Predictive disease modeling for personalized and preventive medicine

Nicola Toschi,[1,2,3] Andrea Duggento,[1] Maria Guerrisi[1]

[1] *Department of Biomedicine and Prevention, University "Tor Vergata", Rome, Italy*
[2] *Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging*
[3] *Harvard Medical School, Boston, Massachusetts, USA*

The past two decades have witnessed enormous advances in terms of high-throughput techniques and technologies in molecular biology fields like genomics,[1] which could potentially provide the terrain for investigations targeting the polygenic and multifactorial nature of complex diseases such as neurodegenerative disorders, chronic inflammatory diseases, or cancer. The highly heterogeneous clinical states of such disorders reflect the uncharacterized interaction of numerous genes, lifestyle and environmental factors. Accordingly, highly coordinated and multicentric research efforts are also underway to collect data at the other end of the spectrum with respect to genomics, e.g. multimodal magnetic resonance imaging (MRI) and positron emission tomography (PET) data in thousands healthy subjects (Human Connectome Project)[2] or Alzheimer patients (Alzheimer's Disease Neuroimaging Initiative).[3] This is expected to increase overall power and accessibility to detect previously inaccessible disease-related biomarkers and mechanisms. In turn, this could provide better disease understanding, hence possibly leading to improved diagnosis, prognosis and prevention.[4]

However, the fundamental question of how effectively such extensive datasets can potentially be translated into forms of clinical applications, which should ideally take place at the point-of-care, remains open. In particular, the goal of personalized and preventive medicine relies fully on the ability to manage, reshape and integrate a multitude of heterogeneous data types (e.g. genetic, clinical/patient history, neuropsychological, biohumoral, molecular) which may contain effects visible only at multiple interacting temporal and spatial scales. This problem alone has been the focus of numerous bioinformatics efforts in the field of personalized medicine,[5] which have begun creating and supporting platform-independent data formats and standard in order to enhance world-wide, transdisciplinary interoperability.

While these efforts are an important and necessary stepping stone, they are not sufficient to tackle the main issue of linking *omics* molecular data to biological pathways as well as their high-level, measurable clinical and subclinical manifestations. In this context, the multifactorial study and interpretation of this rich, multifaceted patient profile lies in the realm of so called *systems biology*, i.e. an integrative modeling approach to the study of interacting biological components. Systems biology is expected to allow integration of multifaceted information into a holistic model(s), able to explain disease phenotype in a personalized fashion. More specifically, systems approaches are precisely aimed at deciphering disease complexity through integrating all possible biological information into models which should be both predictive and actionable. This is also in line with current recommendations and efforts towards developing integrative medical approaches.[6]

In order to make the crucial transition from "descriptive" (i.e. data collection and statistical analysis) to "mechanistic" (i.e. model-based interpretation of heterogenous data) thinking, one should advocate a shift from single node/single modality (i.e. only genome-based or only imaging-based) views to network-based views of human disease and its manifestations. In this context, systems biology approaches have already led to the emergence of paradigms which take advantage of a network-based interpretation of the pathogenesis of complex disorders. For example, the emerging idea of molecular networks, which can describe underlying states of a perturbed biological system underlying disease (often also termed "*biological disease maps*"), can allow the discovery/appearance of associations between entities which perform significantly better than single biological units in providing a clearer picture of the disease mechanism.[7] As an example, an application to Parkinson's disease is already openly accessible.[8]

In general, a model for a biological mechanism can either be derived from (possibly high-throughput) experimental data (i.e. a "data driven" model which makes no prior assumptions about biological mechanisms), or from so called *expert knowledge*, which injects assumptions about unmeasurable quantities/submechanisms (i.e. one builds a "knowledge driven" model). These two approaches are complementary by nature and can (and should) be combined into hybrid approaches[4] which could either explain correlations or even cause-effect relationships in the context of *biomarker* discovery (i.e. the discovery of indicators of biological or pathogenic processes as well as of response to therapeutic intervention). In this context, integrative and predictive approaches which employ network models as the basis for integration have already delivered good performance in subselecting candidate molecular biomarkers from a large combinatorial space,[9] and applications of this paradigm to breast cancer[10] and Alzheimer's disease[11] have recently appeared.

In the context of prevention, disease models can provide great aid in identifying individuals who are at risk in advance of developing symptoms tangible with traditional clinical tools. Accordingly, hybrid model-based approaches can be used to design so-called *preventive biomarkers* which aim at screening the population and stratifying it into risk classes, as has been done (for example) in cardiovascular disease.[12] Another area of application of computational models to disease prevention is the mechanistic study of putative interdependancies between disease appearance

and mechanism of risk,[4] i.e. the underlying drivers of co-morbidities. A recent example can be seen in the association between diabetes and Alzheimer's disease,[13,14] which has been confirmed in a number of large clinical and pharmacological studies.[15-18] The mechanism for this peculiar interaction between two disorders with seemingly different etiology could only be elucidated through an unified computational model which would need to aggregate extremely disparate and inhomogeneous data in order to formulate hypotheses about e.g. genomic hormone interactions underlying dementia.[11] Also, while the idea of personalized medicine stems from the field of genetics, it is now recognized that it should be interpreted as the customizations of all measures related to healthcare to individual patient needs.[19] We therefore need advanced modeling strategies and statistical tools able to stratify individuals based on their putative risk of developing a disease or possible response to therapy – an approach which distances itself from the traditional *one size fits all* therapeutic paradigm. The ability of model-based analysis approaches to design and/or discover predictive and prognostic biomarkers is central to this endeavor. Accordingly, a recent review of biomarker-discovery technologies demonstrates how integrative modeling is an emerging trend in the biomarker-discovery stream of the traditionally more innovative field of oncology,[20] highlighting a shift from correlation-based biomarkers to cause-effect biomarkers. Such information can largely be provided only through some degree of model-based analysis and interpretation.

In summary, integrative, network based disease modeling is establishing itself as a tool of growing importance in the transition from descriptive to mechanistic understanding of disease – a core goal of modern translational research. Given the enormous amount of heterogeneous data which is increasingly becoming available, along with more affordable and distributed (possibly could-based) parallel computing resources, it is expected that computational disease modeling within a systems biology approach will represent a key step in future disease management and prevention as well as drug discovery research.

References

1. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012;336(6090):64-9.

2. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, et al. The WU-Minn Human Connectome Project: an overview. Neuroimage. 2013;80:62-79.

3. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, et al. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers Dement. 2005;1(1):55-66.

4. Younesi E, Hofmann-Apitius M. From integrative disease modeling to predictive, preventive, personalized and participatory (P4) medicine. EPMA J. 2013;4(1):23.

5. Shabo A, Scarpa M. Bridging the informatics gap between bench and bedside: implications to neurodegenerative diseases. In Neurodegenerative Diseases: Integrative PPPM Approach as the Medicine of the Future. 2013:301-8.

6. Golubnitschaja O, Costigliola V. General Report & Recommendations in Predictive, Preventive and Personalised Medicine 2012: White Paper of the European Association for Predictive, Preventive and Personalised Medicine. EPMA Journal. 2012;3(1).

7. Tegnér JN, Compte A, Auffray C, An G, Cedersund G, Clermont G, et al. Computational disease modeling - Fact or fiction? BMC Systems Biology. 2009;3.

8. Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, et al. Integrating pathways of Parkinson's disease in a molecular interaction map. Mol Neurobiol. 2013:1-15.

9. Dudley JT, Butte AJ, editors. Identification of discriminating biomarkers for human disease using integrative network biology. Pacific Symposium on Biocomputing 2009, PSB 2009; 2009.

10. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Molecular Systems Biology. 2007;3.

11. Younesi E, Hofmann-Apitius M. A network model of genomic hormone interactions underlying dementia and its translational validation through serendipitous off-target effect. Journal of Translational Medicine. 2013;11(1).

12. Syed Z, Stultz CM, Scirica BM, Guttag JV. Computationally generated cardiac biomarkers for risk stratification after acute coronary syndrome. Science Translational Medicine. 2011;3(102).

13. Sims-Robinson C, Kim B, Rosko A, Feldman EL. How does diabetes accelerate Alzheimer disease pathology? Nature Reviews Neurology. 2010;6(10):551-9.

14. Ott A, Stolk RP, Van Harskamp F, Pols HAP, Hofman A, Breteler MMB. Diabetes mellitus and the risk of dementia: The Rotterdam Study. Neurology. 1999;53(9):1937-42.

15. Yaffe K, Falvey C, Hamilton N, Schwartz AV, Simonsick EM, Satterfield S, et al. Diabetes, glucose control, and 9-year cognitive decline among older adults without dementia. Archives of Neurology. 2012;69(9):1170-5.

16. Bomfim TR, Forny-Germano L, Sathler LB, Brito-Moreira J, Houzel JC, Decker H, et al. An anti-diabetes agent protects the mouse brain from defective insulin signaling caused by Alzheimer's disease-associated Aβ oligomers. Journal of Clinical Investigation. 2012;122(4):1339-53.

17. McClean PL, Parthsarathy V, Faivre E, Holscher C. The diabetes drug liraglutide prevents degenerative processes in a mouse model of Alzheimer's disease. Journal of Neuroscience. 2011;31(17):6587-94.

18. Moon JH, Kim HJ, Yang AH, Kim HM, Lee BW, Kang ES, et al. The effect of rosiglitazone on LRP1 expression and amyloid β uptake in human brain microvascular endothelial cells: A possible role of a low-dose thiazolidinedione for dementia treatment. International Journal of Neuropsychopharmacology. 2012;15(1):135-42.

19. Simmons LA, Dinan MA, Robinson TJ, Snyderman R. Personalized medicine is more than genomic medicine: Confusion over terminology impedes progress towards personalized healthcare. Pers Med. 2012;9(1):85-91.

20. Deyati A, Younesi E, Hofmann-Apitius M, Novac N. Challenges and opportunities for oncology biomarker discovery. Drug Discovery Today. 2013;18(13-14):614-24.