Opinion

What Information Should Be Required to Support Clinical "Omics" Publications?

Keith A. Baggerly[1]* and Kevin R. Coombes[1]

Results of Many High-Throughput "Omics" Studies Are Difficult to Reproduce, in Part Because the Data and Methods Supplied Are Inadequate

A major goal of "omics" is personalizing therapy—the use of "signatures" derived from biological assays to determine who gets what treatment. Recently, Potti et al. *(1)* introduced a method that uses microarray profiles to better predict the cytotoxic agents to which a patient would respond. The method was extended to include other drugs, as well as combination chemotherapy *(2, 3)*. We were asked if we could implement this approach to guide treatment at our institution; however, when we tried to reproduce the published results, we found that poor documentation hid many simple errors that undermined the approach *(4)*. These signatures were nonetheless used to guide patient therapy in clinical trails initiated at Duke University in 2007, which we learned about in mid-2009. We then published a report that detailed numerous problems with the data *(5)*. As chronicled in *The Cancer Letter*, trials were suspended (October 2, 9, and 23, 2009), restarted (January 29, 2010), resuspended (July 23, 2010), and finally terminated (November 19, 2010). The underlying reports have now been retracted; further investigations at Duke are under way. We spent approximately 1500 person-hours on this issue, mostly because we could not tell what data were used or how they were processed. Transparently available data and code would have made checking results and their validity far easier. Because transparency was absent, an understanding of the problems was delayed, trials were started on the basis of faulty data and conclusions, and patients were endangered. Such situations need to be avoided.

What Should Be Supplied?

We wrote to *Nature* *(6)* to identify the 5 things that should be supplied: (*a*) the raw data; (*b*) the code used to derive the results from the raw data; (*c*) evidence of the provenance of the raw data so that labels could be checked; (*d*) written descriptions of any nonscriptable analysis steps; and (*e*) prespecified analysis plans, if any. We intended these criteria as suggestions for journals, but we see them as requirements before starting clinical trials that use omics signatures to guide treatment.

Lessons from New Information about the Duke Case

Lessons about the role of transparency—and how it might be achieved— can be drawn from events in the Duke case that are being examined in the Institute of Medicine's (IOM)[2] "Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials," which first met on December 20, 2010. This meeting included presentations by Lisa McShane of the National Cancer Institute (NCI) *(7)* and Robert Becker of the US Food and Drug Administration (FDA) *(8)*. The NCI also released documents *(9)* detailing their involvement, which we have annotated *(10)*. These details show that problems were more widespread and severe than we knew: Other clinical trials [e.g., Cancer and Leukemia Group B (CALGB) 30506] and reports *(11)* were affected, and the NCI, which compelled production of the raw data and code, was unable to reproduce the reported results. These documents and testimony (quoted below) strengthen our belief that we need transparent supplying of data and code.

Is Requiring Code Extreme?

Many journals require some posting of raw data on publication, and other reporting standards for biomarker studies have been suggested *(12)*. Dr. McShane notes [37 min, 40 s into the presentation *(7)*], however, that these precautions would not have suffced in this

---

[1] Bioinformatics and Computational Biology, University of Texas M.D. Anderson Cancer Center, Houston, TX.

* Address correspondence to this author at: Bioinformatics and Computational Biology, University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 140, Houston, TX 77030-4009. E-mail kabagg@mdanderson.org.

[2] Nonstandard abbreviations: IOM, Institute of Medicine; NCI, National Cancer Institute; FDA, US Food and Drug Administration; CALGB, Cancer and Leukemia Group B; CTEP, Cancer Therapy Evaluation Program; IDE, investigational device exemption; IVDMIA, in vitro diagnostic multivariate index assay; IRB, institutional review board; GEO, Gene Expression Omnibus; MIAME, Minimum Information about a Microarray Experiment.

case: "the statistical algorithms used to develop these classifiers are quite complex; you basically have to get the computer code." She notes, "some statisticians . . . are calling for the extreme measure that computer code should be provided when . . . articles are submitted to journals." Later, she notes the IOM must decide [50 min, 34 s *(7)*], "was this an extreme situation? And should we not try to guard against all possible extreme situations?" In the first instance, "extreme" refers to whether providing code is practical; in the latter, it asks whether this case was atypical. The severity of the problems encountered in this instance is (in our experience) atypical, but our inability to understand what was done absent the code is not. We do not see the submission of code as impractical (we have posted code ourselves). We concede that checking code at review time is likely impractical, but positive effects may still be seen if investigators know the code could be checked later.

Lack of Information Makes Spotting Basic Mistakes Difficult

We worry about practicality, but we know, empirically, that spotting mistakes in high dimensions is difficult. Identifying unlabeled data is difficult. Dr. McShane notes [26 min, 57 s *(7)*],

Keep in mind that these classifiers went through multiple levels of review. They went through journal review. They went through . . . local review at Duke. They went through NCI review in the NCI study sections. They went through CTEP [Cancer Therapy Evaluation Program] review, where they did not fare so well . . . But there were many, many people who looked at this data and found it satisfactory.

People Repeatedly Make Basic Mistakes

If mistakes were exceedingly rare, supplying extra information might not be an issue. But our own experience, as well as comments from the NCI and the FDA, suggests mistakes are not rare. In discussing studies in which omics results are evaluated retrospectively, Dr. McShane notes [1 h, 1 min, 18 s *(7)*],

it can be an enormous waste of resources and patients' time if the classifiers in fact are not locked down . . . and again, I see this a lot, people claiming they have a classifier just because they can write down a list of genes, and they don't appreciate all the many other steps that go on before you get the results of that classifier.

Dr. Becker notes [17 min, 50 s *(8)*],

it's not uncommon, we have seen in IDEs [investigational device exemptions] that we view, to have a fairly "loose" idea of exactly what the device is going . . . to show you as

a trial or an investigation gets underway. And the need to actually hone that question or hypothesis . . . is surprisingly often something that has to go back and be addressed.

Our Focus Is Not on Complex Issues; It Can Be Done

Dr. McShane notes [47 min, 53 s *(7)*],

This is not rocket science. There is computer code that evaluates the algorithm. There is data. And when you plug the data into that code, you should be able to get the answers back that you have reported. And to the extent that you can't do that, there is a problem in one or both of those items . . . It really wasn't debates about statistical issues. It was just problems with data and changing models.

As Gilbert Omenn puts it in summarizing the objections in *(5)* [40 min, 36 s *(7)*], "many aspects there were not highfalutin' statistics; they were unlabeled tables; they were mislabeled tables; these were things indisputably unacceptable."

Omic Signatures Are Medical Devices

Another factor in considering what to supply is regulatory. In January 2004, Correlogic advertised a proteomic-pattern algorithm, OvaCheck, to diagnose ovarian cancer. The claims were questioned, and the FDA issued restraining letters *(13)*. In June 2004, the FDA ruled that OvaCheck was a medical device subject to FDA review. The FDA has since indicated that omic signatures, or "in vitro diagnostic multivariate index assays" (IVDMIAs), are also devices *(14)*. Thus, IDEs must be obtained before omic signatures are used to guide therapy in clinical trials. An FDA audit team visited Duke in January 2011, reportedly because IDEs had not been obtained *(15)*. We found the following exchange between Larry Kessler and Robert Becker interesting [37 min, 33 s *(8)*]:

[Larry Kessler:] Do you feel that this is an issue of education of both the IRBs [institutional review boards] and NIH-style investigators who (a) don't really understand that some of these algorithms are medical devices and (b) the rules about understanding when you apply for an investigational device exemption . . . that the rules still apply? I have a suspicion that this is **not** *[bold added to show emphasis by speaker] well known, Bob, and you might actually have some insight here.*

[Robert Becker:] To answer your question: yes.

Opinion

How Can Such Information Be Supplied?

We use Sweave *(16)* to document our analyses. Other tools, such as GenePattern *(17)* and Galaxy *(18)*, are available. In terms of supplying raw data and metadata, the Gene Expression Omnibus (GEO) allows both but does not make depositing the latter easy.

Some Tools to Make It Easier

A recurring problem is that connections between clinical characteristics and assay data are not maintained. This problem is exacerbated by the fact that the current standard [Minimum Information about a Microarray Experiment (MIAME) *(19)*] uses just 1 unstructured text field to store sample characteristics (response and so forth). The most commonly used MIAME implementation is the MINiML format used at GEO. MINiML uses XML to store metadata about the raw data, which are stored as tables in external files. Given this structure, it may be easy to add another external file, also documented in XML, for tabulating sample characteristics. To test this possibility, we developed 2 R packages: The MINiML package reads the current format, and ArrayCube creates an R object that matches this format and adds capacity for an extra sample table. Both packages are compatible with Bioconductor and are available at http://bioinformatics.mdanderson.org/Software/OOMPA.

An Advocate's Last Word

Patient advocate Elda Railey notes [54 min, 23 s *(7)*],

I just have to ask: Why are we not asking for the data? . . . we've been calling for recommendations since . . . [2001]? . . . We're still not asking the questions and we're not getting the data? It should not get to patient trials before we ask for the data and get the confirmation.

References

1. Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, et al. Genomic signatures to guide the use of chemotherapeutics. Nat Med 2006;12: 1294–300. Retracted January 7, 2011.

2. Hsu DS, Balakumaran BS, Acharya CR, Vlahovic V, Walters KS, Garman K, et al. Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. J Clin Oncol 2007;25:4350–7. Retracted November 16, 2010.

3. Bonnefoi H, Potti A, Delorenzi M, Mauriac L, Campone M, Tubiana-Hulin M, et al. Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00–01 clinical trial. Lancet Oncol 2007;8:1071–8. Retracted January 29, 2011.

4. Coombes KR, Wang J, Baggerly KA. Microarrays: retracing steps. Nat Med 2007;13:1276–7.

5. Baggerly KA, Coombes KR. Deriving chemosensitivity from cell lines: forensic bioinformatics and reproducible research in high-throughput biology. Ann Appl Stat 2009;3:1309–34.

6. Baggerly K. Disclose all data in publications. Nature 2010;467:401.

7. *The Cancer Letter.* Lisa McShane testimony audio [MP3 file]. NCI official Lisa McShane's testimony before IOM, December 20, 2010. http://www.cancerletter.com/downloads/20110128_1 (Accessed March 2011).

8. *The Cancer Letter.* Robert Becker testimony audio [MP3 file]. FDA official Robert Becker's testimony before IOM, December 20, 2010. http://www.cancerletter.com/downloads/20110128 (Accessed March 2011).

9. National Cancer Institute. Documents referenced in the Cancer Letter: 2011 volume 37. http://www.cancerletter.com/categories/documents (Accessed March 2011). See "Internal NCI documents (zip files 1, 2 and 3) related to the Duke genomics controversy (The Cancer Letter, Jan. 7, 2011)."

10. Baggerly KA, Coombes KR. Annotation and time-line to accompany NCI documents. *The Cancer Letter,* January 14, 2011. http://bioinformatics.mdanderson.org/Supplements/ReproRsch-All/Modified/ (Accessed March 2011).

11. Potti A, Mukherjee S, Petersen R, Dressman HK, Bild A, Koontz J, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. N Engl J Med 2006;355:570–80.

12. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, on behalf of the Statistics Subcommittee of NCI-EORTC Working Group on Cancer Diagnostics. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 2005;97:1180–4.

13. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker driven research. Nat Rev Cancer 2005;5:142–9.

14. FDA. Draft guidance for industry, clinical laboratories, and FDA staff – in vitro diagnostic multivariate index assays. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079148.htm (Accessed March 2011).

15. Goldberg P. FDA auditors spend two weeks at Duke; Nevins loses position in reorganization. The Cancer Letter 2011;37:1–5.

16. Leisch F. Dynamic generation of statistical reports using literate data analysis. In: Hardle W, Ronz B, eds. Compstat 2002—proceedings in computational statistics. Heidelberg: Physika Verlag; 2002. p 575–80.

17. Mesirov JP. Accessible reproducible research. Science 2010;327:415–6.

18. Goecks J, Nekrutenko A, Taylor J, Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 2010;11:R86.

19. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat Genet 2001;29:365–71.