“Omics” research, monetization of intellectual property and fragmentation of knowledge: can clinical epidemiology strengthen integrative research?

Miquel Porta\textsuperscript{a,b,c,*}, Ildefonso Hernández-Aguado\textsuperscript{c,d}, Blanca Lumbreras\textsuperscript{c,d}, Marta Crous-Bou\textsuperscript{a,b,c}

\textsuperscript{a}Institut Municipal \`d Investigació Mèdica, Universitat Autònoma de Barcelona, Carrer del Dr. Aiguader 88, E-08003, Barcelona, Spain
\textsuperscript{b}Facultat de Medicina, Universitat Autònoma de Barcelona, Barcelona, Spain
\textsuperscript{c}CIBER en Epidemiología y Salud Pública (CIBERESP), Spain
\textsuperscript{d}Department of Public Health, Universidad Miguel Hernández, Alacant, Spain

Accepted 5 June 2007

Abstract

An analysis of the contributions of “omics technologies” to human health and clinical care needs to address the relationships between internal issues (e.g., methodological shortcomings in “omics” research and clinical biology) and external influences. Among the latter, monetization of intellectual property (IP) appears to be a powerful force favoring methodological limitations and an excessive reductionism and fragmentation of biological knowledge. Following economic successes in other industries (semiconductors, software, and “dot-coms”), monetization of IP tries to market small fragments of big research “puzzles”; the strategy seems partly responsible for the biotech industry having underperformed methodological, clinical, and economic expectations. Hence, internal, purely scientific reasons can hardly explain failures in the application of long-proven principles of clinical epidemiology to the discovery and validation of diagnostic and prognostic tests. Nevertheless, this paper also sketches methodological proposals that may help integrate microbiological, clinical, and environmental evidence. Clinical and epidemiological reasoning, knowledge, and methods need to be applied on a much wider scale than until now by “omics” studies that aim at making inferences relevant for human beings. Rather than adopting the values and norms of “science business,” “omics research” could apply a diversity of clinicoepidemiological models favoring integrative research.

Keywords: Proteomics; Genomics; Clinical epidemiology; Methods; Relevance; Validity

1. Introduction

A broad analysis of the promises, successes, and failures of “omics” technologies is long overdue in clinical epidemiology. Clearly, epidemiological reasoning, knowledge, and methods need to be applied on a wider scale than until now to critically assess “omics” studies, as well as by “omics” studies that aim at making inferences relevant for human beings. This is the bottom line, in our view: aiming at inferences relevant for human beings. The integration of epidemiology, clinical medicine, and microbiology could hence contribute to improve the internal validity and human significance of “omics research” [2–7].

Scientific processes required to achieve truly integrative research [2] resemble in several ways the interactions that occurred between clinical medicine and epidemiology at a particularly creative pace in the early 1980s: a new partnership that effectively integrated research methods, professional skills, and scientific knowledge [3,6–10]. Much as epidemiology and clinical medicine shaped each other during the past 20 years, epidemiology and biology now need to cross-fertilize each other.

Genomic- and peptidomic-based technologies promised and were expected to make groundbreaking discoveries of markers with high diagnostic and prognostic value in clinical medicine. Public health and environmental applications
are also expected. Although some success has been made in the diagnosis of some genetic diseases, few of the many tests proposed have achieved wide clinical application [11]. A reflection made from the perspective of “business science” [12] is particularly instructive (Box 1). As many others, we were often somewhat puzzled and worried by so much reductionism and fragmentation of knowledge in the microbiological sciences [13]. Pisano’s reasoning (Box 1) [12] is thus noteworthy because it suggests specific reasons—not strictly methodological—for that fragmentation. Specifically, he indicates that a “two-step” external process is causing the fragmentation: first, “monetization of intellectual property” (IP) has been a powerful shaping force in biotech; and second, “the idea behind monetization of IP is that you don’t need to actually develop a product [e.g., a clinically useful diagnostic test or drug], you can just develop a piece of IP, and then capture financial returns through licensing or other market arrangements.” According to the “science business” expert [12], this has worked in fields where breaking up a “big puzzle” into its relatively independent pieces yields particular pieces that can be valued independently. Yet, this is seldom the case when you aim at developing products useful for sick human beings.

In a series of thoughtful and stimulating papers, David Ransohoff has outlined difficult challenges to be faced when developing and validating genome- or peptidome-based diagnostic tests [14–18]. It is worth emphasizing that some of the challenges have direct commercial and social implications [19–21]. “Evidence-based diagnosis” was already difficult to practice before the genomic era [7,22,23], with many reports showing widespread biases and other failures in academic papers on diagnostic research [24–26]. If compliance with the STARD standards for the reporting of diagnostic accuracy research [27] was low for traditional diagnostic tests, it is now even lower in genome-based tests [28]. Thus, we are currently facing not only the problems of any diagnostic research, but these are also compounded by additional difficulties arising from genome-based technologies—and additionally compounded too by the new pressures that the stock market exerts on promising areas of the “new economy” (Box 2).

2. Roles of research phases for diagnostic and prognostic tests

Ransohoff puts forward some solutions to improve “-omics research.” He suggests that guidelines, recommendations, and use of “phases” in the development of a diagnostic marker [29] may improve study design and address important concerns on the performance of a marker. We largely agree with him; for instance, we share his reluctance toward taking “shortcuts” (e.g., using biological samples collected by previous studies with unknown sampling frames or selection criteria). We would also like to stress

Box 1 Fragmentation of knowledge and monetization of intellectual property (IP)¹

- After 30 years, the numbers are in on the biotech business—and it is not what we expected. Though some firms such as Amgen has created dramatic breakthroughs, the overall industry track record is poor—in aggregate, the sector has lost money.

- The biotech industry has underperformed expectations. Biotech has not lived up to its expectations, either in providing outstanding returns for investors or improving research and development productivity.

- Genomics, stem cells, systems biology, and proteomics—all were new streams of science that attracted significant commercial investment long before the science was fully worked out.

- We expected biotech to “work” just like all other high-technology industries, and thus we deployed a lot of the same thinking, models, financial arrangements, and strategies that worked elsewhere, but just did not fit here.

- The sector has indiscriminately borrowed business models, organizational strategies, and approaches from other high-technology industries, and thus we deployed a lot of the same thinking, models, financial arrangements, and strategies that worked elsewhere, but just did not fit here.

- The industry needs to place greater emphasis on long-term learning over short-term monetization of IP. Pisano argues that “monetization of IP” has been a powerful shaping force in biotech. The idea behind monetization of IP is that you do not need to actually develop a product; you can just develop a piece of IP, and then capture financial returns through licensing or other market arrangements. This has worked wonderfully in semiconductors and software; but monetization of IP only works there because of some very specific conditions. You need to have a very modular knowledge base; that is, you need to be able to break up a “big puzzle” into its relatively independent pieces so that a particular piece can be valued independently. So, we have been breaking up the pieces of the puzzle into independent pieces when what matters is the way we integrate the pieces. Integration matters a lot.

Box 2 Basic science trades on the stock market

- In the late 1990s, Celera Genomics, the private company funded by Craig Venter, announced that it would seek patent and “intellectual property (IP) protection” on over 6,000 whole or partial genes. In March 2000, president Clinton announced that the genome sequence could not be patented, and should be made freely available to all researchers. The statement sent Celera’s stock plummeting and dragged down Nasdaq. The biotech sector lost about $50 billion in market capitalization in 2 days.

- Wednesday, June 19, 2000. It is only 6 days before president Clinton and Prime Minister Blair will scrutinize the presentation in society of the first “rough drafts” of the human genome nucleotide sequence. In a section of the newspaper that is still called “New Economy” (Business/Financial desk, Section C, page 4), The New York Times headline is: There’s gold in human DNA, and he who maps it first stands to win on the scientific, software and business fronts.

The business correspondent Tim Race “sees the project to map human genome, which nears completion, as the next big thing in New Economics. The way that New Genomics is being presented to investment world seems modeled on how Internet was introduced.”

- The week before, on June 12, 2000 the title of the “cover story” of Business Week is: The Genome Gold Rush. Who will be the first to hit pay dirt? It is just a reflection of the amazing economic and social expectations that “basic science”/“business science” is at the time stimulating.

- These were indeed times of unprecedented growth of the “dot-com bubble.” Take Nasdaq, the world’s first and largest electronic stock market, for instance. On March 10, 2000, the technology-heavy Nasdaq Composite index reached an all-time historical of 5,048.62. Three years later, the index was 1,300. At the writing of this article (May 2007), it is around 2,500.

- According to Arthur D. Levinson, Chief Executive Officer of Genentech, “Since 1976, when our company was founded, the biotech industry has lost $90 billion in aggregate. I think it’s the biggest money-losing industry of all time. It is hemorrhaging. There are some exceptions: We are doing well, and Amgen is doing well, But for most of the 1,300 to 1,400 companies -300 or 400 of them public- this is a money-losing enterprise.”

---

2 Chase M. How Genentech wins at blockbuster drugs. CEO to critics of prices: “Give me a break”. The Wall Street Journal; 5 June 2007. page B1

that a fine balance is necessary between high reliability and efficiency. On a minor and formal point, we feel “marker research” is an awfully ambiguous expression.....

We concur with David that guidelines and recommendations mainly guarantee the appropriate application of a marker to a population group. However, we do think that the use of guidelines is essential because through editors’ support, research appears more transparent, leading to a complete and informative reporting that can improve decisions in health care. The inclusion of recommendations in the authors’ guidelines has been shown to modify reporting patterns, making published research clearer [30]. No doubt, use of guidelines is no substitute for a thoughtful reflection and insight of the investigator. However, when thinking of a study of biomarker validation, the first question is the research objective. In this sense, the use of research phases, as proposed originally by Feinstein for any diagnostic test [10], and developed by Pepe et al. for cancer biomarkers [29], helps to clarify the validation process by forcing the investigator to make explicit the purpose of the research. The definition of the research question will also help to assess problems related to internal and external validity. For instance, we may explore whether a certain biomarker discriminates well between cancer patients and healthy volunteers, and we may conclude—prematurely, perhaps—that we have a marker with good prospects; this mistake is related to the interpretation of the results and thus, to the external validity; it is not a problem of selection bias due to the sole inclusion of treated cancer patients and healthy volunteers. In this case, we should carry the process of biomarker development to the next phase or stage and assess whether the potential for the biomarker is limited (e.g., because of false positives arising from comorbidities, smoking, or other modifiers).

The formal structure in phases identifies questions and has the advantage of clearly dividing the development of a marker into two stages: discovery and validation. The first part of the process—discovery—ends with the list of the candidate biomarkers and includes the standardization of the analytical procedures [11]. In such study of the assay reproducibility, a rigorous sample collection may not be essential. Analytical methods include the laboratory problems and the study of the biological variability of the biomarker (intra- and inter-individual variation). The contribution of laboratory scientists is crucial here [31]. Field conditions should be reported with sufficient detail. For example, measurable concentrations of various biomarkers may be altered by the type of anticoagulant used when collecting blood specimens. The timing of blood extraction, if it is reported at all, may—wrongly—be mentioned only vaguely, without reference to the relevant clinical endpoints (e.g., time from first symptom of disease to blood draw, or from diagnosis, or from surgery) [32–34].

The validation phase, by contrast, is strongly related with the purpose of the investigation: diagnosis, prognosis, or screening. The assay should be designed and the population
sample collected coherently with the clinical question. If we are interested in clinical diagnosis, sample collection should only include subjects with relevant complaints at the time when the diagnostic problem comes up; individuals may have symptoms and signs but not an established diagnosis or treatment [8–10]. The sample for a prognostic study should involve symptomatic patients in whom testing will try to avoid progression of disease or other subsequent problems. In screening, the sample should include asymptomatic individuals who are at high risk for a disease.

Each specimen from the validation sample must be accurately linked to relevant clinical information. No valid and clinically or socially meaningful “marker research” is ever possible with only “specimens”: we study human beings, people with specific, usually complex diseases; we may thus need valid information on signs and symptoms of the disease, help-seeking pathways, self-care practices, results of exploratory and diagnostic procedures, measures of disease stage and progression, or information on potential confounders and effect modifiers [5,8–10]. In addition to Ransohoff’s well-placed emphasis on access to biologic specimens [18], we believe more emphasis is needed on the importance of access to the corresponding clinical information and, sometimes, to lifestyle and environmental information too [35,36]. The fact that a low percentage of eligible cases is analyzed may partly be due to the a priori, willing exclusion of large sections of patients, by design; for instance, in some studies on pancreatic cancer, patients without surgical tumor material were excluded, even though up to two thirds of eligible patients fell into such category [36]. One practical option would have been to accept cytohistological samples from fine-needle aspiration and endoscopic or laparoscopic procedures, which allow efficient analyses of many genetic variants. In other words, proper attention to actual diagnostic options is also required in “omics” studies.

Ransohoff discusses several uses of already-collected biological specimens [18] (e.g., blood and tumor tissue). We concur on the importance of assessing the validity and efficiency of such uses. As mentioned above, it partly depends on the phase of biomarker development. We all dislike the idea that sometimes one may “skip steps” or take “shortcuts.” Too many shortcuts leading to error are already being taken by some authors. We also believe that when an “early study” demonstrates that a marker can “discriminate” disease, in most circumstances it will be essential to do an “advanced phase” clinicoepidemiological study. Indeed, demonstrating discrimination does not automatically translate to improved clinical outcome [18,37].

3. Biological specimens are never available at random: selection biases must be a concern

In the present scientific context, in which a substantial distance exists between biology and epidemiology, it seems interesting to think about the conclusions that our debate may stimulate among basic scientists. Many of them do not particularly have a “cohort/study base framework” in mind [6], they know little about selection biases and tend to value sample size more than internal validity; that is, laboratory bench scientists often think that it does not matter how many specimens you acquire as long as you end up with a “sufficient” number of them. This is wrong and, in our view, there is a strong need for “omics scientists” to become more aware of the importance of avoiding selection biases (Fig. 1) [38]. Hence, “access to specimens” is not and cannot be made a central issue. Access to specimens must always be assessed in close connection with the—indeed central—issues of internal and external validities. Selection biases can be particularly profound in studies that integrate several types of information, such as those assessing associations among proteomic patterns, clinical outcomes, genetic variants, aspects of the patients’ medical

![Fig. 1. Is “omics” research aware enough of the importance of preventing biases due to patient and specimen selection? (a) A large variety of subjects’ characteristics, lifestyle and social factors, and clinical variables may influence the availability of biological samples and clinical certainty, which in turn may bias the findings of the study. (b) Two different, complementary paths through which heavy alcohol exposure may (in some, not all clinical settings) influence availability of tumor tissue, diagnostic certainty, and the characteristics of patients with exocrine pancreatic cancer included in a study [38].](image-url)

---

**Fig. 1.** Is “omics” research aware enough of the importance of preventing biases due to patient and specimen selection? (a) A large variety of subjects’ characteristics, lifestyle and social factors, and clinical variables may influence the availability of biological samples and clinical certainty, which in turn may bias the findings of the study. (b) Two different, complementary paths through which heavy alcohol exposure may (in some, not all clinical settings) influence availability of tumor tissue, diagnostic certainty, and the characteristics of patients with exocrine pancreatic cancer included in a study [38].
history, the molecular pathology of tumors, etc. [36,39]. Even a small percentage of subjects with missing data for each type of information will eventually result in a high percentage of subjects excluded from multivariate analyses, that is, in a low overlapping of information on all factors included in the statistical analyses. Planning specific ways to avoid selection bias is therefore important when planning the collection of biological samples. This is not easy: completeness of specimens—of what we call the “biological study base” [35,36]—is frequently hampered by ethical, clinical, and logistic factors. Readers of basic and translational articles deserve to be informed about the percentage of patients potentially eligible for inclusion who are finally included in the analyses. If such percent is low, doubts should arise about the validity of findings: while this is obvious to anyone with some “methodological conscience,” it is often unreported in “omics” papers [28].

Moreover, availability of biological samples must always be assumed to be related to characteristics of patients and to clinical procedures. Yet, it is often forgotten that availability of biological specimens does not occur at “random”; rather, availability of samples from human beings is often influenced by the characteristics of the individuals and their context; if the individuals are sick, availability will commonly depend on characteristics of patients and of clinical care (Fig. 1). Thus, for instance, substantial differences often exist between patients with and without serum or tissue available for proteomic analyses. Some studies found a relation between better quality of information on medical records and the availability of histological specimens suitable for molecular analyses [35]. Access to specimens may be linked to socioeconomic status and thus, related to different environmental exposures, quality of medical care, and treatment. Such differences may jeopardize both the internal and external validities of a study. It is seldom possible to assume, a priori, that availability (and misclassification) of specimens is “non-differential”; therefore, we must all beware of threats to internal validity [38].

Translational scientists should not be content either with simple or single measures of disease stage or progression. Rather, different measures should always be included in the study design. Example: in exocrine pancreatic cancer (EPC), it has often been observed that a clinically and statistically significant proportion of patients with more localized tumors have obstructive jaundice [39]. Such visible and unusual sign often brings patients faster to see a physician than less specific signs and symptoms as weight loss and back pain; jaundice often prompts faster exploratory and diagnostic procedures and leads to an earlier diagnosis of EPC. Of course, no clinician will be surprised by this association. But, obstructive jaundice entails a significant number of metabolic alterations whose expression, reflection, or impact upon proteomic patterns is largely unknown. Thus, a study aiming to assess the clinical usefulness of a proteomic test for the early diagnosis of pancreatic cancer cannot be content with just checking that patients have small tumors; it can neither consider enough, a priori, adjusting by stage, because a substantial proportion of patients with stage I will have the profound metabolic and proteomic changes caused by obstructive jaundice [34,39,40]. Therefore, the study will also need to check whether the presence or absence of certain signs and symptoms (e.g., cholestatic, weight loss, and cachexia) influence the discriminatory value of the proteomic or peptidomic test.

4. Conclusions

In conclusion, clinical epidemiology and “clinical biology”—need not be an oxymoron—should together develop integrative frameworks that favor discovery and validation of accurate, reliable, efficient—and above all, clinically relevant—“omics” tests. Applying such frameworks requires both old and new solutions to well-known research problems on diagnostic and prognostic tests. It also requires novel approaches that respond to the complex and dramatically new characteristics of “omics” technologies.

Development of an “omics” test includes discovery, with standardization of analytical procedures, and validation, with a rigorous sample collection according to the clinical purpose of the test, and a complete collection of information from unbiasedly selected study subjects.

Most analyses of the successes and unfulfilled promises of “omics technologies” need to address the relationships between internal issues (e.g., methodological shortcomings) and external influences (e.g., monetization of IP, which appears to powerfully favor taking methodological “shortcuts” and an excessive reductionism and fragmentation of knowledge). Monetization of IP seems partly responsible for the biotech industry having underperformed methodological, clinical, social, and economic expectations.

Internal, purely scientific reasons (e.g., methodological and epistemological) can hardly explain all failures in the application of well-known and long-proven principles of clinical epidemiology to the discovery and validation of diagnostic and prognostic tests. Clinical and epidemiological reasoning, knowledge, and methods need to be applied on a much wider scale than until now by “omics” studies that aim at making inferences relevant for human beings. Rather than adopting the values and norms of “science business,” “omics research” should apply a diversity of clinicoepidemiological models favoring integrative research; eventually, these approaches will yield higher clinical, social, and economic benefits.

Acknowledgments

The authors gratefully acknowledge scientific advice from Pere Ibern, and technical assistance from Silvia Geeraerd and Joaquín García Aldeguer. Supported in part by research grants from “CIBER de Epidemiología y Salud Pública” and Evaluación de Tecnologías Sanitarias (Exp
References


