Author's response to reviews

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

Authors:

Ashkokkumar A Patel (patelaa@upmc.edu)
Dilipkumar Gupta (dkgupta10@hotmail.com)
David Seligson (dseligson@mednet.ucla.edu)
Eyas M Hattab (ehattab@iupui.edu)
Ulysses J Balis (ubalis@partners.org)
Thomas M Ulbright (tulbright@iupui.edu)
Isaac S Kohane (isaac_kohane@harvard.edu)
Jules J Berman (jjberman@alum.mit.edu)
John R Gilbertson (jgi@alum.mit.edu)
Sarah Dry (sdry@mednet.ucla.edu)
Osvaldo Schirripa (oschirripa@mednet.ucla.edu)
Hong Yu (hyu@mednet.ucla.edu)
Michael J Becich (becich@pitt.edu)
Anil V Parwani (parwaniav@upmc.edu)
Shared Pathology Informatics Network (spin) (spin@pop.nci.nih.gov)

Version: 4 Date: 5 December 2006

Author's response to reviews: see over
December 5, 2006

Dear Editorial Team,

We would like to thank your team and the reviewers once again for taking the time to comment on our manuscript, “MS: 1607386739368951 - Availability and Quality of Paraffin Blocks Identified by the Shared Pathology Informatics Network (SPIN): A Multi-institutional Study”.

We would like to thank Drs. M.R. Cooperberg and Johanna Westbrook for helping us improve our manuscript with their input. The valuable comments by the reviewers have all been taken into account in the extensive revision of the manuscript. We would like to highlight some of the changes reflected within this revision of our manuscript:

- Eliminated all references to unsubstantiated claims/conclusions of benefits of SPIN tools. We have focused on only results of this study and how they will benefit future SPIN studies and the research community at large.

- We have given a brief description and references to how the SPIN tools work on p12.

- Addressed why only some of the SPIN sites used the SPIN tools and others did not with this study on p7 (methods) and p13-14 (discussion). The Authors would like to point out that at the beginning of this study, all of the sites were in the process of installing the SPIN software at local facilities. In order to avoid delay with this study, we opted to use “routine” case finding methods that were normally used by the local tissue banks for specimen requests. Today, each of the 4 major centers that are members of the SPIN have over 1 million cases loaded on the nodes of the network.

- Reduced the body of the paper to 15 pages.

In addition, we have answered (in bold) below to the suggestions or questions raised by the reviewers. We hope that we have addressed all of their issues to their satisfaction.

If you have any further questions or concerns, please do not hesitate to get in touch with me. I thank you once again for your time and considering publishing our work.

Regards,
-Ashok

Ashokkumar A. Patel
patelaa@upmc.edu
Phone: 412-623-7839
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. Please clarify the relationship of SPIN to the other related NIH initiatives such as caBIG and caTIEs in the manuscript
   see p12 (discussion)

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.
   Rephrased statement on p9. Please note in your comment (table 2a) that 54% represents the number of prostate cases that had ≥ 1 block with tumor, whereas the 91% in 2c represent the subset of tumor blocks from all the blocks retrieved (benign and tumor blocks) and reviewed for this study.

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.
   See p6-7 of 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.
   See p12-13 (discussion describing the SPIN tools)

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.
   Changed “millions” to “vast amounts”. Also see comment in the bullet points above.

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.
   Eliminated the phrase “limited utility”. (p4)

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.
   Changed “millions” to “vast amounts”. Also see comment in the bullet points above.

8. p5: much pathology material is stored in outpatient or free-standing facilities, not just hospitals
   changed “hospital” to “medical centers”. See top of p5.

9. in methods, please clarify if banked tissue is derived from biopsies, excised specimens, or both
   The cases selected for this study focused on excised specimens, so that we can assess the availability of adequate amount of tumor tissue for research use. See p5-6.
10. there is too much restatement in the text of data given in the tables (esp 2nd and 5th paragraphs of results)  
We have attempted to rephrase some of this text. See p9-11.

11. the gray scale difference between Indiana and UCLA and between Prostate and Colon in the figure 4 is insufficient. Maybe use patterns?  
Changed all figures to 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.  
We have attempted to rephrase some of the discussion and deleted some sections to shorten the manuscript.

13. p8: anonymizer misspelled  
Corrected.

14. p15: reference to triage to womens/childrens hospitals is unclear  
Moved this reference to the methods section on p6.

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.  
We have eliminated all but 2.

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.  
Eliminated this figure.

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)  
Eliminated the % attrition row of the table and reformatted the latest version.