Telepathology for Routine Light Microscopic and Frozen Section Diagnosis

Katherine Chorneyko, MD,1 Ronald Giesler, ART,1 Deborah Sabatino,3 Catherine Ross, MD,1 Francesca Lobo, MD,1 Hafez Shuhaibar, MD,1 Vicky Chen, MD,1 Leela Elavathil, MD,1 Franco Denardi, MD,1 Saira Ansari, MD,1 Samih Salama, MD,1 Victoria LeBlanc, PhD,2 Geoff Norman, PhD,2 Brian Sheridan, MD,3 and Robert Riddell, MD1

Key Words: Telepathology; Light microscopy; Accuracy; Feasibility; Frozen sections

Abstract
Telepathology (TP) uses telecommunication linkages to electronically capture, store, retrieve, and transmit images to distant sites. We assessed the feasibility of a dynamic real-time TP system for light microscopic (LM) diagnosis of anatomic pathology specimens, including frozen sections. Six pathologists, in 2 separate periods, read a set of 160 retrospectively retrieved slides (80 of which were frozen sections) by TP and LM. Reading times were recorded. Diagnoses were compared with the reference diagnosis (established by a group of 5 independent pathologists) and graded on a scale of 0 to 2 (2, correct; 1, incorrect but no clinical impact; 0, incorrect with clinical impact). Overall, LM was more accurate than TP compared with the reference diagnosis (score, 1.68 vs 1.54). There was no difference in accuracy between frozen section and paraffin-embedded tissue. Intraobserver agreement ranged from 82.5% to 88.2%. The average reading time was 6.0 minutes for TP and 1.4 minutes for LM. During the study, reading time decreased for TP but not for LM. These results show that despite marginally lower accuracy and longer reading times, TP is feasible for routine light microscopic diagnosis, including frozen sections.
In our department, anatomic pathology services are provided in 4 geographically separate hospital sites. Although we have had a regionalized laboratory service since 1970, the administration of these laboratories has been consolidated only recently into 1 laboratory medicine program. The long-term goal is to centralize laboratory services, including anatomic pathology. Some services, however, such as an operating room consultation service with frozen sections, will be necessary at each site. Therefore, evaluation of a TP system that would allow provision of a frozen section service at a distance was of interest. The purpose of this study was to assess the feasibility of dynamic real-time TP by evaluating the diagnoses made by TP compared with diagnoses made using conventional light microscopy (LM) and also the time taken to make those diagnoses.

Materials and Methods

Equipment

The Apollo Software (Falls Church, VA) for TP was used in the study. The equipment was packaged in two main configurations: one used at the host site and a second used at the peripheral site. The host configuration consisted of the following:

1. A proprietary microscope fitted with fully robotic X, Y, Z stage control, light intensity control, objective selection (2.5×, 4×, 10×, 40×, 60×), and a digital camera
2. A Pentium II personal computer (PC) operating on Windows '95 (Microsoft, Redmond, WA), loaded with proprietary software and fitted with mouse, keyboard, and touch-pad peripherals
3. A 20" high-resolution computer monitor
4. An integrated gross examination stand (manually controlled, at the host site) fitted with a digital camera
5. An integrated desktop teleconferencing module fitted with full motion audio and video

The peripheral configuration was a duplicate of the host configuration with the exception of the robotic microscope and the gross examination stand. The proprietary software, loaded on the host and peripheral PCs, provided interchangeable control over the robotic microscope and its range of operational controls between operators at either the host or peripheral stations. The operator viewed a 640 × 480 dynamic image of the microscopic slide. The field of view could be selected by point-and-click action of the mouse or by continuous travel controlled via the touch-pad peripheral device. The speed and direction of the dynamic image was controlled electronically by the operator. The capturing, transmitting, and storing of digitized images was operator controlled. The digital camera enabled the capture and display of a 1,520 × 1,144 × 24-bit color image. The software used the JPEG format for image compression. The operators at the host and peripheral sites could converse with one another via the teleconferencing module. “White-boarding” of digitized images was made possible through the proprietary software resident in the host and peripheral PCs. A whiteboard is the equivalent of an on-screen blackboard. A digitized static image of gross tissue or a microscopic area of interest can be viewed simultaneously and discussed by members of the telepathology session. Telepathology participants then can use annotation tools to introduce text, headers, or lines of varying thickness to draw attention to specific features. Once drawn, the annotation tools can be viewed over the original digital image, enhancing discussion or directing dissection.

The host package was stationed permanently in the pathology laboratory of one of the participating hospitals. The peripheral station rotated, on a monthly basis, to the pathology departments of the other 3 participant hospitals. The system method required a minimum of 512 kilobytes per second bandwidth to manage the control of the robotic microscope, transmit and display the dynamic and static digitized images, and enable the teleconferencing module. While the vendor preference was a T1 level carrier system, the equipment could be configured to operate over communication lines other than T1, including IMUX Quad ISDS-BRI, ISDN-PRI, ISO Ethernet, DS3, ATM, NTSC Broadcast Quality Microwave, and Satellite. For the purposes of this trial, the telepathology equipment was operated using a preexisting OC3 backbone networking the 4 major hospital sites. The optical carrier level of the OC3 fiber optic network provided transmission of digital signals at a rate of 155.52 megabytes per second (Mbps) compared with the T1 rate of 1.544 Mbps.

Selection of Cases

A group of 5 pathologists (including C.R., F.D., S.A., and S.S.) representing the 4 hospital sites chose 160 cases, 80 of which were frozen sections. The cases were selected randomly from a routine surgical day in the different sites, and they were representative of a routine spectrum of cases seen across the region. Duplicate diagnoses were rejected. Frozen sections were chosen sequentially from a random surgical day. Each slide was reviewed and discussed by the group. Quality and diagnostic features were assessed. A reference diagnosis was agreed on. For most cases, one representative diagnostic slide (H&E) was chosen, but 5 cases had more than 1 slide (range, 2-4 slides). In 13 cases, additional histochemical or immunohistochemical studies, which were thought to be relevant to the diagnosis, were reviewed and made available to the telepathologist if
requested. For each case, a checklist of diagnoses was generated. The cases were divided randomly into 2 sets of 80 containing equal numbers of frozen section and paraffin-embedded slides.

**Reading of Cases**

A group of 6 pathologists (K.C., R.R., V.C., H.S., L.E., and F.L.), different from the 5 pathologists involved in selection of cases but equally representing the 4 different hospital sites, read the slides by LM and TP, one set of 80 slides by TP and the second set of 80 slides by LM. Pathologists were given code names to ensure anonymity. These code names were maintained throughout the study, including data analysis. Interpathologist discussion was prohibited. To minimize variables in interpretation of answers, they chose one diagnosis from the checklist provided for each case. These were organ-related generic checklists of diseases. The time required to examine the slide and make a diagnosis was recorded for LM and TP. Both sets were read over a period of 1 month. Three months later, the slide sets were reversed and randomized so that the set initially read by TP subsequently was read by LM and vice versa. This was done to minimize the effect of memory. In the second reading period, the telepathologists also rated their confidence in the diagnosis according to the following scale: 0, would not venture a diagnosis based on this slide, defer diagnosis; 1, would give a differential diagnosis but defer for final diagnosis (25% sure of diagnosis and 75% unsure); 2, would give a provisional diagnosis only (75% sure of diagnosis and 25% unsure); 3, confident that diagnosis is correct, likely it would not change (95% sure of diagnosis and 5% unsure); 4, totally confident of diagnosis, would be able to sign out case.

**Data Analysis**

After the slides were read, the checklist diagnoses and comments of each pathologist were transcribed onto a separate data sheet. The diagnoses made by TP and LM for each case were scored by a group of 3 pathologists (C.R., F.D., and S.A.) from the original group of 5 who chose the slides. The group that analyzed the data was blinded as to pathologist and modality of diagnosis (ie, TP or LM). The diagnoses were given a score of 0, 1, or 2: 0 if the diagnosis by TP or LM was incorrect and would alter patient management; 1 if the diagnosis was not completely correct, but the error was such that it would have no consequence for patient management; or 2 if the diagnosis agreed completely with the reference diagnosis. The TP and LM answers were scored separately. Deferred diagnoses were scored as 1. The values were entered into a database and analyzed. The data were submitted to a repeated analysis of variance. The slide number (1 to 160) was the “subject” in the analysis, with type (frozen, fixed) as a between-subject factor and pathologist (1 to 6) and technique (TP, LM) as within-subject factors.

**Results**

The degree of accuracy compared with the reference diagnosis was high with both techniques: the percentage of diagnoses scoring 1 or 2 by LM and TP were 91.1% and 86.6%, respectively. Accuracy was not significantly different for the type of section examined (ie, frozen section vs paraffin-embedded section); however, it was significantly different for individual pathologists. Half of the pathologists were just as accurate by TP as by LM, while the other half were significantly more accurate by LM. Overall, LM was more accurate than TP compared with the reference diagnosis (1.68 vs 1.54; P < .01). In 6.3% of cases, the diagnosis (TP and LM) was deferred. These were initially scored as 1 since technically they were not correct. However, they were not entirely relevant in the assessment of LM vs TP since no diagnosis was given. Therefore, the statistical analysis was repeated with deferred diagnoses removed. In this analysis, LM remained more accurate than TP (1.76 vs 1.68; P < .01), and accuracy was not affected by the type of slide (frozen section vs paraffin-embedded) or by pathologist. In addition, the order of the slide sets had no effect on accuracy.
Intraobserver agreement ranged from 82.5% to 88.2%. In other words, on a case-by-case basis, individual pathologists were in agreement with themselves more than 80% of the time. There was a smaller number of cases in which the agreement was very close (a score of 1 by TP and 2 by LM or vice versa).

The average reading time was 5.9 minutes for TP (range, 3.8-7.8 minutes) and 1.3 minutes for LM (range, 0.8-1.7 minutes). The average reading time for TP decreased from 6.6 minutes per slide to 5.4 minutes per slide between the first and second reading periods. The average reading time per slide for LM stayed approximately the same between the first and second reading periods (1.4 minutes and 1.3 minutes, respectively).

For TP, the average deferral rate was 9.1% with a large range (0.63%-22.5%), while for LM, the average deferral rate was 3.5% with a smaller range (0.63%-6.3%). In 44 of 160 cases, all 6 telepathologists scored 2 (correct) by TP. The type of cases for which telepathology seemed to perform best were skin samples for margins. In 16 cases, 3 or more pathologists scored a 0 by TP. In 7 of these cases, 3 or more of the pathologists correctly diagnosed the cases by light microscopy. These 7 cases were a lymph node with metastatic carcinoma, an acinic cell tumor (frozen section), a breast biopsy specimen with calcification (frozen section), a bowel biopsy specimen for Hirschsprung disease in which ganglion cells were absent (frozen section), Paget disease of skin, a parathyroid adenoma (frozen section), and carcinoma in situ in a biopsy specimen from the large bowel.

The correlation between confidence in diagnosis and accuracy was 0.321 for TP and 0.345 for LM (P < .01).

Discussion

Many studies have shown good to excellent correlation between LM and TP. TP has been used in the diagnosis of gastrointestinal biopsy specimens, breast specimens, prostate biopsy specimens, skin specimens, congenital heart malformations, and other tissue and specimen types. In our study, LM was more accurate than TP, and this difference was statistically significant. This result is different from other published reports assessing the accuracy of TP. In an early study that assessed differences between LM and video microscopy in frozen section diagnosis of breast lesions, there was no significant difference in accuracy between the techniques (area under the receiver operating characteristic curve, 0.9967 for LM vs 0.9962 for TP). There was, however, a trend toward better results by LM. Subsequent studies also using a dynamic telepathology system have shown a concordance rate between TP and LM ranging from 88.1% to 100.0%. There were many factors in the present study that could account for these discrepancies. The training session for pathologists on the system was relatively short. The pathologists did not have the opportunity to test themselves with a wide variety of cases to gain experience with the system and the types of cases they individually might have trouble with. The pathologists were reading the TP slides in an office that was not their own, and they were without the benefit of their own reference material. They did not consult other pathologists when, in reality, for some cases they might have done so. The cases were a mixture of cases from all sites and included cases with large areas to be examined, such as endometrial curettage specimens and transurethral resections of prostate. In the study by Dunn et al, these types of cases were deferred routinely for LM because of the long time required to adequately examine such cases. All of these factors likely contributed to the lower accuracy in our study compared with other reports in the literature, but they represent real considerations and problems that should be considered when considering implementation of a telepathology system.

One of the main purposes for undertaking this study was to assess the performance of TP for frozen section diagnoses. Despite better overall accuracy of LM than TP, the type of section examined, frozen or fixed, did not affect accuracy, which has positive implications for the provision of a frozen section service to remote sites. In fact, TP already has been used successfully for this purpose in Norway and Japan.
**Figure 3** Paired accuracy results for telepathology (TP) and light microscopy (LM), pathologists 1 to 6. Accuracy score: 0, error; 1, error with no effect on patient management; 2, correct.

**Table 1** Accuracy Scores of Telepathology (TP) and Light Microscopy (LM) by Type of Case

<table>
<thead>
<tr>
<th>Site</th>
<th>No. of Cases</th>
<th>Telepathology Score</th>
<th>Light Microscopy Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Frozen section</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>9</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>Skin†</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>6</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Genitourinary</td>
<td>19</td>
<td>11</td>
<td>26</td>
</tr>
<tr>
<td>Head and neck</td>
<td>25</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>Soft tissue/bone</td>
<td>5</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>Lung/mediastinum</td>
<td>6</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Lymph nodes</td>
<td>5</td>
<td>13</td>
<td>30</td>
</tr>
<tr>
<td>Bone marrow</td>
<td>0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Permanent section</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>8</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>Skin</td>
<td>8</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>14</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td>Genitourinary</td>
<td>29</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Head and neck</td>
<td>8</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>Soft tissue/bone</td>
<td>6</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>Lung/mediastinum</td>
<td>7</td>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td>Lymph nodes</td>
<td>6</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Bone marrow</td>
<td>3</td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>

* Data are given as percentages of total scores for each organ system for each modality, TP or LM. Accuracy score: 0, error; 1, error with no effect on patient management; 2, correct.

† All cases were for margins.
and in many cases, a one-line diagnosis is insufficient for clinical management. For example, prostatic adenocarcinoma in a needle biopsy specimen is reported with a Gleason score, percentage of the biopsy area involved, the presence or absence of perineural invasion, and any other pertinent findings. A checklist diagnosis made it easier to perform statistical analyses for the study and permitted the assessment of diagnostic accuracy; however, it did not permit assessment of the overall accuracy of TP compared with conventional LM. This assessment was beyond the scope of the present study but is an important aspect of TP that should be addressed in future studies.

All of the reading pathologists noted that cases with small areas of tissue to be examined seemed better suited for TP than those with larger areas. There are many reasons these latter types of cases were difficult. The real-time resolution of the system was inferior to a regular light microscope, which made scanning at low power for small focal abnormalities difficult. If the image was captured as a still digital image, the resolution was comparable to a light microscope, but it took approximately 7 to 8 seconds to acquire a high-resolution digitized image. Therefore, for most samples, it was impractical and time-consuming to digitize many areas of the slide or the entire slide. The slower stage speed compared with a light microscope also made it more difficult to examine tissue samples with large areas. One slide of a prostatic curettage sample could take up to 90 minutes in the real-time mode, and even after spending this amount of time, most of the pathologists would not be confident that they had assessed the slide accurately unless they had digitized virtually the entire area. It also was noted that cases that required frequent changes between low and high power were frustrating owing to the time necessary to change objectives and then reposition the slide to the area of interest and refocus. Technologic advances that will improve real-time resolution and make stage manipulation comparable to that of a light microscope will, no doubt, advance the use of TP.

Such technologic improvements also will decrease the amount of time necessary to read a slide by TP compared with regular LM. This was a major concern of the reading pathologists in the present study, in which it took on average 4 to 5 times longer to read a slide by TP. It was encouraging, however, that there was a learning curve in the use of the technology as evidenced by the decrease in reading time per slide by TP between the first and second reading sessions. The reading time per slide did not change during the study for LM. In the implementation of a TP system, the implications of the greater amount of time necessary to read a slide by TP compared with LM would need to be considered in light of potential costs and time involved in travel by a pathologist to a remote site or in having a pathologist stationed at a remote site. This sort of
cost analysis of TP was undertaken by Agha et al. using the TP service between the Veteran Affairs Medical Center, Iron Mountain, MI, and the Veterans Affairs Medical Center, Milwaukee, WI. In their base analysis, they found that a courier method was the most economic but that TP was less costly than on-site pathology. Their analysis cannot be directly extrapolated to other sites, particularly other countries, since remuneration, technical costs, and health care systems vary widely. However, their findings show that in certain scenarios, TP can be economically feasible, particularly when one considers that for the purposes of frozen section, courier methods are not an option.

In the present study, the confidence in a diagnosis was not highly reflective or predictive of the accuracy of the diagnosis for either TP or LM, although the correlation between confidence in the diagnosis and accuracy was slightly better for LM. This low correlation may reflect the large spectrum of cases assessed in the study. The cases were culled equally from all of the 4 sites, and for all pathologists in the study, there were certain types of cases with which they were unfamiliar; for example, only 2 of the 6 reading pathologists regularly encounter head and neck cases, which in our area, are performed at only 1 hospital site.

This study has shown that TP is feasible for routine light microscopic diagnosis, including frozen sections, and it has provided valuable insights into the operational characteristics of a TP system, its strengths, and its weaknesses. This sort of data is important for planning further validation studies and when planning to implement such a system in an anatomic pathology laboratory setting. Technologic advances are certain to improve or eliminate many of the weaknesses identified in this assessment, making telepathology an important part of most laboratories in the near future.

References


