Published in Journal of internal medicine, 2013, vol. 274, no. 4, pp. 321-328 which should be cited to refer to this work

How to improve data and knowledge management to better integrate healthcare and research

Montserrat Cases[1,*], Laura I. Furlong[1,*], Joan Albanell[2], Russ B. Altman[3], Riccardo Bellazzi[4], Scott Boyer[5], Angela Brand[6], Anthony J. Brookes[7], Søren Brunak[8], Timothy W. Clark[9], Joaquim Gea[10], Peter Ghazal[11], Norbert Graf[12], Roderic Guigó[13,14], Teri E. Klein[15], Núria López-Bigas[1,16], Víctor Maojo[17], Barend Mons[18], Mark Musen[19], José L. Oliveira[20], Anthony Rowe[21], Patrick Ruch[22], Amnon Shabo (Shvo)[23], Edward H. Shortliffe[24], Alfonso Valencia[14,25], Johan van der Lei[26], Miguel A. Mayer[1], Ferran Sanz[1,14,**]

[1] Research Programme on Biomedical Informatics (GRIB), IMIM. DCEXS, Universitat Pompeu Fabra, Barcelona, Spain.
[2] Servei d'Oncologia, Hospital del Mar – IMIM. DCEXS, Universitat Pompeu Fabra. Barcelona, Spain.
[3] Departments of Bioengineering, Genetics, and Medicine, Stanford University, CA, USA.
[4] Dipartimento di Ingegneria Industriale e dell'Informazione, Università di Pavia, Italy.
[5] Safety Assessment, AstraZeneca, Mölndal, Sweden.
[6] Institute for Public Health Genomics, Maastricht University, Netherlands.
[7] Department of Genetics, University of Leicester, UK.
[8] Center for Protein Research, University of Copenhagen. Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark.
[9] MassGeneral Institute for Neurodegenerative Disease, Massachusetts General Hospital, Cambridge, Massachusetts, USA.
[10] Servei de Pneumologia, Hospital del Mar – IMIM. DCEXS, Universitat Pompeu Fabra. CIBERES. Barcelona, Spain.
[11] Division of Pathway Medicine, Synthesys, University of Edinburgh, Scotland, UK.
[12] Department of Pediatric Oncology and Hematology, Saarland University, Germany.
[13] Centre de Regulació Genòmica (CRG), Barcelona, Spain.
[14] Spanish Institute of Bioinformatics (INB), Spain.
[15] Department of Genetics, Stanford University, CA, USA.
[16] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
[17] Biomedical Informatics Group, Universidad Politécnica de Madrid, Spain.
[18] Department of Human Genetics, Leiden University Medical Center. Netherlands Bioinformatics Center, Nijmegen, Netherlands.
[19] Stanford Center for Biomedical Informatics Research, Stanford University, CA, USA.
[20] DETI/IEETA, University of Aveiro, Aveiro, Portugal.
[21] Janssen Research & Development, UK.
[22] BITEM, University of Applied Sciences, Geneva, Switzerland.
[23] IBM Research Lab in Haifa, Israel.
[24] Arizona State University, Phoenix, AZ, USA. Department of Biomedical Informatics, Columbia University, NY, USA.
[25] Spanish National Cancer Research Centre (CNIO), Madrid, Spain.
[26] Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands.

* Equally contributed to this manuscript
** Correspondence to F. Sanz, e-mail: ferran.sanz@upf.edu

ABSTRACT:

Medicine is an increasingly data intensive discipline, with a growing need to link individual patient health records to rapidly changing research knowledge for better differential diagnosis, prognosis, and prediction of treatment response. Equally, biomedical research will gain massively from the integrative analysis of clinical and multi-omics information. Capitalizing on these opportunities must be guided by a precise understanding of many complex issues related to the integration of large amounts of diverse information. Partly this involves overcoming barriers between different disciplines, such as biology, medicine and computer sciences – implying a key role for 'knowledge engineers'. This paper summarizes recent expert debates on such matters, leading to suggestions for concrete actions that should improve and better synergize both research and healthcare.

KEYWORDS:

Biomedical informatics; knowledge management; knowledge engineering; translational research; translational bioinformatics; electronic health records; decision support systems; personalized medicine; P4 medicine; genomic medicine

ABBREVIATIONS & ACRONYMS:

CDSS: Clinical Decision Support System
CPOE: Computerized Physician Order Entry
EHR: Electronic Health Record (European synonym for the US Electronic Medical Record, EMR)
GWAS: Genome-Wide Association Studies
IT: Information Technology
KE: Knowledge Engineering
RRS: Reproducible Research Systems

"Once upon a time, some engineers, biologists and clinicians realized that a lot of information in biomedicine was partitioned into silos that do not intercommunicate. These silos were a side effect of the existence of different disciplines required to, for example, develop new drugs.

The engineers decided to make the silos go away, and to put the information in axiomatic form to facilitate automatic reasoning over multiple data sources. They also decided to do this in a very open way so that effort was not duplicated. This seemed like a very reasonable step and was welcomed by all.

After they had done a lot of axiomatization, the engineers found that there were still issues. They found lack of agreement on many seemingly uncomplicated 'facts'. They had to employ curators to resolve the 'facts', and then people said the curators were losing the plot. They also found out that there were not only discipline silos, but also intra-discipline silos. These were the partitions between evidence and the assertions developed from the evidence, and the earlier assertions these cited, based on even earlier assertions, and so on. There were not only webs of disagreement, but also *chains of error. And they found that connecting facts from various silos was not so uncomplicated after all, even after axiomatization. Why was this? Because the results of scientific experiments are not axioms, even if they may be treated in this way to perform isolated bits of reasoning." (T.W. Clark)*

This story illustrates the challenge that scientists and clinical practitioners face: the world contains a vast array of complex and diverse data, but locating and connecting the information is difficult[1-3], and deriving definitive knowledge from the data to guide research and/or for clinical practice is even harder. The many road blocks that make it difficult to progress this field were recently discussed at a scientific debate held in Barcelona on July 3-4, 2012 (http://www.bdebate.org/debat/beyond-omics-revolutions-integrative-knowledge-management) under the general title "Beyond omics revolutions: Integrative Knowledge Management for Empowered Healthcare and Research". The meeting was organized around six topics: "Dealing with biomedical knowledge explosion for better healthcare: Identifying actionable knowledge items at the point of care", "Exploiting patient information to enrich basic biomedical research", "Standards for clinical-omics integration: the semantic challenge", "New IT is supporting massive biomedical data management", "Systems medicine: Making systems biology translational", and "Integrative knowledge management for improving drug R&D". The main ideas and conclusions arising from this event are presented below.

Translating research findings into actionable knowledge in the clinical setting

New biomedical discoveries emerge at an ever-increasing rate, but their translation into healthcare typically occurs slowly or not at all. There is a lack of sufficient systems that can astutely identify, distil and hand on these advances to the relevant practitioners, in useable formats. For example, thousands of biomarkers exist, comprising a few truly useful ones intermingled with many other less or non-actionable items. Valuable new biomarkers (diagnostic, prognostic or therapeutic) are therefore not effectively being taken forward into healthcare. The gamble of knowing which ones to progress with is simply too onerous - given the cost of modern clinical trials and a deficiency of incentives and expertise amongst researchers who would be best placed to progress markers down the development path. Hence, when this translation does happen it is usually because of a major 'pull' from the clinical world, rather than 'push' from researchers.

Clearly then, there is a need for methods and systems that can reliably and routinely identify and connect the most informative, reliable, and useful information (not least biomarkers) generated by the research community. Efforts to better structure scientific knowledge, for instance by means of nanopublications[4] or the ISA commons[5], could provide key parts of this solution. But the challenge is magnified by the fact that the relevant information is spread not only across research resources (e.g., literature, patents, laboratory reports, market data, medical reports, biobanks, etc), but also in

realms with less professional rigor such as social networks and patient communities (e.g., wikis, blogs and other social media platforms). Progress will therefore necessitate addressing cross-language and cross-jargon barriers, as well as all the traditional targets of interoperability such as standards for data syntax and semantics.

Beyond connecting and integrating research findings, there lies the challenge of understanding this information. Education is important here, and indeed it has been proposed that a lack of appropriate training explains the slow uptake of companion diagnostics into clinical practice[6]. Tackling this will require robust guidelines on how to use pharmacogenomics information, and also the provision accompanying pharmacokinetics, metabolic and drug interaction knowledge derived from the latest biomedical research. Arguably then, researchers have a responsibility to make their clinically relevant research findings more understandable to the healthcare sector, perhaps in the form of user-friendly web portals or other software[7,8]. Electronic Health Record (EHR) developers, Computerized Physician Order Entry (CPOE) designers, and Clinical Decision Support Systems (CDSS) creators and vendors likewise need to be involved in bringing forth additional content for such portals, and in connecting such platforms to the intended end users.

Putting all the above issues together, and thinking also about the core data interpretation challenges, several experts concluded that the overall challenge is one of "knowledge engineering", rather than simply a need for better informatics, research, or medical practice. Hence, it may be difficult to make real progress with biomedical researchers and clinical practitioners alone, and so there is a need for a new breed of multidisciplinary engineers[9]. This echoes back to the tradition of Knowledge Engineering (KE) for Health, a field that stemmed from Artificial Intelligence research in the 1990s[10]. However, contrasting to previous KE approaches that aimed at organizing all the data to reveal absolute knowledge (which is a flawed concept, as our lead story illustrated), there is a need for a far more pragmatic approach ('KE 2.0') – aimed at identifying and making directly useful the very limited set of data and knowledge items that are both reliably proven and clinically actionable. The aim would be to explicitly address the two core information problems faced by clinicians: (i) having too much existing and new data to keep up with, and (ii) not having time or resources to discern reliable from uncertain and erroneous information.

As listed in Supplementary Table 1, many international projects now exist that aim to integrate various types of data related to specific diseases or their pharmacological treatments. In general, however, these are not doing KE 2.0 but developing new methodologies and tools for data integration and exploitation, or novel strategies for massive data storage and handling. But as these sorts of projects make progress in consolidating and unifying the relevant data, KE 2.0 approaches can then begin explored. However, for this to succeed, the data must be of suitable quality and breadth, as discussed in the following section.

Data quantity and quality

Sadly, in many situations today, petabytes of potentially useful biomedical data are not captured in a structured format and/or made available for use by others. This includes molecular *omics* profiles (genomes, transcriptomes, proteomes, epigenomes, etc.), exposure to environmental chemicals (exposomes), phenotype data (e.g., as recorded in clinical settings), and dynamic data (e.g., measurements at different time or space points) – all of which could contribute to improved research and healthcare. For instance, in the research world, primary data from high-throughput studies on a large number of subjects (e.g., genome-wide association studies, GWAS)[11] typically never escape from the lab where they were generated, and in the healthcare world, molecular profiles of individual patients, sometimes recorded per time period, are starting to be recorded by then poorly exploited[12]. Of course, simply handling this diversity and scale of data is a challenge in itself, but that should motivate focusing much effort upon it, rather than providing a reason for letting the data be lost.

Many considerations pertain to the quality, completeness, reliability and reproducibility of primary data and the knowledge derived from them. Relevant judgments may well be context dependent – such as whether a biopsy from a heterogeneous tumor might be considered usefully representative of the whole tumor. Contextual metadata (data about the data) are therefore important, but such information is often not properly collected or recorded. This is directly related to current discussions about the reproducibility of research findings and the comparability of different analytical procedures. Approaches that allow consistent and repeated analysis on datasets are therefore becoming important (e.g., Galaxy, GenePatterns). Question about reproducibility concern both the data (how it was produced) and the knowledge gleaned from data (how it was derived). In this respect, we refer the reader to important studies about statistical and experimental design problems in contemporary scientific publications[13,14].

One notable problem in applying KE to biomedical research data is the nature of the knowledge being engineered. Specifically, active as opposed to consolidated scientific knowledge, consists of assertions supported by evidence. What we consider knowledge is a snapshot of the consensus of the scientific community on a particular subject at a given time, but this active knowledge is subjected to continuous re-evaluation where new findings change our perspective, and 'facts' may be refuted after some years. Essentially, no knowledge is truly absolute. A particular complication here is that of human bias or error underlying citation distortion, not least in review articles. An example can be seen in a recent review suggesting a role for inclusion body myositis in the etiopathology of Alzheimer disease. Following the chain of assertions to the grounding evidence, it was found that in some cases there was no such grounding evidence, and in other cases its meaning had been distorted or the results misapplied or misconstrued[15]. These issues contribute to the existence of intra-discipline silos, which disconnect facts and assertions from the underlying evidence. In other cases, we face discrepancies between data collected from different sources. This clearly argues the need for more information accessibility and structure, and less reliance on subjective human opinion. But this itself must be tempered against the risk of drawing too many hypotheses from extensive and high-throughput data, which could easily lead to spurious associations.

In this context, ongoing multi-party curation efforts from different initiatives are appreciated as a way to identify and organize relevant information, but they represent very costly and time-consuming tasks. Efforts on harmonization and standardization, as well as the development of software for supporting curation tasks, are therefore needed to improve and assist curators in their work.

An important point to emphasize is that we need very different levels of evidence for CDSS compared to what is required for research grade knowledge discovery. Medical reasoning may be represented by epistemological models, which are amenable to partial automation[16,17], and in all cases the data should be generated or chosen to fit a purpose. Researchers, for example, must design their experiments and simulations to record as much detailed information as possible, to facilitate a comprehensive exploration of the biomedical question. In contrast, clinicians must carefully define healthcare questionnaires and register just the salient medical variables pertaining to their patients to aid in clinical decision-making. Ideally, however, to avoid continuing with silos of data, both groups should always also consider the possible or likely reuse of their data. As part of this, data provenance should be carefully recorded to make possible the retrieval of the original sources and to ensure its reliability and reproducibility, which will undoubtedly have an effect on the generation of useful predictions[18].

Time constraints at the Barcelona meeting precluded extending this discussion into areas of ethical and legal frameworks, and so further information on the matter can be found elsewhere[19,20]. It should also be noted that the European Parliament is currently discussing a data protection directive that will underpin a new legal framework (http://ec.europa.eu/justice/data-protection/index_en.htm).

Standards to facilitate translation

Increasingly, genomic information is likely to be relevant to healthcare, and as such it should ideally be stored within medical records. A current use case would be that of personalized drug dosing. Some pharmacogenomic tests are now being used in routine clinical practice, however they are vastly under-used. Key biological data on individuals should be encapsulated in its native format in clinical data structures, with 'bubbled-up' items being associated with phenotypic data using clinical data standards. This then spawns the question as to what standards are required to allow the efficient translation of key research findings into clinical practice, and what IT paradigms will be needed to support biomedical data management. Of course part of the answer concerns controlled vocabularies and ontologies for the integration of diverse and heterogeneous biomedical information. Fortunately, several initiatives today support the development of ontologies to describe different aspects of biology and biomedicine (e.g., NCBO (http://www.bioontology.org/), OBO (http://www.obofoundry.org/) Ricordo (http://www.ricordo.eu/)). But yet more needs to be done. For instance, it is difficult to reconcile medical records with disease descriptions associated with public molecular data. This is due to the inherent complexity of diseases and the way they have been traditionally classified and described. Also, disease descriptions are very heterogeneous and often dynamic, as in the case of mental illness[21].

Beyond ‘standards’ perhaps there is actually an equal need for ‘understandards'. In other words, efforts that aim to deliver the standardization capabilities required for KE 2.0, not just standards for semantic integration irrespective of common understanding. We need to make sure that the next generation of *in cerebro* and *in silico* reasoning strategies understands what is 'meant' by any node and edge in a network of associations. To resolve the tension that the more expressive a standard is the less interoperable it is, constraining the standards is crucial, which also enables capturing similarities while preserving disparities. More specifically, health data semantics and context cannot be faithfully represented using flat structures (e.g., a list of entries), rather it requires a compositional language that meaningfully connects various data entries.

Furthermore, health data standards need to accommodate unstructured data and text (e.g., clinician's narrative), while having links to structured data entries. A life-time comprehensive recording of personal health information including *omics* data is certainly desirable. This arguably calls for a new model of data stewardship: the Independent Health Record Banks vision (http://independenthealthrecordbanks.blogspot.co.il/), which would support the implementation of lifelong, cross-institutional and interoperable EHR. This would constitute an escape from 'legacy systems' fixation. As long as healthcare providers are also record-keepers, we will continue to have poor archives, proprietary-based, isolated in silos, with most of the data semantics not represented explicitly – making it hard or impossible for CDSSs to be really effective. So instead it is proposed that there should be a limited number of independent and regulated third parties specialized in sustaining the individual life-time EHR, continuously curating it and running various analyses to prepare the right info-structure for CDSS. These tasks require unique specialization, a new kind of archive, which should provide the most complete and coherent information framework to support the individual health.

Fostering literacy in health information management

The challenge of improving biomedical knowledge management goes hand in hand with the need for suitable education and training for all the relevant stakeholders: patients, clinicians, researchers, and regulators and policy makers. In particular, clinicians need more support to improve their ability to interpret and use research findings, and researchers must learn how to take actionable findings closer to the clinicians. Concomitantly, researchers need to better comprehend the problems raised in clinical practice that can be solved in the laboratory or by intensive use of

information technologies (IT). This reinforces the need for forums of interaction with the active participation of biomedical researchers, bioinformaticians and physicians with experience in clinical research. Hence, we should move from a one-size-fits-all to a stratified medicine education, and from this towards a truly individualized clinical exercise, following the paradigm shift towards the predictive, preventive, personalized and participatory medicine (P4) concept[22]. Finally, the active participation of citizens, via blogs and other social networks, provides a way to improve the general level of health literacy, and thereby to empower all individuals regarding their role in the health care system.

OUTCOMES OF THE DEBATES

The experts that took part in the aforementioned debates held in Barcelona, also offer the following consensus statements:

- There is an urgent need to promote communication and collaboration between experts from different disciplines in order to overcome current information silos and to setup integrated knowledge frameworks required for better managing health problems. In this regard, patients' voices have to be considered as well.

- The current rate of growth of data exceeds that of computational power (e.g., throughput of sequencing instruments will grow faster than the capacity of computers, and this can become a limitation for the spread of Next Generation Sequencing data use in medical practice). It should be considered malpractice to fund data generation without an adequate data exploitation and stewardship plan. Research funding must seriously consider the need for data storage and analysis, which may well comparable to the effort needed for data generation. When data is generated on human subjects, the stewardship of those data might be handled within each subject's EHR, if a cross-institutional and lifelong record is available.

- Efforts should be made to improve the methodological and technological background to allows the integrative analysis of complex information (KE 2.0), with the aim of distilling and delivering clinically actionable information and supporting computational predictions to facilitate the prevention and treatment of diseases.

- Maximizing data sharing should be an imperative. Not all healthcare data need to be protected under a controlled access regime and not all research need to be open access. Most current barriers for data sharing and reuse are not technical but social. In this respect, we acknowledge novel initiatives (e.g., *altmetrics*, http://altmetrics.org/manifesto/) that seek to go beyond the classical narrative articles as the only source of scientific knowledge to be taken into account.

- It is important to address language and jargon barriers to connect the worlds of traditional scientific reporting (peer-reviewed articles) and web sources (patients blogs, twitter) as sources for knowledge discovery.

- To facilitate the effective reuse of information, elements of provenance and context along with the basic assertions have to be captured from text, databases and EHR systems.

- The current classification of diseases is largely based on signs and symptoms, and in general does not take into account current and evolving knowledge on the molecular pathways that lead to any particular illness. A diseases classification based on the molecular biology or the genomics of the diseases would help in the identification of relevant therapeutic interventions.

- Proper guidelines are needed to help clinicians understand how the results of available genetic tests should be used to optimize patient care, rather than whether tests should be ordered. Here researchers have a role in preparing these guidelines. Disease and/or domain specific

'Knowledge portals' could provide a key part of the overall solution, facilitating and driving analysis of data, regulating and tracking data access, and providing an optimum balance and scale in terms of the centralization-federation challenge.

- Citizens (including health professionals) must be enabled, individually and cooperatively, to access, understand, appraise, and apply information that will facilitate the use of genome-based information for the benefit of individuals and their communities. In addition, we have also to consult with citizens and patients on "donating their data".

- All clinical and research data related to an individual's health should be stored in, or linked to, a single lifelong personal (electronic) health record, which would overcome current institutional borders. The development of independent Health Records Banks may be a way of implementing this vision (e.g., http://independenthealthrecordbanks.blogspot.co.il/).

ACKNOWLEDGEMENTS

The present paper is based on the debates held in Barcelona on July 3-4, 2012 with the active participation of the authors. The debates were organized by B-Debate (an initiative of Biocat and Obra Social 'La Caixa') and Universitat Pompeu Fabra (Barcelona). The event was held in the framework of the European INBIOMEDvision project (funded by the EU FP7 under grant agreement no. 270107). In addition, we also received support from EU FP7 project no. 200754 (GEN2PHEN) and the IMI JU under grant agreements no. 115002 (eTOX), and no. 115191 (Open PHACTS), resources of which are composed of financial contribution from the EU FP7 and EFPIA companies' in kind contribution. L.I. Furlong received support from ISCIII FEDER [CP10/00524].

Supplementary Information

Table 1. Recent initiatives related to data and knowledge management in healthcare and biomedical research.

REFERENCES

1. Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, et al. (2007) Advancing translational research with the Semantic Web. BMC Bioinformatics 8 Suppl 3: S2.

2. Antezana E, Kuiper M, Mironov V (2009) Biological knowledge management: the emerging role of the Semantic Web technologies. Brief Bioinform 10: 392-407.

3. Butte AJ (2008) Translational Bioinformatics: Coming of Age. J Am Med Inform Assoc 15: 709-714.

4. Mons B, van Haagen H, Chichester C, Hoen PB, den Dunnen JT, et al. (2011) The value of data. Nat Genet 43: 281-283.

5. Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, et al. (2012) Toward interoperable bioscience data. Nat Genet 44: 121-126.

6. Wilffert B, Swen J, Mulder H, Touw D, Maitland-Van der Zee AH, Deneer V (2011) From evidence based medicine to mechanism based medicine. Reviewing the role of pharmacogenetics. Int J Clin Pharm 33: 3-9.

7. Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan R (2012) INDI: a computational framework for inferring drug interactions and their associated recommendations. Mol Syst Biol 8: 592.

8. Preissner S, et al. (2010) SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. Nucleic Acids Res 38: D237-D243.

9. Beck T, Gollapudi S, Brunak S, Graf N, Lemke HU, et al. (2012) Knowledge engineering for health: A new discipline required to bridge the “ICT gap” between research and healthcare. Hum Mutat 33: 797-802.

10. Warner HR, Sorenson DK, Bouhaddou O (1997) Knowledge engineering in health informatics. New York: Springer-Verlag.

11. Hardy J, Singleton A (2009) Genomewide association studies and human disease. N Engl J Med 360: 1759-1768.

12. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, et al. (2012) Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes. Cell 148: 1293-1307.

13. Fernandes-Taylor S, Hyun JK, Reeder RN, Harris AHS (2011) Common statistical and research design problems in manuscripts submitted to high-impact medical journals. BMC Res Notes 4: 304.

14. Prinz F, Schlange T, Asadullah K (2011) Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov 10: 712.

15. Greenberg SA (2009) How citation distortions create unfounded authority: analysis of a citation network. BMJ 339: b2680.

16. Sparkes A, Aubrey W, Byrne E, Clare A, Khan MN, et al. (2010) Towards Robot Scientists for autonomous scientific discovery. Automated Experimentation 2: 1.

17. Riva A, Nuzzo A, Stefanelli M, Bellazzi R (2010) An automated reasoning framework for translational research. J Biomed Inform 43: 419-427.

18. Ekins S, Waller CL, Bradley MP, Clark AM, Williams AJ (2012) Four disruptive strategies for removing drug discovery bottlenecks. Drug Discov Today [epub ahead of print].

19 Hudson KL (2011) Genomics, Health Care, and Society. N Engl J Med 365: 1033-1041.

20. Jensen PB, Jensen LJ, Brunak S (2012) Mining electronic health records: towards better research applications and clinical care. Nat Genet 13: 395-405.

21. Tabarés-Seisdedos R, Dumont N, Baudot A, Valderas JM, Climent J, et al. (2011) No paradox, no progress: inverse cancer comorbidity in people with other complex diseases. Lancet Oncol 12: 604-608.

22. Hood L, Friend SH (2011) Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol 8: 184-187.