Reproducibility of histopathologic findings in experimental pathology of the mouse; a sorry tail

Jerrold M. Ward, 1 Paul N. Schofield, 2 and John P. Sundberg 3

1 Global VetPathology, Montgomery Village, MD, USA.
2 Dept of Physiology Development and Neuroscience, University of Cambridge, Downing Street, Cambridge, CB2 3EG, UK.
3 The Jackson Laboratory, Bar Harbor, ME, USA

Correspondence should be addressed to J.M.W. (veterinarpethology@gmail.com)

Reproducibility of in vivo research using the mouse as a model organism is dependent on many factors including experimental design, strain or stock, experimental protocols, and methods of data evaluation. Gross and histopathology are often the endpoints of such studies and there is increasing concern about the accuracy and reproducibility of diagnoses in the literature. In order to reproduce histopathologic results, the pathology protocol, including necropsy methods and slide preparation, should be followed by interpretation of the slides by a pathologist familiar with reading mouse slides and familiar with the consensus medical nomenclature used in mouse pathology. Likewise, it is important that pathologists are consulted as reviewers of manuscripts where histopathology is a key part of the investigation. The absence of pathology expertise in planning, executing, and reviewing in vivo research using mice leads to questionable pathology-based findings and conclusions from studies, even in high impact journals. We discuss the various aspects of this growing problem, give some examples from the literature, and suggest solutions.

Histopathological descriptions of the frequency and nature of lesions and disease entities are very often the endpoints in biomedical research conducted in model organisms such as the mouse. In contrast to clinical pathology where endpoints are usually assessed using biochemical and molecular assays, histopathological assessment, whilst using molecular markers and imaging as adjunct qualitative and quantitative techniques, is highly dependent on the individual expertise of trained expert pathologists, not only in recognizing lesions but also in
their knowledge of the background diseases of the mice and understanding of the meaning of the pattern of disease in the whole mouse\textsuperscript{1-4}. Reproducibility of histopathological endpoints therefore depends on the implementation of a common standardized vocabulary, competent work-up, and an in-depth knowledge of the mouse strains under investigation so that, for example, background lesions are not mistaken for those which are experimentally induced. Such knowledge is critical in the design of experiments, as well as in understanding the impact of husbandry, the microbiome, and diet on the interpretation of results\textsuperscript{2-5}.

In recent years increasing concern has been expressed by funding agencies and scientific communities alike about the lack of reproducibility of experiments in the biomedical domain. Attention was initially drawn to this by pharmaceutical companies which rely on preclinical, precompetitive research for drug development pipelines\textsuperscript{6-8}. Identification of this problem has been followed by an outpouring of concern from funding agencies such as the U.S. National Institutes of Health\textsuperscript{9-11} and to an extent journals and professional bodies\textsuperscript{12-17}.

While much attention has been paid to the reproducibility of molecular assays, \textit{in vitro} (cell culture) assays and the inappropriate application of statistical methods (as well as simple errors), only recently have the issues surrounding reproducibility in animal experimentation been discussed in depth\textsuperscript{13}. Much of these discussions have concerned husbandry and the effect of diet and microbiome on experimental outcomes\textsuperscript{18-20}, particularly in neuroscience\textsuperscript{14}, but recent papers have addressed the problem of what a sound histopathological investigation should look like, how to use knowledge of pathology in experimental design, based on the ARRIVE and related guidelines, and more recently the confounding impact of the environment and the gut and skin microbiomes\textsuperscript{5}.

In this paper we address some of the issues impacting on the reproducibility of histopathological findings: (1) lack of pathology expertise, in author lists and in peer-review, (2) poor standards of reporting—illustrated with common errors seen in papers—, and inconsistent pathology nomenclature; and (3) availability of primary data, without which it is impossible to assess a paper without attempting to undertake a compete experimental replication\textsuperscript{21-23}. Most importantly, we emphasize that if pathologists are not involved in designing mouse experiments and interpreting lesions, the accuracy of the diagnoses reported and conclusions drawn may be questionable.

**The importance of pathologists**

Investigators often do not have enough funds to pay for research pathology services and/or believe that they can perform histopathology interpretations themselves. Pathology is a medical specialty that requires years of training, experience, and board certification as a minimum. Although pathology has many sub-disciplines, such as mouse pathology, a general pathologist is much more expert than a non-pathologist, and is often sufficient to provide substantial benefit to an animal research study\textsuperscript{24}. Lack of pathology expertise by investigators
leads to inaccurate histopathological descriptions of lesions and often missed or spurious reporting of pathologic findings in publications. Indications may be noticed in the figure legends which often do not describe the lesions displayed, or in some rare cases images are replicated in various orientations for different lesions or mice[^25]. In some cases, a pathologist was not involved in late or final edits or did not review the galley proofs of an accepted manuscript[^26]. In addition to accurate interpretation of data, pathologists are important to ensure proper nomenclature is used when reporting on results. The use of generally accepted pathology nomenclature for unexpected and novel findings leads to publications that can be interpreted by readers, including other pathologists. Rodent pathology terminology often mirrors that used for humans but differences do occur. Pathology of genetically engineered mice often requires interpretation of novel findings since each mouse may have unique lesions not previously reported, especially where the study is the first for a novel gene knockout or treatment. A classic example is the relatively common lesion in mouse hearts that pathologists diagnose as epicardial and myocardial mineralization and fibrosis but non-pathologists often call “dystrophic cardiac calcinosis” or a variety of other names[^27-31]. Investigators without pathology backgrounds often over-interpret their research findings, the temptation being to fit results to their hypotheses. Over-diagnosis of lesions as malignant when they may be, in fact, benign, hyperplastic, or even normal is a common problem. This latter point emphasizes the value of knowing anatomical differences between the species.

**Some examples of errors seen in reported histopathological diagnoses**

Besides incomplete reporting of the experimental design, including the pathology protocol, there are common questionable diagnoses that can be found (Table 1)^[^32-34]. Often, the figure legends do not describe what is illustrated by the figure, normal tissues are misidentified as lesions, non-neoplastic lesions are reported as cancer, or benign lesions are diagnosed as malignant neoplasms. In addition, inflammatory lesions may be described as neoplasms or “tumors”, benign or malignant.

Our interpretations of histopathology figures in published reports, which we will give as examples, are based solely on our interpretation of what was present in the published figures and not based on microscopic slide review. Slide review may reveal different findings than what is in the published figures. Often the published histopathology figures are small and when enlarged in the pdf file, they often lose resolution to the point of being uninterpretable. One of many approaches to this problem is to post additional digital images at a variety of magnifications or whole slide images as supplemental data and/or in other public websites such the Mouse Tumor Biology Database (http://tumor.informatics.jax.org/mtbwi/index.do[^35, 36] , Gene Expression Database (http://www.informatics.jax.org/expression.shtml[^37] , Pathbase (http://pathbase.net/)[^38, 39] , and many others[^5].

In order to evaluate any organs, such as skin, clear understanding of the normal anatomy is absolutely necessary in order to recognize any type of change, be it disease or just subtle changes in normal physiology. For the skin, a lack of knowledge of the normal hair cycle is a
commonly reported source of misinterpretation. All hair follicles regularly go through anagen, the normal growth phase; catagen, the transition stage to telogen, the long term resting stage; to exogen, when the old hair shaft is lost. This process then starts over again and is repeated throughout life. The cycle varies by hair type (vibrissae cycle is different compared to body hair) and species (mice cycle in waves while humans cycle in a mosaic pattern)\(^1\). The hypodermal fat layer in the skin changes thickness through the cycle. When thinnest, in telogen\(^40\), this is often reported as an abnormal phenotype. Sebaceous gland size also changes through the hair cycle making estimation of the size of this gland an unpredictable feature, and also commonly misinterpreted\(^41\). Changes in numbers of hair follicles can be misinterpreted owing to artifacts of section orientation (Fig. 2)\(^42\).

Male mice have modified sebaceous glands (with a large excretory duct along the penis) known as preputial glands or clitoral glands in the inguinal area of females (Fig. 3). These tissues have been diagnosed as teratomas or skin tumors in some publications\(^33,34,35\), and an erratum published for one of the publications\(^33\). Mouse accessory sex glands include various prostate lobes, seminal vesicles, and other structures, the architecture of which differs from that of humans. Tissue artifacts have been diagnosed as early stage prostate cancer\(^45\) (Fig. 4).

Immunohistochemistry findings are commonly a problem in publications. Authors often report positive labeling (often with a brown chromogen) of cells and tissues which appear to represent nonspecific background staining, due to omission of proper controls\(^46\). A good example was reported in mouse prostate epithelial cells and connective tissue\(^47\) using an anti-human antibody that never was reported (even by the company selling the antibody) to work in mouse tissues. Mast cells are often nonspecifically positive in mouse tissues using peroxidase based reagents\(^48\).

**When is a neoplasm not a neoplasm?**

Neoplasms and their preneoplastic/precancerous lesions are commonly found in mouse experiments involving chemical carcinogens and/or in genetically engineered mice. The majority of papers on mice with these induced or spontaneous lesions do not refer to the publications that focus on standardized nomenclature for the organ or disease under investigation such as those noted above and in our reference list. While many journals list in their “instructions for authors” that they require authors to use standardized nomenclature, this requirement is often not enforced by editors. This policy holds not only for diagnostic terms but also for mouse strain and allele designations, as mouse genetic nomenclature is very uniformly standardized\(^49,50\). The questionable diagnosis of preneoplastic and neoplastic lesions of mice is commonly found in publications; some examples are given below.

The most widely used prostate cancer mouse model (commonly called TRAMP) is an example of misuse of standardized nomenclature. A search of Mouse Genome Informatics (http://www.informatics.jax.org/genes.shtml) yielded 4 matches (Table 2), only one of which is the transgenic line used for prostate cancer research (Tg(TRAMP08247Ng))\(^51\). With over 600 publications, this transgenic line is often considered the best mouse model of human prostate
cancer. The Tg(TRAMP)08247Ng line was noted to have a high incidence of prostate adenocarcinoma metastases suggesting that this was a good model for human prostate cancer. However, these were later shown to be of neuroendocrine origin and not of the typical epithelial origin most common in humans (Fig. 5)\textsuperscript{51-53}. Phyllodes prostate carcinomas, initially diagnosed in the first Tg(TRAMP)08247Ng publication, were later shown to be adenomas or benign epithelial-stromal tumors of the seminal vesicles (Fig. 6)\textsuperscript{51, 54}. Metastatic prostate carcinoma to bone marrow was described in a new genetically engineered mouse model but a pathology nomenclature consensus committee determined that the cases were merely direct invasion from a large prostate mass\textsuperscript{51, 55}.

Inflammatory lesions caused by bacteria have sometimes been reported as tumors (neoplasms)\textsuperscript{56} as have been other types of inflammation\textsuperscript{57}. This may be technically correct, as some textbooks define “tumor” literally as any type of swelling and one of the five cardinal signs of inflammation, but inflammation should and can be easily differentiated from neoplasia. Lymphomas and leukemias are often difficult to diagnose accurately. Tail tumors in transgenic mice were reported to be large granular lymphocytic (LGL) leukemia\textsuperscript{58} but other investigators working with the same mice found sarcomas of various types including those arising in tendons and nerves in the tail (Figs. 7-9). These tail tumors may have developed accompanying inflammatory responses which included LGLs. Large spleens have been diagnosed as myeloproliferative disorders and leukemias, especially in mice with ulcerative skin lesions which cause reactive myeloid hyperplasia in the spleen of mice\textsuperscript{59}. Using a Helicobacter felis mouse gastric model, a research group developed a model of chronic gastritis that eventually was reported to develop gastric lymphomas. These changes appeared histologically unconvincing as the lesions described in the initial publication. However, in this case, a subsequent publication did provide molecular proof that these were indeed lymphomas\textsuperscript{60}.

**How to increase reproducibility by improving nomenclature usage**

Problems in pathology evaluation may occur at various stages of the study leading to questionable pathology interpretation in the manuscript submitted or published. In order to increase reproducibility in mouse studies, a trained pathologist knowledgeable in rodent pathology nomenclature should be involved, either in study design, manuscript writing, or chosen by editors during the peer-review process.

Pathology nomenclature in the paper should follow general guidelines for mouse pathology as published by international committees and experts, as discussed below.

The ability of a pathologist to accurately diagnose lesions in laboratory animals, especially rodents, depends on training and experience. Experience includes the use of widely acceptable veterinary pathology and species specific nomenclature often provided through publications by expert groups of pathologists and by international or national committees\textsuperscript{33, 61-65} and in books with multiple authors\textsuperscript{61, 66-69}. Also others have proposed formal ontologies for data capture and analysis\textsuperscript{39, 70, 71}, which are also based on international nomenclatures and informatics standards.
There are numerous publications on neoplastic diseases in mice, especially in genetically engineered mice (GEM). The NCI Mouse Models of Human Cancer Consortium over the past 15 years has established pathology committees to develop nomenclatures for several important organs\textsuperscript{33, 51, 61}. The publications on neoplastic diseases in mammary gland, prostate, lung, intestine, brain, skin and pancreas provide important guidelines for investigators (http://emice.nci.nih.gov/camod-models-by-organ-site). The INHAND pathology nomenclatures have similarly created detailed terminological recommendations for both proliferative and non-proliferative lesions under the auspices of the committees established by a consortium of societies of toxicologic pathology, which can be found on https://www.toxpath.org/inhand.asp. These, together with detailed publications on individual classes of lesion make up a significant terminological corpus and pathologists making diagnoses should be familiar with these.

**Conclusion**

In comparison to factors confounding reproducibility of mouse biological experiments originating in experimental design, husbandry, and microbiome, the problem of reliable histopathological interpretation of experimental animals is perhaps one of the most tractable sources of error. Enrolment of an experienced pathologist onto a study early in its inception and planning stages, then its subsequent analysis, is clearly highly desirable, and many of the problems we discuss above are unlikely to arise under the guidance of appropriately trained personnel. The issue of finding experienced mouse pathologists is one which has been discussed at length elsewhere\textsuperscript{2, 24, 34}, though the authors feel that the mouse pathology community is sufficiently interactive that good advice can easily be sought out by motivated investigators.

Changes in priorities at journals and funding agencies are also needed to significantly improve the reliability of pathology in mouse model studies. Availability of the primary images on which experimental conclusions are based should be mandatory at journals and in line with the FAIR guidelines\textsuperscript{5, 72}. Similarly, funding agencies need to pay more attention to the intended use of histopathology in grant applications and insist on provision of appropriate expertise with an appropriate budget. Increased stringency surrounding the processes of funding and publication might represent more effort for researchers, reviewers, and journal editors, but will reduce the instances of flawed histopathology we see in many journals today. The pragmatic answer has to be the education of investigators in the critical assessment of histopathology data, but in the final analysis, as ever; *caveat emptor.*
Figure Legends

FIGURE 1 Numbers of papers published between 1948 and 2015 indexed in Pubmed as (red squares) "Mouse histopathology", and (blue diamonds), "Mouse model." (data accessed 12.9.16)

FIGURE 2 Skin of mouse in anagen with abundant hair follicles. This normal stage of the cell cycle has been reported to be hyperplasia. H&E, X10.

FIGURE 3 Normal mouse preputial gland showing glandular tissue with central ducts. Two publications reported normal glands as teratomas or carcinomas. H&E, X4.

FIGURE 4 Normal prostate of young mouse with artefactual folds of the acinar epithelium which were misdiagnosed as early stage prostate cancer in a publication. H&E, X40.

FIGURE 5 Neuroendocrine carcinoma in the prostate of a TRAMP (Tg(TRAMP)8247Ng) mouse. These mice were reported to develop highly metastatic prostate adenocarcinoma. H&E, X40.

FIGURE 6 Benign (epithelial-stromal) tumor in the seminal vesicles of a TRAMP mouse. Note tumor growth into the lumen and no invasion. These lesions were reported as phylloides prostate carcinomas. H&E, X4.

FIGURE 7 Tail of a HTLV-I tax transgenic mouse with early tumors (on left and right side) of tendon origin. This mouse was reported to develop leukemia and not tendon tumors. H&E, X4.

FIGURE 8 Tail of a HTLV-I tax transgenic mouse with an early tumor of tail tendon origin (darker tumor in the lower portion of the figure beneath the skin). This mouse was reported to develop leukemia. H&E, X4.

FIGURE 9 Early tail tendon sarcoma showing tumor infiltration by myeloperoxidase positive neutrophils. Immunoperoxidase, X40.
TABLE 1 Evidence of Questionable Pathology Interpretation in Publications

Figure legends do not accurately reflect what is in the figure
Figure legends do not describe anything in the figure
Lack of complete or appropriate necropsies and histopathology
Misidentification of normal organs and tissues as lesions
Diagnoses of non-neoplastic lesions as neoplasms
Diagnoses of tumors with unconventional terminology
Reporting of benign lesions as malignant
Reporting of inflammatory lesions as tumors
Reporting of novel lesions incorrectly
Use of incorrect (accepted) terminology/diagnoses

TABLE 2. Partial table from Mouse Genome Informatics (Accessed 19 Sept. 2016) to illustrate three genes and one transgene identified by the term TRAMP, three of which are unrelated genes. Source: http://www.informatics.jax.org/genes.shtml

<table>
<thead>
<tr>
<th>Genetic Location</th>
<th>Symbol</th>
<th>Why Matched</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr4 82.89 cM</td>
<td>Tnfrsf25, tumor necrosis factor receptor superfamily, member 25</td>
<td>synonym: TRAMP</td>
</tr>
<tr>
<td>Chr1 4.18 cM</td>
<td>Tram1, translocating chain-associating membrane protein 1</td>
<td>synonym: TRAMP</td>
</tr>
<tr>
<td>Chr Unknown</td>
<td>Tg(TRAMP)8247Ng, transgene insertion 8247, Norman M Greenberg</td>
<td>currentSymbol: Tg(TRAMP)8247Ng</td>
</tr>
<tr>
<td>Chr1 72.12 cM</td>
<td>Dpt, dermatopontin</td>
<td>humanSynonym: TRAMP</td>
</tr>
</tbody>
</table>

ACKNOWLEDGEMENTS

Supported, in part, by NIH Grants R01 AR049288 and R21 AR063781 (to JPS) and by The Warden and Fellows of Robinson College, Cambridge (to PNS).

COMPETING FINANCIAL INTERESTS

Dr. Sundberg has a research contract with BIOCON LLC for preclinical trials not involving any part of this article. The other authors declare no competing financial interests.
References
