The Banff classification revisited

Kim Solez and Lorraine C. Racusen

1Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, Alberta, Canada and 2Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

From small beginnings in 1991, the Banff working classification of renal allograft pathology has grown to be a major force for setting standards in renal transplant pathology, and is widely used in international clinical trials of new antirejection agents. The meeting, classification, and consensus process have unique history, and look poised to continue for another several decades as the embodiment of the process for setting global standards in pathology. The Banff meetings have expanded from renal allograft pathology to most other areas of solid organ transplantation, and increasingly incorporate international working groups, so that productive collaborative activity is ongoing, creating an important dynamic process enhancing clinical success in transplantation. On the other hand, despite the successes of the working classifications and ongoing collaborative efforts, there are limitations in this and other pathological classifications, related to potential for sampling error, issues of reproducibility when implemented globally, and lack of widespread application of morphometry with true quantification, and lack of formal integration of molecular and genomics data into the classification. These are issues to be addressed in the ongoing international Banff conferences and related activities, described below.

BACKGROUND

The need for a new classification was quite obvious in 1991 when the Banff Classification of Allograft Pathology began. Most people learning kidney transplant pathology had no direct mentor and were learning from outdated textbooks. Although a few individual classifications for renal allograft rejection had been developed, none were in general use. The Banff Classification of Kidney Allograft Pathology had its origin in a meeting in Banff, Canada, held on 2–4 August 1991. This original Banff meeting in 1991 was part of the activities of the International Society of Nephrology Commission on Acute Renal Failure, which also included international disaster relief. The initiative was inspired by the then recent development of a consensus grading system for diagnosis of rejection in cardiac allografts led by Dr Margaret Billingham, a key participant at the first Banff meeting. Looking back now 21 years later at that meeting and the working classification developed there, it is clear that the classification, as it has evolved since then, has made a critical contribution to many advances in the field of transplantation.

The Banff Allograft Pathology consensus process ultimately not only led to revisions and expansion in the schema for kidney allograft pathology but also extended to development of classifications of allograft pathology in the liver, pancreas, and composite tissue grafts, as well as contributing to advances in the existing classifications of the heart and lung allograft pathology.

Counterbalancing the successes of the classification are its limitations—limitations that apply to any biopsy- and histopathology-based classification system. These include inherent potential for sampling error, suboptimal reproducibility when implemented globally, lack of widespread application of morphometry with true quantification, and lack of formal integration of molecular and genomics data into the classification. These are issues to be addressed in the ongoing international Banff conferences and related activities, described below.
potential categories of rejection that would be clinically relevant for guiding therapy.

The new classification used lesion scoring and guidelines to provide rigor in the evaluation of renal allograft pathology and the assignment of biopsies into diagnostic categories. Nonrejection pathology in the allograft was also considered and described in some detail, and this continues to be a focus in ongoing Banff meetings.

**THEMES OF THE BANFF MEETINGS**

A previous article on the history of the Banff Classification focused on the participants. Here, we will concentrate on the ideas espoused by the Banff meetings (Table 1) and the Banff consensus process. The flavor of the Banff meetings from the beginning was one of flexibility and openness to the ideas of others. At the original meeting of 20 people in 1991, the kidney transplant pathology classification that emerged was very different from any of the drafts that individuals had created before the consensus discussions began. It was a true creation of the meeting itself, based on international expertise, the existing literature, and facilitated discussions. The classification was designed as a dynamic working document that could be modified as the need for future changes was demonstrated.

At the second meeting in 1993, a similar effort was initiated to create a classification for liver transplant pathology, and this became a major focus of the Third Banff Conference held in July 1995. The third conference had four times as many attendees as the first conference and was conducted with an air of positive expectation that could not have been anticipated in 1991. The 1995 meeting also brought Banff lesion scoring into line with the CADI scoring system, so there was no difference between the two classifications.

By the time of the 1995 meeting, the original Banff Classification of Renal Allograft Pathology was in wide use, and there had been extensive studies showing its clinical validity and reproducibility. Extrapolating from these studies, it seemed likely that the classification had already resulted in improvement in patient care. The Banff classification had been endorsed by the FDA and other regulatory agencies, and had enabled the use of objective histological endpoints for international clinical trials of new antirejection agents and other scientific studies, a process that continues to this day.

**Table 1 | Banff Allograft Pathology Meetings since 1991 with key themes**

<table>
<thead>
<tr>
<th>Location</th>
<th>Length (days)</th>
<th>Links</th>
<th>Key subjects debated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991 Banff, Canada</td>
<td>1.5</td>
<td>ISN</td>
<td>Classification established, lesion scoring, diagnostic categories, physician-led consensus</td>
</tr>
<tr>
<td>1993 Banff, Canada</td>
<td>3</td>
<td>ISN, CAP</td>
<td>Liver classification, chronic rejection, first presentation on molecular pathology approaches</td>
</tr>
<tr>
<td>1995 Banff, Canada</td>
<td>4</td>
<td>ISN</td>
<td>Pancreas classification, glomerulitis, first international medical meeting on CD-ROM, first Banff conference with microscope sessions. Lesion scoring normalized with CADI.</td>
</tr>
<tr>
<td>1997 Banff, Canada</td>
<td>5</td>
<td>ISN</td>
<td>Merging of Banff and CCTT classifications, establishing basis for current Banff classification, post-transplant lymphoproliferative disorder, first Banff conference with posters</td>
</tr>
<tr>
<td>1999 Banff, Canada</td>
<td>5</td>
<td>ISN, NKF, NIH</td>
<td>Protocol biopsies, chronic rejection, and viral diseases, clinical practice guidelines. First conference supported by an NIH grant.</td>
</tr>
<tr>
<td>2001 Banff, Canada</td>
<td>5</td>
<td>ISN, NKF</td>
<td>AMR, donor biopsies, genomics, CAN, heart transplantation</td>
</tr>
<tr>
<td>2003 Aberdeen, Scotland</td>
<td>4</td>
<td>ISN, NKF</td>
<td>C4d, macrophages, tolerance, accommodation, immunodepletion</td>
</tr>
<tr>
<td>2005 Edmonton, Canada</td>
<td>6</td>
<td>NKF</td>
<td>Genomics and molecular markers, B cells, chronic allograft injury with elimination of CAN, establishment of criteria for chronic rejection</td>
</tr>
<tr>
<td>2007 La Coruna, Spain</td>
<td>6</td>
<td>UofA</td>
<td>Protocol biopsies, transcriptome, mechanisms of rejection, ptc grading, new total inflammation score; working groups for v-lesion, genomic integration, pancreas and composite tissue rejection schemas</td>
</tr>
<tr>
<td>2009 Banff, Canada</td>
<td>5</td>
<td>UofA</td>
<td>Viruses, quality assurance, AMR in kidney, heart, and pancreas, liver allograft accommodation, endothelial cells, surrogate markers. Working groups.</td>
</tr>
<tr>
<td>2011 Paris, France</td>
<td>5</td>
<td>UofA</td>
<td>Sensitized patient, C4d, isolated v-lesion, the future, genomics, glomerulitis, epithelial injury/epithelial mesenchymal transformation, operational tolerance monitoring in liver grafts</td>
</tr>
</tbody>
</table>

Abbreviations: AMR, antibody-mediated rejection; CAN, chronic allograft nephropathy; CAP, Canadian Association of Pathologists; and Future of Pathology/Laboratory Medicine in Canada Consortium; ISN, International Society of Nephrology; NKF, National Kidney Foundation (US); NIH, National Institutes of Health (US); ptc, peritubular capillary; UofA, University of Alberta.
Interestingly, in 1995, an energetic attempt was mounted to remove glomerulitis—the ‘g’ lesion (Figure 2)—from the classification, as the lesion appeared to have little clinical significance, based on the existing literature at that time. This was successfully resisted, and today glomerulitis is recognized as an important sign of microcirculatory alloinflammation and is an integral part of the classification. With the recent recognition of C4d-negative antibody–mediated rejection (AMR), the presence of ‘the g lesion’ is now sometimes the main evidence supporting that diagnosis along with peritubular capillaritis (Figure 3).13 The 1995 meeting also saw establishment of precepts guiding changes in the classification: ‘no changes should be made unless supported by well conducted studies from two different investigative groups’.14

At the 1997 meeting, the most useful and validated parts of the Banff Classification were merged together with the most useful parts of the CCTT classification,15 which had been developed as a simplified version of the Banff criteria, defining tubulointerstitial, vascular, and severe rejection with clinical correlations. These were merged to create a new classification: Banff ‘97.16 There have been subsequent modifications and additions, but the classification today is still built on the framework established at the 1997 meeting. A CD-ROM was created capturing video high points of the meeting.17

Although AMR was rarely diagnosed at the time of the inaugural Banff meeting, development of the classification, and the 1997 modified schema, by the mid-to-late 1990s, characteristic histological changes had been identified and AMR was being increasingly recognized in renal allografts. With the addition of immunohistological staining for C4d to the pathologists’ tool box,18 definite criteria enabling diagnosis of AMR could be established. On the basis of the growing literature and consensus discussions, an addition to the Banff classification outlining criteria for diagnosis of AMR was published.19 In the years since then, AMR in other solid organ allografts has been discussed at the Banff meetings, and criteria for diagnosis developed and published (e.g., in pancreas).20

As immunosuppressive strategies became more effective, acute rejection was no longer a common cause of graft loss, and investigators began to focus on late allograft changes and fibrosis in the allograft. The 1997 Banff working classification had guidelines for semiquantitative assessment of fibrosis in allografts and moved from labeling all chronic changes in the allograft ‘chronic rejection’ to a new concept, ‘chronic allograft nephropathy’. Chronic allograft nephropathy included a subgroup because of chronic rejection, but with recognition of the other numerous causes of fibrosis in the allograft. In the Banff meeting in 2001, there was extensive discussion of fibrosis and atrophy in the allograft, with discussion of a new paradigm.21 Chronic allograft nephropathy, while correlating with graft outcomes, in itself became a blanket term that was almost meaningless in its lack of specificity. This topic remained a major issue at subsequent Banff meetings, and by the Banff meeting in 2005 and 2007, precise definitions of lesions led to elimination of the term chronic allograft nephropathy22 to be replaced by specific diagnoses, including chronic rejection, both cell- and antibody-mediated.23 Only a relatively small subset of patients remained with interstitial fibrosis and tubular atrophy of undefined cause.

A major advance of the 2009 meeting was the creation of the Banff working groups, which aim to address issues in transplantation pathology. These include the significance of isolated intimal arteritis without tubulointerstitial inflammation (the ‘isolated v-lesion’), accurate scoring of fibrosis, detection and accurate scoring of acute and chronic glomerular lesions, incorporation of molecular pathology into diagnostic schemas, a possible new classification system for polyomavirus nephropathy, quality assurance in histology and immunohistology, and diagnosis of AMR in kidney independent of C4d. These international Banff working groups collect data from multiple centers in areas where Banff criteria are problematic, and validate the results, ultimately leading to refinement and improvement of the working classification, thus ensuring that the Banff process is evidence-based and continues to lead to improvements in patient care and management.24,25 At the most recent Banff
meeting in 2011, the criteria for diagnosis of AMR in the kidney were revisited with the recognition that one or more diagnostic criteria may be absent in cases that subsequently developed chronic microvascular injury. Banff working groups are currently studying this issue and developing recommendations for a change in the working classification.

LIMITATIONS OF THE CLASSIFICATION
From the beginning, the Banff classification has included semiquantitative lesion grading, which made it simple and quick to apply. However, some have emphasized the desirability of true quantitation (e.g., Howie). Reproducibility of lesion scoring, an issue in any histopathology scoring system, is another limiting factor but seems highly dependent on the group in which the reproducibility is tested. Initial studies on reproducibility in the early years of the classification were reasonably encouraging, with things looking less favorable as studies expanded more widely. It is fascinating that with fibrosis scoring, the same first author who demonstrated excellent reproducibility in fibrosis scoring in the native kidney demonstrated much poorer reproducibility in the transplanted kidney where participants were more widely scattered. Ultimately, reproducibility is influenced by true biological variability, by the exact ways in which lesions are scored, and by the experience of the pathologists performing the scoring.

Early on in deliberations at the Banff meetings, standardization and streamlining of morphometry techniques were a focus, so that they could be routinely applied to diagnostic material with quantitative results obtained in a timely manner. However, such strategies have never been developed for standardized application to biopsy assessment. This is an issue currently being addressed by the graft fibrosis Banff working group. Other problems include the persistence of the ‘borderline’ category, when inflammation is below threshold to confidently diagnose acute cellular- or AMR using current histopathological criteria. Diagnostic thresholds are set to avoid overtreatment with attendant toxicities but may reduce sensitivity. Integration of molecular and genomics data into the classification remains an ideal and a working group focus, although implementation of this throughout the world may be problematic, especially in under-resourced areas.

Whether one understands the failure to disseminate and standardize morphometry, genomics, and molecular studies as a failure of medical politics and vision, or as simple economics, it is clear that it may have retarded progress to an ideal diagnostic paradigm. The success of recent morphometry, genomics, and molecular studies in experimental models and in the native kidney may ultimately lead to standardization and implementation in individual transplant centers worldwide, and these strategies remain a major focus of the ongoing Banff meetings.

A recent article related to the Banff classification suggests that tubulitis is severely underdetected by conventional microscopy. In this study, in which lymphocytes in the tubular epithelium, which define tubulitis, were identified by immunostaining for CD3, the authors conclude that tubulitis is missed very frequently, but the Banff classification seems to be ‘calibrated’ to allow for this and it does not seriously affect the identification of clinically significant acute rejection. Immunostaining is therefore not indicated in routine practice because (by Banff criteria) it would result in overdiagnosis of rejection. Intimal arteritis can indicate acute rejection even if extremely mild and the ‘isolated v-lesion’ can be clinically meaningful. The importance of isolated intimal arteritis/‘isolated v-lesion’ in the absence of tubulointerstitial inflammation is an ongoing focus of a multicenter study organized by a Banff working group, and will continue to be debated. One recent presentation suggests that patients with this lesion may actually have a worse prognosis than when the arteritis is associated with tubulointerstitial inflammation.

Two recent articles relating molecular changes to Banff lesion scoring directly show the potential in the long run for molecular studies to greatly improve the classification.

THE BANFF CONSENSUS PROCESS—WHAT HAS BEEN ACHIEVED?
The Banff meetings are designed for optimal interaction, with many small group sessions and working group meetings, and attendees feel very much a part of the intellectual process. It is also an excellent forum for junior and senior participants to mingle, and for organization of collaborative studies around issues discussed at the Banff meetings. Meetings are open to all, and poster sessions that frequently feature the work of young investigators are an important component of the meetings. Following consensus discussions and literature review, numerous formal multiauthored publications have resulted from the Banff meetings. The Banff working groups were established in 2009. The work of many of these groups has been presented at meetings in 2012, with full-length papers expected in 2013. It appears that the Banff collaborative process will be productive well into the future.

As shown in Figure 1, there are now 717 articles in the transplantation literature that reference the Banff process. Activities of and interest in the Banff Allograft Pathology consensus process continue to grow. These 717 articles are an underestimation of the impact of the Banff Allograft Pathology consensus process, as they include only those articles where Banff was specifically listed as a key word or search term. Of the 717 articles, most articles are in journals with impact factor of 3.2 or higher, and one-third have impact factors of 6–9. The Banff meeting reports and main meeting papers have been cited over 4300 times in the medical literature. The 1999 paper “The Banff 97 Working Classification of Renal Allograft Pathology”, an update of the original classification, is a citation classic in the field having been cited 1764 times. This paper reflects a basic philosophy of the Banff process—that this and other Banff schemas are working classifications intended to be continuously refined and updated as the field of solid organ transplantation evolves (e.g., Demetris et al.).
The strong interest in global consensus generation and electronic communications that began with the Banff process led to the establishment of the cyberNephrology and cyberMedicine initiatives and eventually to the creation of the Technology and Future of Medicine course at the University of Alberta. It can be difficult to focus on the future in the field of medicine when the problems of the present already seem so challenging. It is heartening to see presentations about the Banff classification begin to be more future oriented as seen in this slide from a May 2012 presentation by Thangamani Muthukumar (Figure 4). We are fortunate that the next (2013) Banff meeting will be in Brazil where the application of strategic foresight to science is official government policy. Updated information about future meetings planned in Canada, Turkey, and Spain in 2015, 2017, and 2019, respectively, is available on line.

DISCLOSURE
All the authors declared no competing interests.

REFERENCES