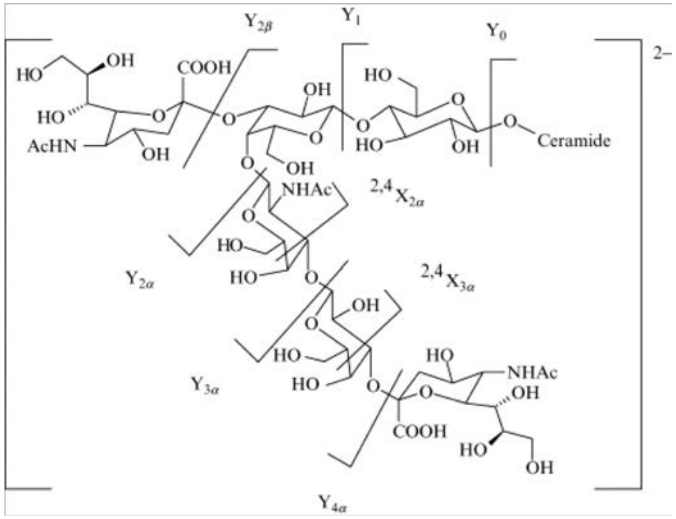
TARGETED ANALYSIS

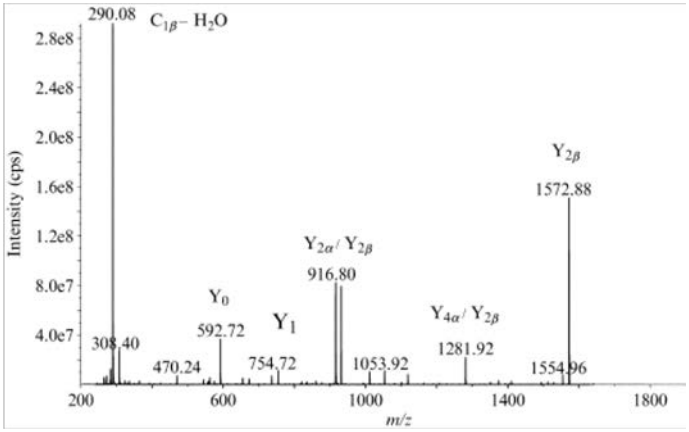


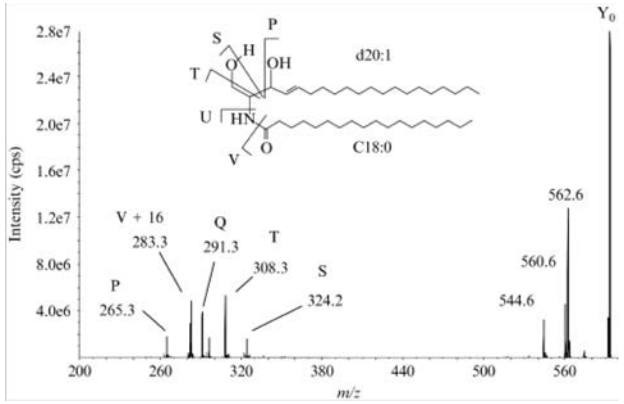
- quantification of a small number of related metabolites for **hypothesis testing** or systems biology modelling
- specificity provided by extensive sample preparation and MS/MS
- absolute quantification using isotopic internal standards or standards addition
- **BOTTOM-UP strategy**

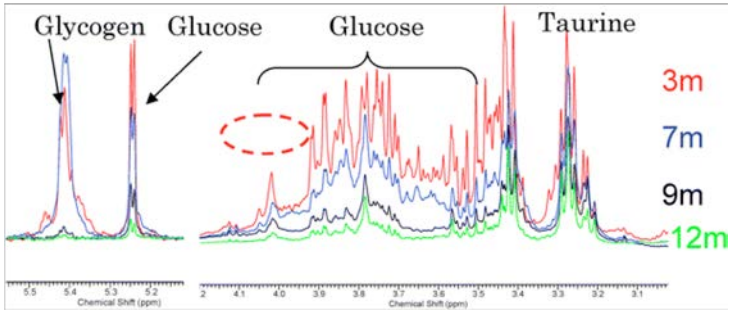












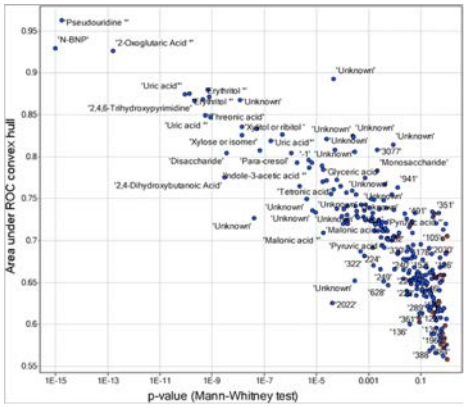


Fig. 13 An example of plots describing the relationship between area under ROC curve and p-values for various metabolites. These plots are applicable when comparing univariate biomarkers or multiple model predictions. The more effective biomarkers approach the top left hand corner of the plot (i.e., low p-value and high AuROC). Kindly reprinted from a study related to heart failure203 with permission from Springer.

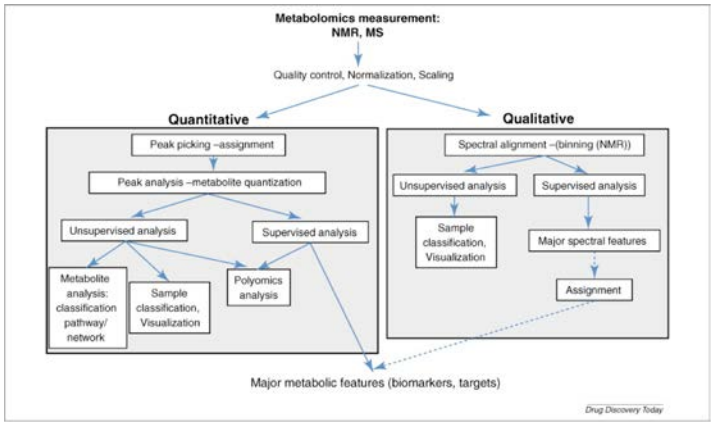


TABLE 1

Comparison of characteristics of major experimental methods for metabolomic analysis.

Analysis	NMR	MS
High throughput – metabolites	No	Medium
High throughput – samples; automation	Yes	No
Quantitative	Yes	Yes
Availability in clinic	No	No
Equipment cost	High	High
Maintenance cost	Medium	High
Per sample cost	Low	High
Required technical skills	Yes	Yes
Sensitivity	Medium	High
Reproducibility	High	Low
Data analysis automation	Yes	Yes
Identification of new metabolites	Difficult	Possible
Chemical exchange analysis	Yes	No
Stereoisomers analysis	Yes	Difficult
Sample preservation	Yes	No
In vivo measurement	Possible	Impossible

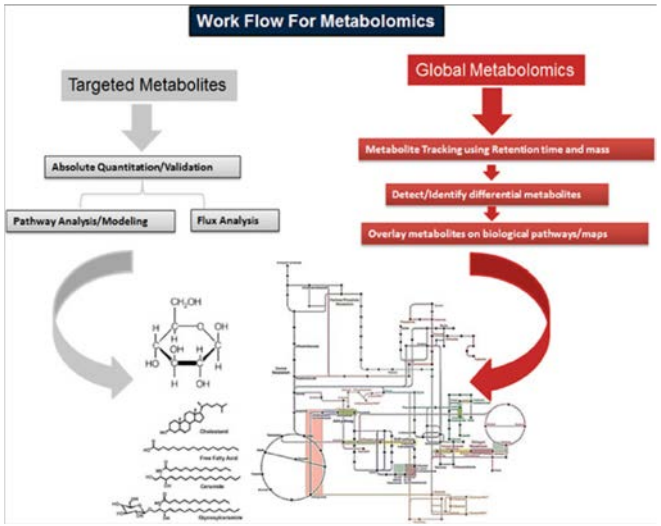
TABLE 2

Some major non-commercial databases of metabolomic standard data for quantification and assignment.

Name and availability	Instrument	Additional information
Human Metabolome Project [81] (http://www.hmdb.ca)	NMR, MS	Biological data; chemical and clinical data specific to humans
BMRD (http://www.bmrb.wisc.edu)	NMR	Database search for NMR peaks assignment
Prime (Akiyama [82]) (http://prime.psc.riken.jp)	MS, NMR	
Golm metabolome database (http://csbdb.mpimp-golm.mpg.de)	MS	Specific to plants
METLIN metabolite database (http://metlin.scripps.edu)	MS	Drug and drug metabolites; specific to humans
NIST Chemistry WebBook (http://Webbook.nist.gov/chemistry)	NMR, MS, IR	
Madison metabolomics database (http://mmcd.nmrfam.wisc.edu)	MS, NMR	
NMR Lab of biomolecules (http://spinportal.magnet.fsu.edu)	NMR	Database search for NMR peaks assignment

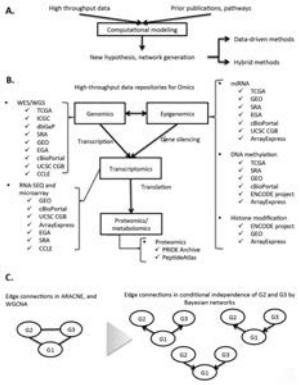
TABLE 3. Summary of medical metabolomics applications cited in this review.

Disease	Species	Material	Method	Approach	Specific biomarker species	Reference
Myocardial ischemia	Human	Plasma	LC/MS/MS	Targeted	Gamma-amino-butyric acid, uric acid, citrate	90,91
Type 2 diabetes	Mouse	Urine	NMR	Targeted	Mannose, 1,5-anhydroglucitol, phenylacetylglutamine	92
Type 2 diabetes	Human	Plasma	GC/MS, LC/MS/MS, NMR	Targeted	3-Indoxyl sulfate, glycerophospholipids, bile acids	93
Obesity	Human	Serum	LC/MS/MS	Non-targeted	Lysophosphatidylcholine	94
Obesity	Human	Serum	MS/MS	Targeted	Phosphatidylcholine	95
Cardiovascular disease	Human	Plasma	LC/MS/MS	Non-targeted	Trimethylamine N-oxide, choline, betaine	96
Ovarian carcinoma	Human	Tumor tissue	GC/MS	Non-targeted	Alpha-glycerolphosphate, uracil, glycine	97
Lung cancer	Human	Tissue, plasma	GC/MS, NMR	Stable isotope resolved analysis	(^{13}C-enrichment in lactate, alanine, succinate)	98
Pancreatic cancer	Human	Serum	GC/MS	Targeted	Thiodiglycolic acid, lactic acid, 7-hydroxyoctanoic acid	99
Hepatocellular carcinoma	Human	Urine	GC/MS	Non-targeted	Xylitol, urea, hydroxy proline dipeptide	100
Colorectal cancer	Human/rat	Urine/tissue	GC/MS	Targeted	Succinate, N-acetyl-aspartate, 2-hydroxyhippurate	101
Oral cancer	Human	Saliva	CE/MS	Non-targeted	Pyrroline, leucine + isoleucine, taurine	102
Breast cancer	Human	Saliva	CE/MS	Non-targeted	Taurine, putrescine, leucine + isoleucine	102
Pancreatic cancer	Human	Saliva	CE/MS	Non-targeted	Leucine + isoleucine, phenylalanine, alpha-amino butyric acid	102
Schizophrenia	Human	Cerebrospinal fluid	NMR	Non-targeted	Lactate, citrate, glucose	103
Parkinson's disease	Human	Plasma	NMR	Targeted	Threonate, myoinositol, suberate	104
Huntington's disease	Human/mouse	Serum	GC/MS	Non-targeted	Glycerol, urea, valine	105
Schizophrenia	Human	Plasma	LC/MS/MS, GC/MS	Targeted	Free fatty acids, triglycerides, phosphatidylethanolamine	106
Depression	Rat	Plasma	GC/MS	Targeted	Glucose, glutamine, butanedioic acid	107



Databases

Databases and tools for constructing signal transduction networks in cancer.
BMB Rep. 2017 Jan;50(1):12-19.
Nam S.



Systems biology, databases, and network generation. (A) The diversity of types of high-throughput data (genomics, epigenomics, transcriptomics, proteomics, metabolomics) available. The relationships among the data types are connected by edges. (B) The flow (represented by "edges") of genetic information from DNA to protein is aligned with the diverse data types. Public repositories corresponding to each data type are listed (further description in Table 1). (C) Network differences between correlation-based approaches and Bayesian networks approaches. The correlation (or mutual information) oriented tools, ARACNE (39) and WGCNA (36), do not report directions of edges in networks. Bayesian-driven networks naturally reveal directed edges among the network entries. In other words, the undirected network (in left of the grey-shaded triangular) having G1, G2, and G3 entries by ARACNE and WGCNA can be differentiated into directed networks (in the right of the grey-shaded triangular), using Bayesian networks tools (48–51).

16
Databases

Summary

With the rapid increase of biological data, it has become even more important to organize and structure the data in a way so that information can easily be retrieved. As a result, the number of databases has also increased rapidly over the past few years. Most of these databases have a web interface and can be accessed from everywhere in the world, which is an enormously important service for the scientific community. In the following, various databases are presented that might be relevant for systems biology.

Moreover, the journal *Nucleic Acids Research* offers a database issue each year in January dedicated to factual biological databases and in addition to this a web server issue each year in July presenting web-based services.

Databases Sources

- National Center for Bioinformatics NCBI
- European Bioinformatics Institute
- EMBL
- Ensembl
- Interpro
- Protein databank
- Bionumbers
- Gene Ontology
- Pathway- KEGG
- Consensus Path DB

Omics data set

A generic term that describes the genome-scale data sets that are emerging from high-throughput technologies. Examples include whole-genome sequencing data (genomics) and microarray-based genome-wide expression profiles (transcriptomes).

Data mining

An analytical discipline that is focused on finding unsuspected relationships and summarizing often large observational data sets in new ways that are both understandable and useful to the data owner.

***In silico* prediction**

A general term that refers to a computational prediction that usually results from the analysis of a mathematical or computational model.

Unsupervised analysis: Unsupervised analysis includes methods used for grouping of features (sample, metabolites and spectral features) according to the molecular data measured. These methods are used for the analysis of features when no prior information is available about the system. Depending on the method, the analysis might or might not require the user to define the number of clusters. In terms of cell culture metabolomics, this method is ideal for discovery of novel classes.

Supervised analysis: Supervised analysis defines methods for sample grouping or classification and for selection of major sample defining features. In supervised analysis, a set of features is pre-assigned to a class and it is used as a training set for the method of choice to define a classifier that will be used for classification of an unknown sample. Supervised analysis creates a model from the training set and, thus, can only be accurately used for classification of a different dataset (i.e. supervised analysis requires application of cross-validation for the determination of accuracy of the classifier).

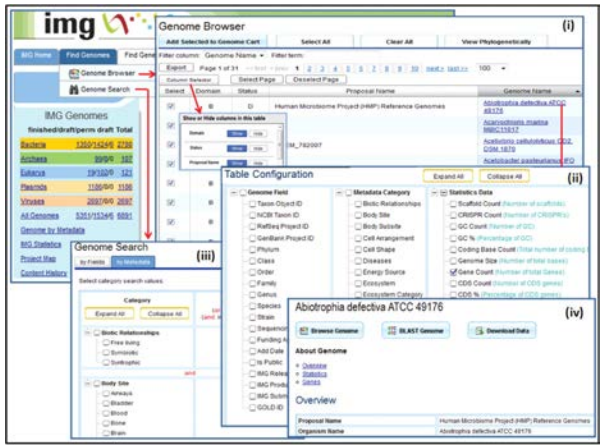
Table 1 | **'Omics' data repositories***

Data types	Online resource	Description	URL
Components			
Genomics	Genomes OnLine Database (GOLD)	Repository of completed and ongoing genome projects	http://www.genomesonline.org
Transcriptomics	Gene Expression Omnibus (GEO)	Microarray and SAGE-based genome-wide expression profiles	http://www.ncbi.nlm.nih.gov/geo
	Stanford Microarray Database (SMD)	Microarray-based genome-wide expression data	http://genome-www.stanford.edu/microarray
Proteomics	World-2DPAGE	Links to 2D-PAGE data	http://us.expasy.org/ch2d/2d-index.html
	Open Proteomics Database (OPD)	Mass-spectrometry-based proteomics data	http://bioinformatics.icmb.utexas.edu/OPD
Lipidomics	Lipid Metabolites and Pathways Strategy (LIPID MAPS)	Genome-scale lipids database	http://www.lipidmaps.org
Localizomics	Yeast GFP Fusion Localization Database	Yeast genome-scale protein-localization data	http://yeastgfp.ucsf.edu
Interactions			
Protein–DNA	Biomolecular Network Database (BIND)	Published protein–DNA interactions	http://www.bind.ca/Action/
	Encyclopedia of DNA Elements (ENCODE)	Database of functional elements in human DNA	http://genome.ucsc.edu/ENCODE/index.html
Protein–protein	Munich Information Center for Protein Sequences (MIPS)	Links to protein-protein-interaction data and resources	http://mips.gsf.de/proj/ppi
	Database of Interacting Proteins (DIP)	Published protein–protein interactions	http://dip.doe-mbi.ucla.edu
Functional states			
Phenomics	RNAi database	*C. elegans* RNAi screen data	http://mai.org
	General Repository for Interaction Datasets (GRID)	Synthetic-lethal interactions in yeast	http://biodata.mshri.on.ca/grid
	A Systematic Annotation Package For Community Analysis of Genomes (ASAP)	Single-gene-deletion microarray data for *E. coli* phenotypes	http://www.genome.wisc.edu/tools/asap.htm

*This table details some of the databases that store and distribute genome-scale omics data sets through publicly accessible Web sites. Some omics technologies do not yet have associated data-dissemination resources — notably metabolomics, glycomics and fluxomics — and are therefore not included in this table. It should also be noted that this table does not represent all publicly available omics data resources, but, rather, provides a reasonably broad sample of the data that are readily accessible to researchers today. *C. elegans*, *Caenorhabditis elegans*; 2D-PAGE, two-dimensional polyacrylamide-gel electrophoresis; *E. coli*, *Escherichia coli*; GFP, green fluorescent protein; RNAi, RNA interference; SAGE, serial analysis of gene expression.

Table 1: Useful online resources for systems biology and modeling of the human microbiome

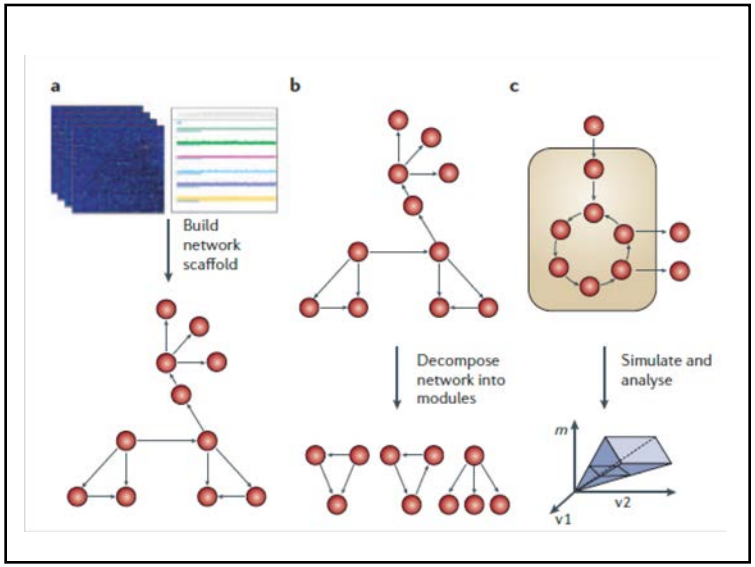
Resources	References
Microbial genomic data and analysis	
IMG	[80]
DACC	[81]
GOLD	[31]
Microbes online	[82]
RAST	[83]
Metagenomic data and analysis	
IMG/M	[84]
MG-RAST	[85]
METAREP	[86]
Metabolic databases	
KEGG	[23]
MetaCyc	[24]
Brenda	[87]
Metabolic model reconstruction, visualization and analysis	
The Model Seed	[34]
Systems Biology Research Group	[88]
iPath	[89]
Pathway Tools	[90]
Cytoscape	[91]
Cobra	[92]
Reverse ecology software	
NetSeed	[44]



Networks

Network scaffold
Refers to the structure of a network that specifies the components of the network and the interactions between them, and represents the end product of the network-reconstruction process.

Network module
A portion of a biological network that is composed of multiple molecular entities (such as genes, proteins or metabolites) that work together as a distinct unit within the cell, for example, in response to certain stimuli or as part of a developmental or differentiation programme.



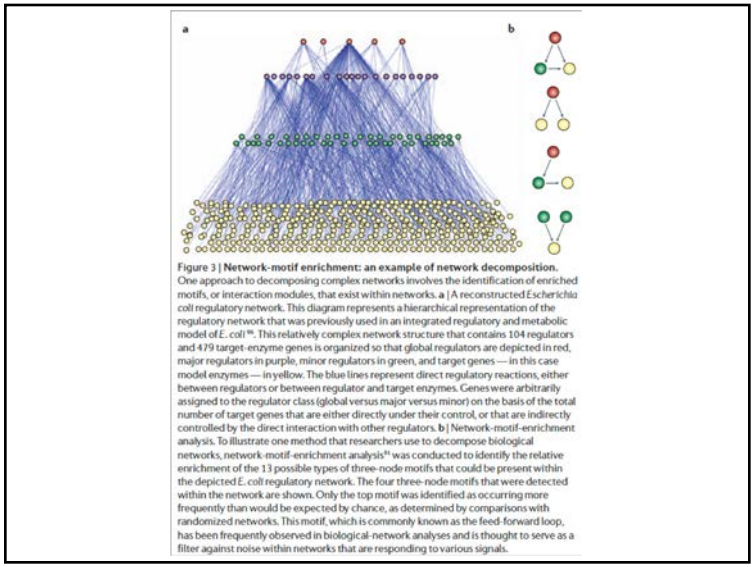
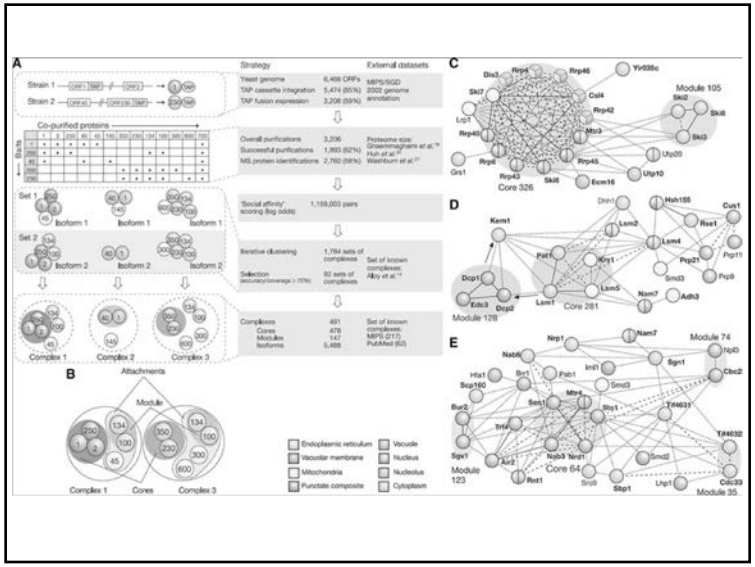


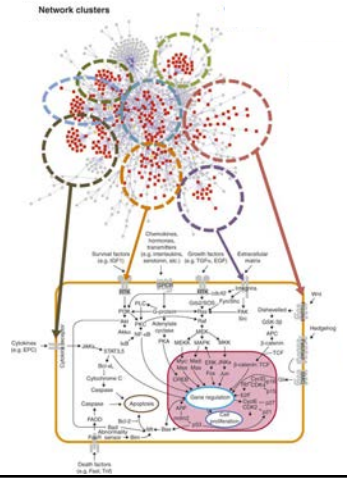
Figure 3 | **Network-motif enrichment: an example of network decomposition.** One approach to decomposing complex networks involves the identification of enriched motifs, or interaction modules, that exist within networks. **a** | A reconstructed *Escherichia coli* regulatory network. This diagram represents a hierarchical representation of the regulatory network that was previously used in an integrated regulatory and metabolic model of *E. coli* [86]. This relatively complex network structure that contains 104 regulators and 479 target-enzyme genes is organized so that global regulators are depicted in red, major regulators in purple, minor regulators in green, and target genes — in this case model enzymes — in yellow. The blue lines represent direct regulatory reactions, either between regulators or between regulator and target enzymes. Genes were arbitrarily assigned to the regulator class (global versus major versus minor) on the basis of the total number of target genes that are either directly under their control, or that are indirectly controlled by the direct interaction with other regulators. **b** | Network-motif-enrichment analysis. To illustrate one method that researchers use to decompose biological networks, network-motif-enrichment analysis[81] was conducted to identify the relative enrichment of the 13 possible types of three-node motifs that could be present within the depicted *E. coli* regulatory network. The four three-node motifs that were detected within the network are shown. Only the top motif was identified as occurring more frequently than would be expected by chance, as determined by comparisons with randomized networks. This motif, which is commonly known as the feed-forward loop, has been frequently observed in biological-network analyses and is thought to serve as a filter against noise within networks that are responding to various signals.



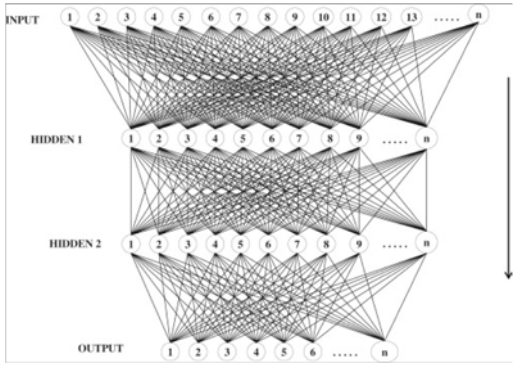
Network reconstruction

The process of integrating different data sources to create a representation of the chemical events that underlie a biochemical reaction network.

Network analysis: a new approach to study endocrine disorders.
Stevens A, et al.
J Mol Endocrinol. 2013 Dec 19;52(1):R79-93.



Artificial Neural Networks, and Evolutionary Algorithms as a systems biology approach to a data-base on fetal growth restriction.
Street M, et al.
Prog Biophys Mol Biol. 2013 Dec;113(3):433-8.

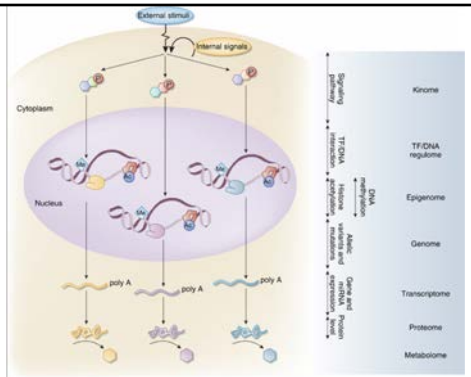


Typical neural network architecture. The basic elements of ANN are the nodes, also called processing elements (PE), and their connections. Each node has its own input, from which it receives communications from other nodes and/or from the environment and its own output, from which it communicates with other nodes or with the environment. Finally, each node has a function through which it transforms its own global input into an output.

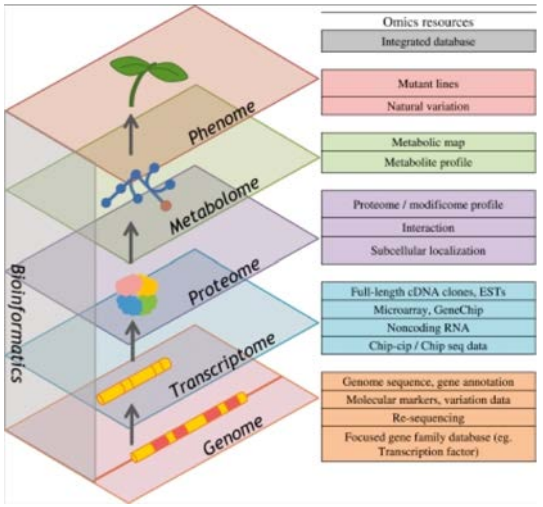
Omics Data Integration

Omics data integration

The simultaneous analysis of high-throughput genome-scale data that is aimed at developing models of biological systems to assess their properties and behavior.



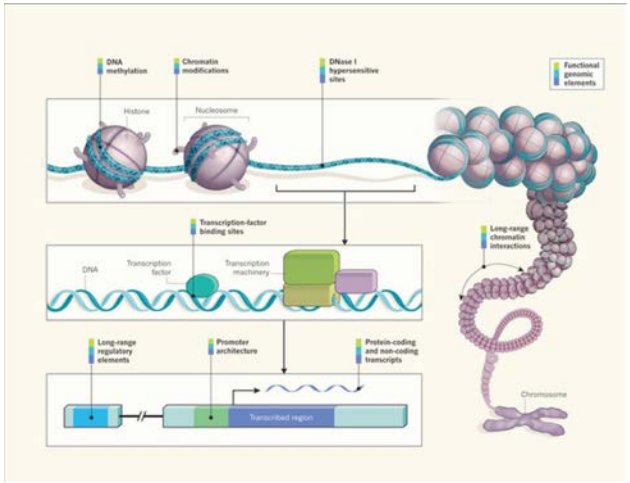
'Omic' approaches to profile cellular processes. Both external stimuli and internal signals can initiate the activation of downstream signaling pathways through post-transcriptional modification of signaling molecules, such as protein kinases, by phosphorylation. Activation of the signaling cascade leads to either transcriptional activation by recruitment and assembly of the transcriptional complex on the DNA, transcription factor/DNA (TF/DNA) interaction, or inhibition of transcriptional activation by exclusion of key TFs from the complex. Covalent modifications of the DNA by methylation and of histones by acetylation have a major role in the regulation of gene expression. Variations in the coding region of the genome result in mutations of the gene transcripts, and changes in non-coding region could affect the regulation of gene expression. Gene transcripts as messenger RNAs in the cytoplasm are translated into proteins. The protein products synthesized in response to the initial signal serve to maintain cellular integrity and react to perturbations in the systems by initiating additional signals or catalyze reactions that generate specific metabolites as a by-product. Characterization of these processes at the systems level is known as the 'omics' approaches: kinome for phosphorylation of proteins, TF/DNA regulome for regulation of transcription by the interaction of TF with DNA, epigenome for modification of histones and DNA, genome for the sequence of DNA, transcriptome for the mRNA transcripts of the expressed genes, proteome for the protein composition, and metabolome for the metabolites that are generated in a specific tissue or cell (Red P=phosphorylation, Me=methylation, Ac=acetylation, line with poly A tail=mRNA). miRNA, microRNA.

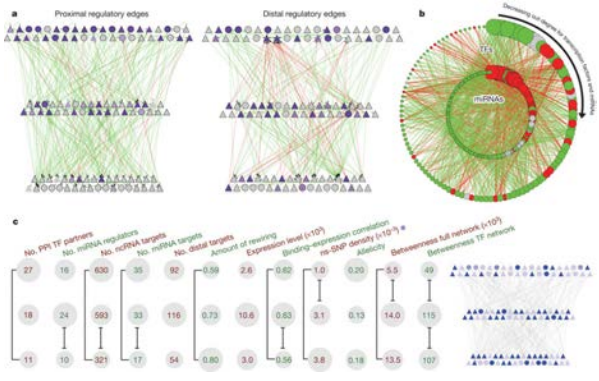


FORUM: Genomics

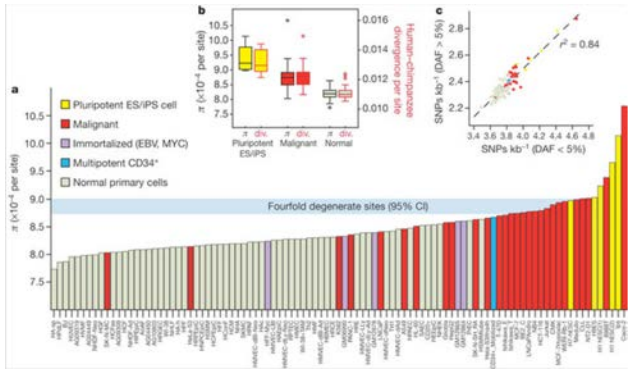
ENCODE explained

The Encyclopedia of DNA Elements (ENCODE) project dishes up a hearty banquet of data that illuminate the roles of the functional elements of the human genome. Here, five scientists describe the project and discuss how the data are influencing research directions across many fields. SEE ARTICLES P.57, P.75, P.83, P.91, P.101 & LETTER P.109

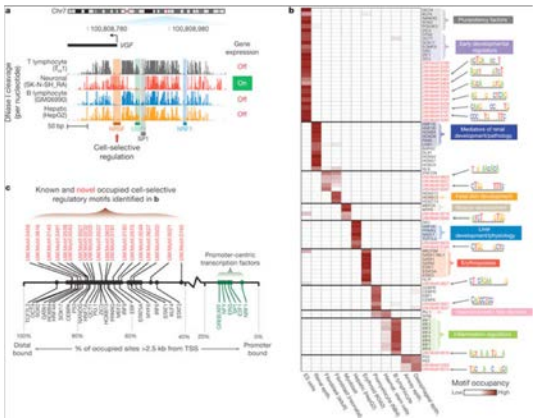




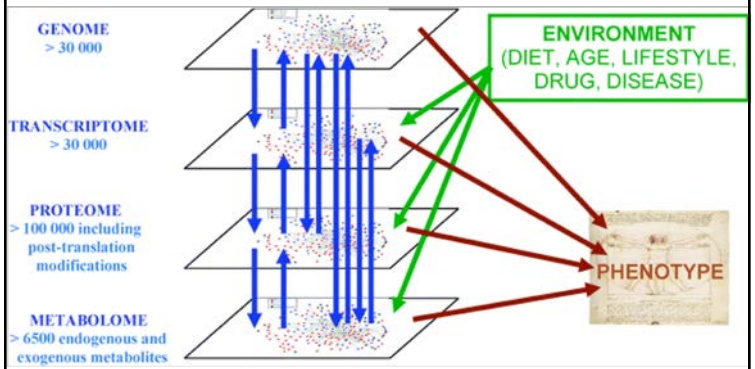
Overall network. a, Close-up representation of the transcription factor hierarchy. Nodes depict transcription factors. TFSSs are triangles, and non-TFSSs are circles. Left: proximal-edge hierarchy with downward pointing edges coloured in green and upward pointing ones coloured in red. The nodes are shaded according to their out-degree in the full network (as described in Table 1). Right: factors placed in the same proximal hierarchy but now with edges corresponding to distal regulation coloured green and red, and nodes re-coloured according to out-degree in the distal network. The distal edges do not follow the proximal-edge hierarchy. **b**, Close-up view of transcription-factor–miRNA regulation. The outer circle contains the 119 transcription factor, whereas the inner circle contains miRNAs. Red edges correspond to miRNAs regulating transcription factors; green edges show transcription factors regulating miRNAs. Transcription factors and miRNAs each are arranged by their out-degree, beginning at the top (12:00) and decreasing in order clockwise. Node sizes are proportional to out-degree. For transcription factors, the out-degree is as described in Table 1; for miRNAs, it is according to the out-degree in this network. Red nodes are enriched for miRNA–transcription factor edges and green nodes are enriched for transcription factor–miRNA edges. Grey nodes have a balanced number of edges (within ±1). **c**, Average values of various properties (topological, dynamic, expression-related and selection-related—ordered consistently with Table 1) for each level are shown for the proximal-edge hierarchy. The top, middle and bottom rows correspond to the top, middle and bottom of the hierarchy, respectively. The sizing of the grey circles indicates the relative ordering of the values for the three levels. Significantly different values ($P < 0.05$) using the Wilcoxon rank-sum test are indicated by black brackets. The proximal-edge hierarchy depicted on the right shows non-synonymous SNP (ns-SNP) density, where the shading corresponds to the density for the associated factor. (See Supplementary Fig. 4 for more details.)



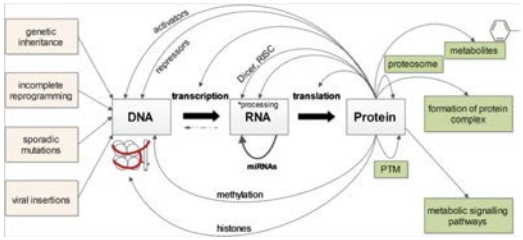
Genetic variation in regulatory DNA linked to mutation rate. a, Mean nucleotide diversity (π, y axis) in DHSs of 97 diverse cell types (x axis) estimated using whole-genome sequencing data from 53 unrelated individuals. Cell types are ordered left-to-right by increasing mean π. Horizontal blue bar shows 95% confidence intervals on mean π in a background model of fourfold degenerate coding sites. Note the enrichment of immortal cells at right. ES, embryonic stem; iPS, induced pluripotent stem. **b**, Mean π (left y axis) for pluripotent (yellow) versus malignancy-derived (red) versus normal cells (light green), plotted side-by-side with human–chimpanzee divergence (right y axis) computed on the same groups. Boxes indicate 25–75 percentiles, with medians highlighted. **c**, Both low- and high-frequency derived alleles show the same effect. Density of SNPs in DHSs with derived allele frequency (DAF) <5% (x axis) is tightly correlated ($r2 = 0.84$) with the same measure computed for higher-frequency derived alleles (y axis). Colour-coding is the same as in panel a.

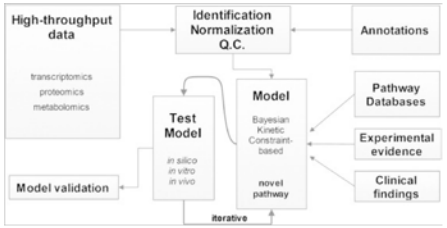


Multi-lineage DNase I footprinting reveals cell-selective gene regulators. a, Comparative footprinting of the nerve growth factor gene (*VGF*) promoter in multiple cell types reveals both conserved (NRF1, USF1 and SP1) and cell-selective (NRSF) DNase I footprints. **b,** Shown is a heat map of footprint occupancy computed across 12 cell types (columns) for 89 motifs (rows), including well-characterized cell/tissue-selective regulators, and novel *de novo*-derived motifs (red text). The motif models for some of these novel *de novo*-derived motifs are indicated next to the heat map. **c,** The proportion of motif instances in DNase I footprints within distal regulatory regions for known (black) and novel (red) cell-type-specific regulators in b is indicated. Also noted are these values for a small set of known promoter-proximal regulators (green). ES, embryonic stem.

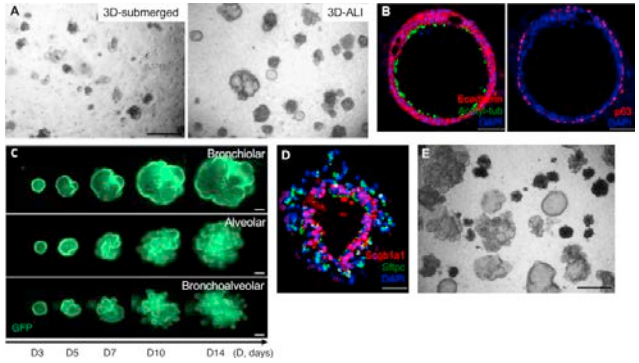


Current advances in systems and integrative biology.
Robinson SW, Fernandes M, Husi H.
Comput Struct Biotechnol J. 2014 Aug 27;11(18):35-46.

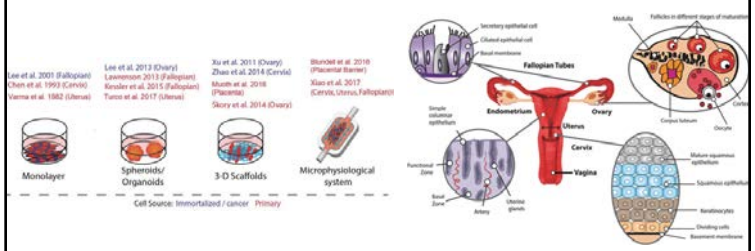




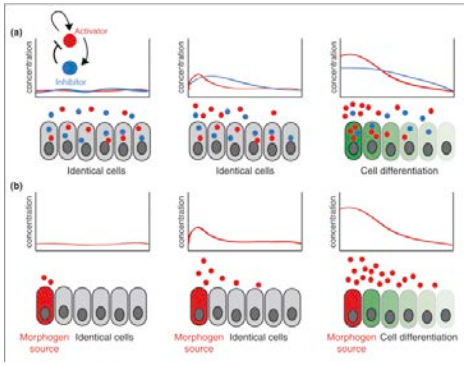
Organogenesis of adult lung in a dish: Differentiation, disease and therapy.
Choi J, Iich E, Lee JH.
Dev Biol. 2016 Dec 15;420(2):278-286.



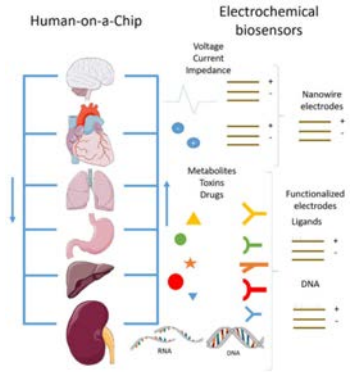
Organ-on-a-Chip Systems for Women's Health Applications.
Adv Healthc Mater. 2018 Jan;7(2).
Nawroth J, Rogal J, Weiss M, Brucker SY, Loskill P.



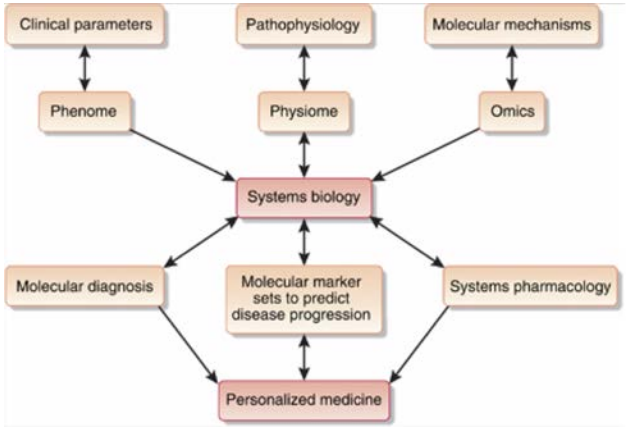
Self-organization in development, regeneration and organoids.
Curr Opin Cell Biol. 2017 Feb;44:102-109.
Werner S, Vu HT, Rink JC.



Organs-on-a-Chip Module: A Review from the Development and Applications Perspective.
Sosa-Hernández JE, Villalba-Rodríguez AM, Romero-Castillo KD, et al.
Micromachines (Basel). 2018 Oct 22;9(10).

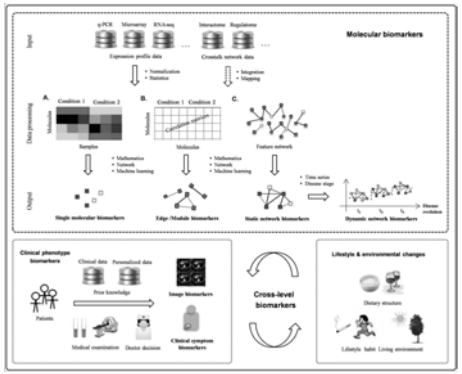


Systems Biology and Medicine



Personalized genomic medicine
The idea that genome-scale technologies will allow clinicians to apply treatment regimens that are tailored specifically to an individual patient on the basis of their genetic makeup and associated predispositions.

Computer-aided biomarker discovery for precision medicine: data resources, models and applications.
Brief Bioinform. 2017 Nov 29. doi: 10.1093/bib/bbx158. [Epub ahead of print]
Lin Y, Qian F, Shen L, Chen F, Chen J, Shen B.



Biomarker classification based on systems biology viewpoints. Molecular biomarkers described here can be categorized into three subtypes: single molecular biomarkers, edge/module biomarkers and network biomarkers. The integration of molecular biomarkers, clinical phenotype biomarkers and lifestyle/environmental factors constitutes the concept of cross-level biomarkers, which are of great significance for precision medicine. q-PCR: quantitative polymerase chain reaction.

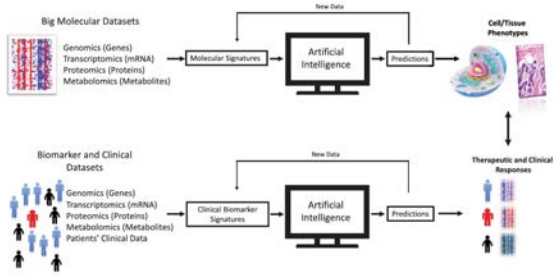
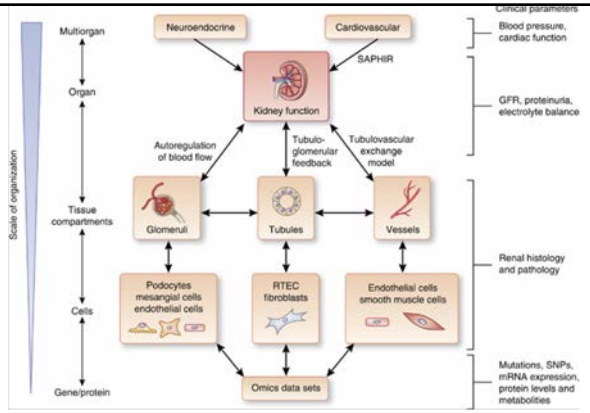
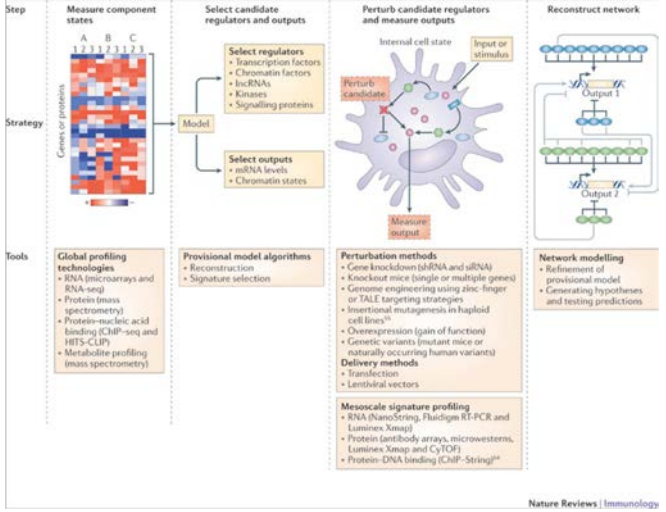
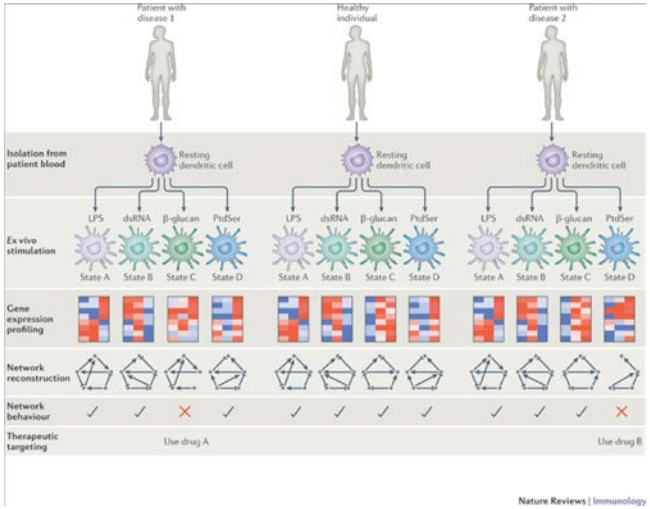


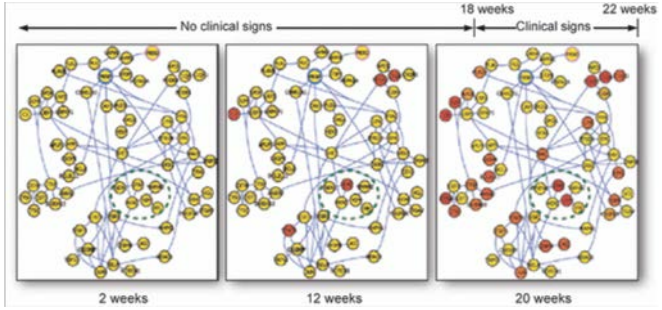
Figure 3. AI in systems biology and precision medicine
AI and machine learning approaches are helpful tools for finding molecular signatures from high-throughput measurement datasets and assigning them to higher level phenotypes and cellular functions. They also can be used to tailor treatment based on an individual patient's molecular markers for precision medicine. Clinical data and omics data from patients are used to extract clinical biomarkers for training sets and to build predictive models of the course of the disease and patient responses to treatment. Verified predictions are used as additional training sets to make the predictive models progressively more accurate.



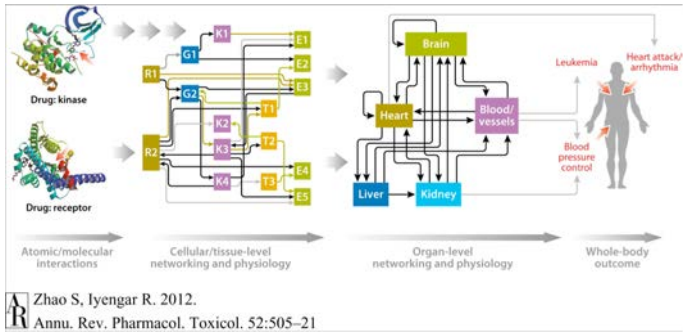
Multiscale analysis of kidney function. Maintenance of renal function requires the coordinated regulation from other organ systems (neuroendocrine and cardiovascular) and various tissue compartments and cells within the kidney. To recapitulate normal renal physiology in biological models, this multiscale organization from organ systems down to cell/gene level will need to be determined. The interactions at several levels have been described: Systems Approach for Physiological Integration of Renal, Cardiac, and Respiratory (SAPHIR) models, as well as models of autoregulation of glomerular blood flow, tubuloglomerular feedback, and tubulovascular exchange. Clinical parameters that we can use to assess and infer the function of organ systems and organs include blood pressure and cardiac function for cardiovascular and neuroendocrine input into the kidney, glomerular filtration rate (GFR), and proteinuria as determinants of the filtration function of the glomeruli, balance of electrolytes as an indicator of tubular function, podocyte number, foot process effacement, mesangial deposition, glomerulosclerosis, and tubulointerstitial fibrosis on renal histology as indicators of disease severity, and genetic variations (mutations, single-nucleotide polymorphisms (SNPs)) and mRNA and protein expression levels as indices of cellular response to internal and external stimuli. RTEC, renal tubular epithelial cell.







A schematic of the network perturbations of one neural degenerative network over the 20 weeks of the progression of this disease in a mouse model. The red nodes indicate mRNAs that have become disease perturbed as compared with the brain transcripts of normal mice. The spreading of the disease-perturbed networks at the three different times points is striking – indicating the progressive disease perturbation of this neurodegenerative network.



Zhao S, Iyengar R. 2012.
Annu. Rev. Pharmacol. Toxicol. 52:505–21

Application system biology

Systems Biology Approaches to a Rational Drug Discovery Paradigm.
Prathipati P, Mizuguchi K.
Curr Top Med Chem. 2016;16(9):1009-25.

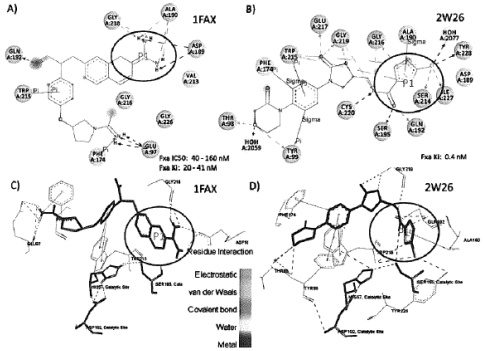


Fig. (1). The advantages of structure-based approaches augmenting ligand-based approaches are exemplified using FXa inhibitors Dabigatran (A and B) and Rivaroxaban (C and D). Using a ligand-based approach, a basic fragment at P1 site (A and B) was shown to be essential for FXa inhibition. Hence, the optimization strategy involved exploring other basic fragments at this site, which led to unfavorable ADMET profiles. A structure-based approach was used for replacing the basic amidine fragment of Dabigatran at the P1 pocket with a neutral chlorothiophene fragment by exploiting FXa's binding site information and the full range of intermolecular interactions such pi-pi, cation-pi and anion-pi, in addition to the usual hydrogen bonding, hydrophobic and salt bridge interactions. This strategy has led to the design of the new inhibitor Rivaroxaban (C and D) with favorable ADMET profiles. For more details, the readers can refer to a review by Nar *et al.* [4].

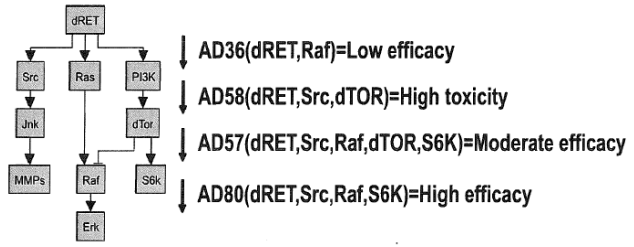


Fig. (2). The cancer model built using the fruit fly Drosophila pathway† data was used to rationally arrive at a compound (AD80) that was shown to inhibit four of the 10 selected cancer targets, dRET, Src, Raf and S6K. This chemoproteomic profile (inhibiting these four targets out of the 10) was associated with high efficacy and low toxicity in whole animal screening. Further SAR analysis also led to the identification of dTOR as an anti-target responsible for toxicity. AD80 proved far more effective and less toxic than standard cancer drugs, which generally focus on a single target. This study by Dar *et al.*[6] was the first time that whole-animal screening has been used in a rational, step-wise approach to identifying favorable chemoproteomic profiles and laid the case for a rational systems biology approach to drug discovery. For a more general discussion of chemoproteomics, see section 3.5.

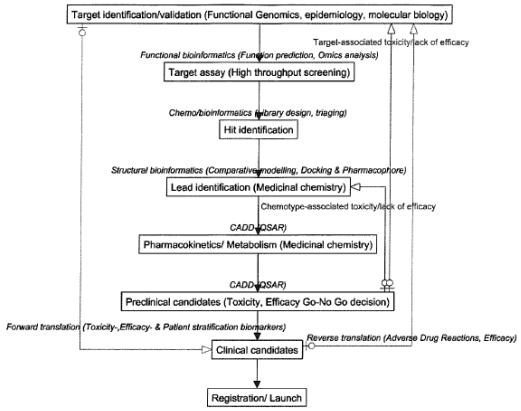


Fig. (3). A schematic flow chart summarizing the process of drug discovery, including major *in silico* contributions (italicized) to chemistry, biology† and ADMET.

Plant synthetic biology for molecular engineering of signalling and development.
Nemhauser JL, Torii KU.
Nat Plants. 2016 Mar 2;2:16010.

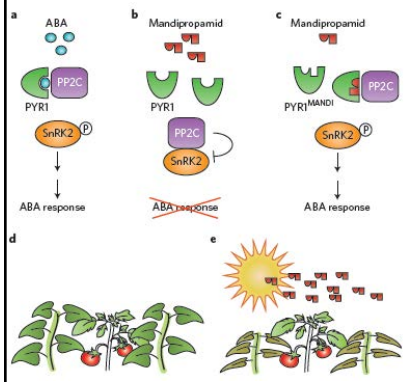
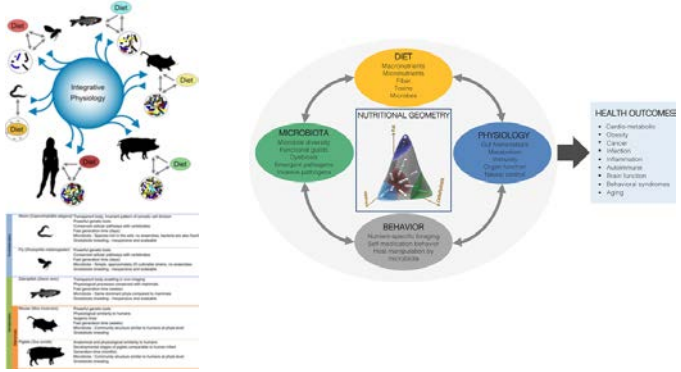
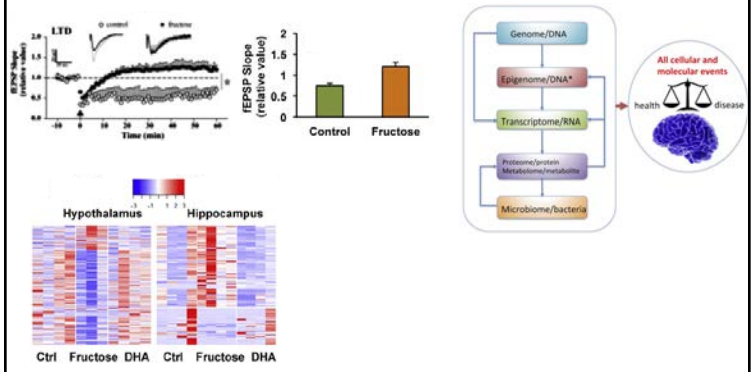


Figure 2 | An engineered ABA receptor can perceive a fungicide and trigger an ABA response. a, Endogenous ABA signalling. Binding of ABA to the PYR1 receptor promotes PYR1–PP2C association, which in turn activates the downstream protein kinase SnRK2. Active SnRK2 triggers ABA responses, including stomatal closure and drought tolerance. **b**, Mandipropamid treatment does not elicit an ABA response in wild-type plants. **c**, In a plant expressing the engineered $PYR1^{MANDI}$ receptor, mandipropamid treatment triggers ABA response and, consequently, drought tolerance. **d**,**e**, Potential and idealized translational application of a synthetic ligand–receptor system in a crop field. Here, a transgenic tomato crop plant expressing $PYR1^{MANDI}$ is outcompeted by nearby weeds (**d**). During drought season, mandipropamid application triggers ABA response to the transgenic tomato plant, thereby boosting its drought tolerance. Surrounding non-transgenic weeds do not respond to the chemical spray. Mandipropamid has already been approved by the Environmental Protection Agency as a fungicide (reg. no. 100-1281) for field application.

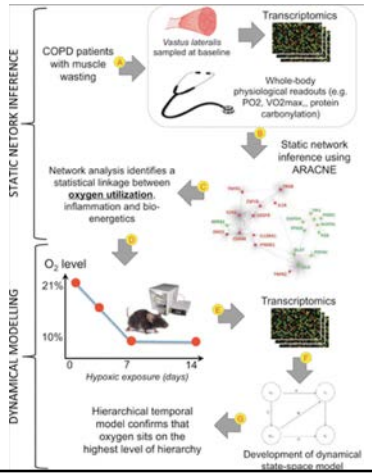
Integrative Physiology: At the Crossroads of Nutrition, Microbiota, Animal Physiology, and Human Health.
Cell Metab. 2017 Mar 7;25(3):522-534.
Leulier F, MacNeil LT, Lee WJ, Rawls JF, Cani PD, Schwarzer M, Zhao L, Simpson SJ.



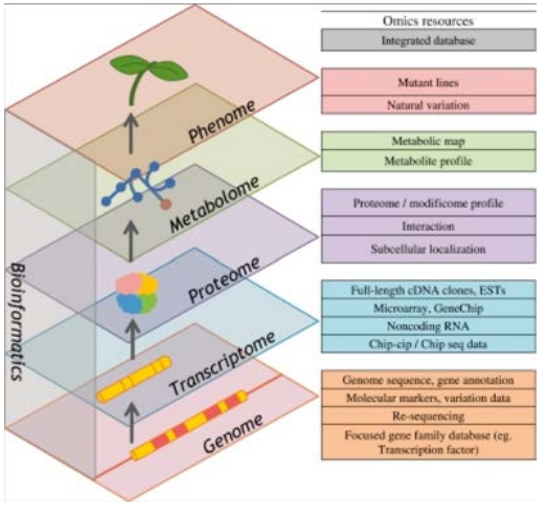
System biology approach intersecting diet and cell metabolism with pathogenesis of brain disorders.
Prog Neurobiol. 2018 Oct;169:76-90.
Gomez-Pinilla F, Yang X.

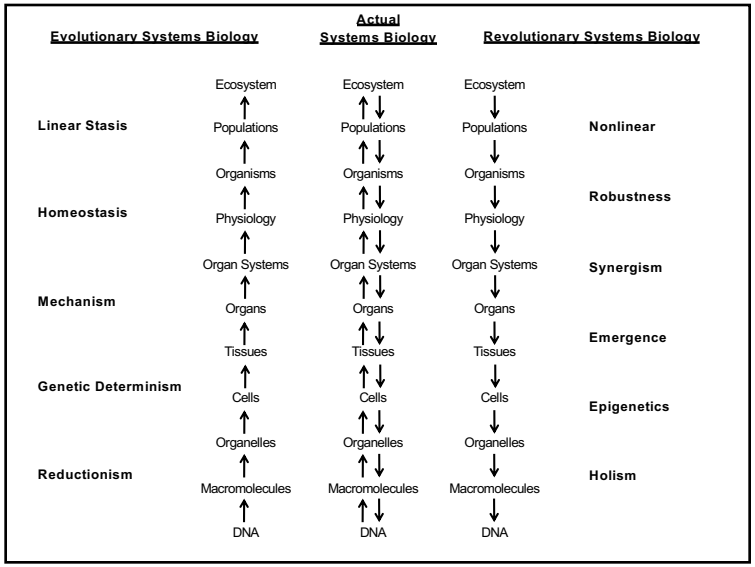


Multilevel functional genomics data integration as a tool for understanding physiology: a network biology perspective.
Davidsen PK, Turan N, Egginton S, Falciani F.
J Appl Physiol (1985). 2016 Feb 1;120(3):297-309.



Schematic representation of the analysis strategy used in the case study, highlighting how the inferred static multiscale network from the clinical chronic obstructive pulmonary disease (COPD) cohort (A–C) can be bridged to the inference of a dynamic network representing the temporal progression of events following an experimental challenge (hypoxic exposure) in a murine animal model (D–G). Having identified a clinical condition with known outcome (exercise intolerance in patients with respiratory disease), we could target unknown mechanisms by focusing on one likely source of functional limitation (skeletal muscle dysfunction ± central limitation on O2 supply) and generate data characterizing the phenotype. Both genomic and physiological readouts were used to construct a network of inferred interactions, which was then interrogated to identify statistically robust linkages among broad biological functions. While very useful in providing a list of useful biomarkers, there remains a potential limitation with single-point associations. The dynamic nature of relationships is captured by repeated measures across a suitable time scale (which will vary for different molecular, physiological, and structural responses) using an animal model of respiratory distress, where the transcriptome-based model demonstrated the central importance of oxygen in the response. $\dot{V}O_{2max}$, maximum O_2 uptake; ARACNE, Algorithm for the Reconstruction of Accurate Cellular Networks.




Evolutionary Systems Biology
Actual Systems Biology
Revolutionary Systems Biology
Linear Stasis
Homeostasis
Mechanism
Genetic Determinism
Reductionism
Nonlinear
Robustness
Synergism
Emergence
Epigenetics
Holism
Ecosystem
Populations
Organisms
Physiology
Organ Systems
Organs
Tissues
Cells
Organelles
Macromolecules
DNA