Microarrays and molecular research: noise discovery?

The promise of microarrays has been of apocalyptic dimensions. As put forth by one of their inventors, "all human illness can be studied by microarray analysis, and the ultimate goal of this work is to develop effective treatments or cures for every human disease by 2050".[1] All diseases are to be redefined, all human suffering reduced to gene-expression profiles. Cancer has been the most common early target of this revolution[2] and publications in the most prestigious journals have heralded the discovery of molecular signatures conferring different outcomes and requiring different treatments. Yet, in today's *Lancet*, Stefan Michiels and colleagues show that, on close scrutiny, in five of the seven largest studies on cancer prognosis, this technology performs no better than flipping a coin. The other two studies barely beat horoscopes. Why such failure?

Give me information on a single gene and 200 patients, half of them dead, please. I bet I can show that this gene affects survival ($p < 0.05$), even if it does not. One can do analyses: counting or ignoring exact follow-up; censoring at different timepoints; excluding specific causes of death; exploiting subgroup analyses; using dozens of different cut-offs to decide what constitutes inappropriate gene expression; and so forth.[3] Without highly specified a-priori hypotheses, there are hundreds of ways to analyse the dullest dataset. Thus, no matter what my discovery eventually is, it should not be taken seriously, unless it can be shown that the same exact mode of analysis gets similar results in a different dataset. Validation becomes even more important when datasets become complex and analytical options increase exponentially. Typically, patients are split into separate training and validation sets. In another common approach, each patient is left out in turn, a model is built, and then checked against the excluded patient.[4]

Validation is still an analysis and can be manipulated as can any analysis. Several variants of inadequate or incomplete validation have been described.[2,5] Furthermore, when the same team does both the original analysis and validation thereof, one might consciously or unconsciously select the best-performing pair of training-validation data and analytical mode. Against this licence-to-analyse, one can use always and strictly the same method, generate randomly many training and validation sets, and examine whether results are stable. But then, as Michiels and colleagues show, the selected "important" genes rarely coincide across random replicates. Published estimates often seem excessively optimistic, probably due to serendipitous selection bias either in the analysis mode or in the validation process.

Microarrays produce information of unparalleled wealth. This information is their great, fascinating advantage—and their downfall. Let us suppose for a moment that no gene is important for any disease outcome and that it is all random noise. That scenario is scary: this noise is so data-rich that minimum, subtle, and unconscious manipulation can generate spurious "significant" biological findings that withstand validations by the best scientists, in the best journals. Biomedical science would then be entrenched in some ultramodern middle ages, where tons of noise is accepted as "knowledge". However, hopefully, some biological variables must indeed be important—but how do we suppress surrounding noise? If 30 genes determine the outcome of a specific cancer, we expect upfront that each gene (of 30 000 tested) has a 1:1000 chance on average to be truly important. The same caveat applies not only in gene-related applications, but also in proteomics,[6] and all discovery-oriented molecular research where enormous databases can be rapidly generated from just a handful of patients.[7] With such massive information, usually there cannot be any strong a-priori hypothesis that specific biological factors are more important than others. Any confident claims of "biological plausibility" sit on very slippery ground.[8]

True discovery remains a challenge in the molecular era. Routinely repeated random sampling for multiple validations is useful. Perhaps more importantly, validations should be done by several completely independent teams. I cannot stress "completely" enough here. Some journals, dismayed at the questionable replication of some molecular research,[9] propose that papers should also contain inde-

Rights were not granted to include this image in electronic media. Please refer to the printed journal.

pendent replications.[10] Yet do same-team approaches ensure independence? Any intermingling of the process for generating and replicating the hypothesis entails the danger of somehow diluting the independence of the replication.

Sample size is also essential. A recent editorial hailed the advent of "small studies with high density of data".[11] Well, I think there is no free lunch in good research. Microarrays need evidence and this cannot be obtained from a couple of small studies, no matter how high-tech. Small sample sizes might actually hinder the identification of truly important genes.[2] Molecular medicine may eventually fulfil its arrays of promises.[12] However, we should aim for many independent studies with a total of several thousand patients, a hundred-fold more than the current standard. If we truly believe that microarrays and molecular research in general are important, we should not settle for less.

John P A Ioannidis

Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece; Biomedical Research Institute, Foundation for Research and Technology–Hellas, Ioannina, Greece; and Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, USA

jioannid@cc.uoi.gr

1   Schena M. Microarray analysis. New York: Wiley-Liss, 2003: 21.
2   Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates. *Lancet* 2003 ;**362:** 1439–44.
3   Counsell CE, Clarke MJ, Slattery J, Sandercock PA. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? *BMJ* 1994; **309:** 1677–81.
4   Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000; **19:** 453–73.
5   Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 2003; **95:** 14–18.
6   Boguski MS, McIntosh MW. Biomedical informatics for proteomics. *Nature* 2003; **422:** 233–37.
7   Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004; **4:** 309–14.
8   Ioannidis JP. Genetic associations: false or true? *Trends Mol Med* 2003; **9:** 135–38.
9   Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001; **29:** 306–09.
10   Huizinga TW, Pisetsky DS, Kimberly RP. Associations, populations, and the truth: recommendations for genetic association studies in *Arthritis & Rheumatism*. *Arthritis Rheum* 2004; **50:** 2066–71.
11   Liu ET, Karuturi KR. Microarrays and clinical investigations. *N Engl J Med* 2004; **350:** 1595–97.
12   Lander ES. Array of hope. *Nat Genet* 1999; **21** (suppl 1): 3–4.

Early supported discharge: a valuable alternative for some stroke patients

In this issue of *The Lancet*, Peter Langhorne and colleagues report the benefits of early supported discharge (ESD) teams as an effective health-service option for a selected group of stroke patients. Since 1997, when Langhorne and his co-workers in the Stroke Unit Trialists Collaboration reported the advantages of integrated stroke care,[1] this care is now embedded in most stroke services worldwide. However, between countries the organisation of stroke care differs; different types of institutions participate, which have their own specific treatment modalities. Several treatment options are usually available, depending on the patient's need for further rehabilitation therapy; each option has specific admission criteria, and consists of some form of inpatient or outpatient treatment in specialised care centres. District sick bays or similar institutions might serve as hotel facilities where the patient can recover from the stroke during a limited period without receiving specific rehabilitation treatment. Some countries have inhospital rehabilitation facilities, which lengthen hospital stay, while in other countries, transfer to a specialised rehabilitation centre will take place as soon as the patient is medically stable. In their recent meta-analysis Langhorne and his fellow Early Supported Discharge Trialists refer to treatment at home given or coordinated by a multidisciplinary team as ESD;[2] in today's *Lancet* they evaluate the effects of this new health-service product.

Earlier Cochrane reviews concluded that, compared with standard care, integrated stroke-unit care reduces the odds of death, aftercare in specialised institutions, and dependency recorded at final (median 1 year) follow-up.[3] When researchers could not identify a specific factor responsible for the remarkable improvement in outcomes from integrated stroke-unit care, they switched their focus to the organisation of health care. Because of this shift in focus, the Cochrane review from 2000 showed that people who entered the ESD programme and benefited most from it tended to come from a selected group of elderly patients with discrete disabilities.[4] In the current updated Cochrane review, Langhorne and colleagues conclude that ESD should be considered part of a comprehensive stroke service. They show that a coordinated multidisciplinary ESD team yields the best results in stroke patients with moderate disability; and again they added valuable information to the organisation of health services for stroke patients. ESD might become a new health-service product in many countries.

However, many questions remain. What patients are eligible for ESD? Selection of patients in the studies included in the latest review was based on need (persisting disability), practicability (living within the local area), and stability of the medical condition. Prespecified subgroup analyses