Reviewer’s report

Title: Availability and quality of paraffin blocks identified in pathology archives: A multi-institutional study by the Shared Pathology Informatics Network (SPIN)

Version: 3 Date: 20 August 2006

Reviewer: Matthew R Cooperberg

Reviewer’s report:

General
Overall the manuscript is significantly improved in both language and content. A few more changes should be made, after which it will be a strong contribution to the literature.

-------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

-------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

(some I would consider major but do trust the authors to address)

1. Please clarify the relationship of SPIN to the other related NIH initiatives such as caBIG and caTIEs in the manuscript
2. The relationship of the data in table 2c to the data in 2a and 2b is unclear to me. For example, 2a shows that 54% of prostate blocks had tumor; but 2c gives a 91% retrieval rate for blocks with tumor.
3. Figure 4 needs to be discussed in the results section, and the reason that Harvard cases were 2000-2004 and the other 3 sites were 1990-1999 should be stated in methods.
4. How do the SPIN tools process electronic pathology records to glean clinically useful information? Path reports from different institutions will have widely different formats, information content, etc.
5. One limitation still missing from the discussion is that you are extrapolating without basis from 4 major academic centers to countless pathology repositories of unknown size and structure who are expected to provide the “millions” of specimens. Whether the quality of tissue preservation and access is the same in smaller centers remains unknown. Alternatively, if SPIN is realistically going to be a collaboration among major centers (this would be my guess barring a major incentive for voluntary participation among smaller non-research labs), then this should be explained.
6. I still feel the statement in the first paragraph that institution-specific tissue banks have "limited utility" is unfair and untrue. There obviously are limitations to these resources, but the extensive publication record from institutional tissue banks frankly disproves this statement.
7. The number of centers participating in SPIN, if more than the 4 centers in this study, should be stated. My understanding of SPIN is that so far it is still a relatively small pilot collaboration among academic centers. Unless it is a much larger number, I would still advise that you not state the SPIN utilizes “millions” of specimens (in abstract and intro). That may be its potential, but it does not seem to be there yet.
8. p5: much pathology material is stored in outpatient or free-standing facilities, not just hospitals
9. in methods, please clarify if banked tissue is derived from biopsies, excised specimens, or both
10. there is too much restatement in the text of data given in the tables (esp 2nd and 5th paragraphs of results)
11. the gray scale difference between Indiana and UCLA and between Prostate and Colon in the figure 4 is insufficient. Maybe use patterns?
12. The discussion is still a bit too verbose; there is too much restatement of arguments already made in intro or earlier in discussion.
13. p8: anonymizer misspelled
14. p15: reference to triage to womens/childrens hospitals is unclear

-------------------------------------------------------------------------------

Discretionary Revisions (which the author can choose to ignore)
1. In general there is too much use of quotation marks (e.g., archived tissue, tissue bank collections, etc., and again on p15. This is distracting and unnecessary.
2. I do not feel figure 1 adds much to the manuscript – most readers will be familiar with Excel spreadsheets; the description of the spreadsheet in methods can also be made more concise.
3. In tables 2a and 2b, the 2nd line (% attrition) is not necessary, as these data are evident from the % retrieval line (i.e., 100-retrival = attrition by definition)

**What next?:** Accept after minor essential revisions

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No

**Declaration of competing interests:**

I declare that I have no competing interests