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:

Ashkookumar 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)
Issac S Kohane (isaac_kohane@harvard.edu)
Jules J Berman (jberman@alum.mit.edu)
John R Gilbertson (gi@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: 3 Date: 25 July 2006

Author's response to reviews: see over
Dear Editorial Team,

We would like to thank your team and the reviewers 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 A.G. Glass 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:

- Change title to “Availability and quality of paraffin blocks identified in pathology archives: A multi-institutional study by the Shared Pathology Informatics Network (SPIN)” in hopes to draw focus to the paraffin archives rather than the SPIN tools themselves.

- Shorten the body of the paper from 20 pages to 16 pages.

- Focused this manuscript more on each institution’s ability to retrieve slides/blocks with significant presence of tumor material that would be useful for research studies. Emphasis was given to the quality and quantity of the paraffin blocks retrieved rather than how cases were initially identified or the utility of the SPIN tools themselves. The authors plan to address the later issues in a follow up manuscript on how using the SPIN tools uniformly at each participating institutions will resolve these biases.

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

It was important for our preliminary data to highlight that adequate tumor material was retrievable from the paraffin archives, if tools, such as those from the SPIN group are to be successful and beneficial to the wider research community. We believe this work to be very critical for future collaborations and comparative research activities that will utilize powerful health information technology tools.

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

1. Throughout the introduction and discussion, the authors seem somewhat over-enthusiastic about SPIN, and unfairly derogatory regarding existing institutional tissue banks. As described, SPIN will facilitate identification of tissue resources stored at multiple institutions. It will coordinate rather than replace institutional repositories. The characterizations in the first paragraph, for example, that the utility of existing tissue banks is “limited to a very narrow focus” and that they are “usually abandoned after their funding period terminates” are simply not true; individual institutions are developing tissue banks to answer a wide range of scientific and clinical questions. These have a wide focus and are often funded from multiple sources over the long term; indeed, without existing tissue banks, SPIN would have no raison d’etre. SPIN is not a mechanism for sharing tissue resources per se, and identification of resources at a given institution is by no means a guarantee that that institution will share those resources.

Although we have made appropriate changes to reflect the reviewer’s comments, we agree to the fact that many individual institutions are developing tissue banks to answer a wide range of scientific and clinical questions. The authors’ intention was to highlight that many NCI funded projects dealing with tissue banking are primarily focused on a specific disease or organ type (i.e. SPORE tissue banks, Collaborative Breast Cancer Tissue Resource, Collaborative Prostate Cancer Tissue Resource, etc). Other groups, such as Collaborative Human Tissue Network, have a wider focus on the types of tissues collected, but are mostly prospective targeted collections based on a researchers needs. The authors hope to highlight the vast number of paraffin tissue archives that are processed at every major pathology lab nationwide for diagnostic workup; they are processed and stored for at least 7 years (for med-legal reasons). Being able to tap into electronic pathology reports using the SPIN tools would allow every hospital to attain an “instant” tissue bank. The SPIN tools do not need existing tissue banks for it to achieve its goals. It is focused on utilizing the electronic pathology reports for identifying paraffin blocks from cases of interest based on information written in them.

It is also important to note that although individual institutions often get funded from multiple sources for their tissue banks over the long term, the amount of funding and banking activities will vary based on the amount of funding received. If there is not a constant level of funding stream, the ability to meet the research demand will not be met.

The guarantee that any given institution will share their archived material to outside researchers is beyond the scope of the SPIN project. Those that are members of the SPIN study are in agreement that proper IRBs are in place to share de-identified electronic pathology reports within SPIN. The scope of the SPIN project is to develop technologies that would essentially allow “peer-to-peer” sharing of data with participating institutions within the network. Although, the developers from the SPIN group have processes in place to safeguard patient privacy, HIPAA, and other regulatory requirements, requiring every participating institution to share their tissues with other members of the network is beyond the objectives of this study. As more collaborative or “team” science approaches are pushed by NCI/NIH, such as projects like caBIG, individual institutions will have to work with each other and policy making agencies to resolve these issues on a broader scope. However, for now, individual researcher initiated collaborations will be necessary with colleagues where tissues resources are available.
2. There is wide variation in the site-specific methods for query, collection, review, etc.. What is the rationale for not standardizing these methods? The point is made in the conclusions that a future study will use SPIN-only queries, but the wide variation in this study, especially the fact that two institutions used local system queries and two queried SPIN indirectly, requires both justification in the methods and exploration of differences in the results/discussion.

This is addressed in the third paragraph of the methods section (p.6).

…It is important to note that each institution has different databases and specimen query mechanisms to identify cases of interest by default, and as such, the different methods used to identify cases are part of the routine workflow and represent each site's optimal search procedure. An advantage the SPIN tools intend to bring will be the uniformity or at least functional inter-communication between each member institutions surgical pathology reports within a common information technology framework.…

As well as part of the discussion (p. 13):
… Using the local preferred methods for case identification, such as the LIS or cancer registry tools was easier because it utilized the existing workflows and personnel without involving the SPIN tool and personnel…. these assorted “normal” methods used at various institutions bring to surface one of the key barriers for advancing translational research, the lack of quality biospecimens and its access. This highlights the importance of implementing standardized tools, such as those from the SPIN group, for identifying cases within a network of institutions and reference labs that house large archives of paraffin blocks. As a follow up to this study, we intend on solely using SPIN tools in order to estimate the resources required to use SPIN alone and to determine what sample yield that restriction generates. The two results can then be compared for a more complete picture of the viability of the SPIN tools...

3. Why were methods of specimen collection and review not standardized? For example, outside referrals of rare cases were included for Indiana but not for UCLA. This seems like a source of significant potential bias.

Similar to the above question, this was addressed as part of the discussion (p. 13).

4. One crucial component of institutional tissue banks is correlated clinical data. A list of 100 prostate cancer cases, for example, is of limited utility without details regarding stage, grade, treatment, and ideally outcomes. Is there information accessioned for SPIN outside of what is available in the pathology report? What fields are in the SPIN Excel template?

We agree with the reviewer that biospecimens by themselves have limited utility. It is important that readers know that the SPIN tools are primarily for allowing cancer researchers to query standardized and de-identified electronic pathology record information to identify tissue blocks that could be used for research studies. Thus, the data associated with these samples will be limited to what is on the pathology report themselves. However, with collaborations with other large informatics initiatives such as the Cancer Bioinformatics Grid (CaBIG), the advances made
by the SPIN group can be leveraged with other clinical databases, such as the cancer registry, to expand the associated data sets available with biospecimens.

The SPIN Excel template was a data collection form used for this study to standardize the pathology review process. A section describing the SPIN study data form was added in the methods (p.7).

5. The manuscript in general, and the methods section in particular, is too verbose. I am not sure how the fine details of the query mechanisms, for example, are relevant to the reader.

The methods section was revised and shortened to give an overview at all of the participating SPIN institutions. Detailed methods of case identifications at each institution was eliminated to remove the confusion and focus drawn to it. Although, we acknowledge that there are biases to the case identification process, the primarily goal of this manuscript is to highlight the quality and quantity of blocks retrievable for potential researchers by using or not using SPIN-like tools.

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

p4. It seems somewhat early to propose that SPIN will coordinate “millions” of tissue samples. There are only four institutions included in this study; even were the project expanded significantly, it would presumably only incorporate academic centers, and perhaps major private sector pathology reference labs. Will the expansion be funded wholly by the NCI? What are the incentives for participation?

There is potential for any size institute, private or academic, that has an electronic pathology records system that can participate within SPIN. With current limited funding at the NCI, it is undetermined if the expansion of this network can be funded wholly by the NCI. However, with projects like caTIEs, within the caBIG program, leveraging on some of the technologies initiated by the SPIN group, some funding might be available through alternative methods directed by the NCI/NIH.

As the reviewer stated, the current incentive for participation with the network is primarily driven by research demands at large academic centers and some major private reference labs. In addition, there could potentially be incentives for large pharmaceutical companies to provide funding to support growth of these networks in order to gain access to biospecimens where otherwise they would have limited tissue samples to validate potential biomarkers within their pipeline.

p4. Figures are given regarding national pathology resources. Are there any data on the proportion of these specimens which include tumors? See below.

p6. What is the reference / basis for the statement that 10-20% of pathology reports include a cancer diagnosis?

We were not able to find a strong literature reference that would give data on the proportion of pathology specimens which include the presence of tumor. This information was based on the consensus of the practicing SPIN pathologists participating in this study. Thus, to eliminate any confusion, we have taken out any reference to this statement.
p6. The discussion states that some tumors types initially investigated as rare tumors were discarded due to heterogeneity among sites. This is an important issue to discuss, and should be mentioned early, in methods and/or results not just in discussion. How were the original and replacement rare tumors selected – at random, by consensus of the investigators, etc.?

We have addressed this in the second paragraph of the methods section (p.6) and in the discussion (p.14).

p6. What does the “n” refer to in “One hundred random n cases”?

Deleted “n”. Typo.

p28. Figures 2, 4, and 6 are tables, not figures.

Corrected.

p27. The mix of 2D and 3D bar graphs is distracting and unnecessary. 2D should be used consistently unless there is a specific reason for 3D.

Corrected.

p29. Figure 3 should be formatted as a standard stacked column chart, not a 100% percent chart.

Corrected.

p31. In figure 5, why are there only Harvard cases after 1999?

Study pathologist pulled cases that would allow for the fastest retrieval by support staff; primarily pulling any locally available slides/blocks with off-site resources not being utilized to meet the study deadline.

===============================================================================

Peer Reviewer: Andrew G Glass

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

As stated above, the focus of this paper needs to be on the SPIN activity, how it has/will advance the field and how the current study fits into that. In any case, the text needs to be significantly shortened at the time of focusing on the key questions. If these things can be done, the paper is worth publishing.

Although the reviewer would like to see this manuscript to be more focused on the SPIN tools themselves, we are only able to highlight the feasibility of implementing such tools. As mentioned
above, this manuscript’s primary message is to determine the number of paraffin blocks that can be retrieved and access the quality of the presence of tumor material that potentially could be used by researchers. If this cannot be proven to be significant, then no tools, including those of the SPIN, would be worth developing. We authors are currently underway to examining how many of the issues raised by both reviewers with the case identification process can be resolved by standardizing and using the SPIN tools uniformly at all participating institutions. We have also attempted to shorten the manuscript without compromising the take home message.