Barcode Label for E-poster

Deep Learning-Based Image Analysis Model for Classification and Quantification of Testicular and Epididymal Lesions in Rats

Taishi Shimazaki[1], Kyotaka Muta[1], Yuzo Yasui[1], Rohit Garg[2], Pranab Samanta[2], Amogh Mohanty[2], Tijo Thomas[2] and Toshiyuki Shoda[1]

1: Toxicology Research Laboratories, Yokohama Research Center, Central Pharmaceutical Research Institute, Japan Tobacco Inc. 2: AIRA Matrix Private Limited

COI Disclosure Information :
We declare no conflicts of interest associated with this poster.
The content of this poster is currently being submitted for publication.

Introduction

Spermatogenic staging and assessment of testicular toxicities in rat tissue sections are time-consuming and requires the expertise of well-trained pathologists.
Deep Learning (DL)-based image analysis is increasingly being used for preclinical safety-assessment studies in the pharmaceutical industry.
Here we present a DL-based solution for classifying 11 stage groups of spermatogenesis namely stages I, II-III, IV-VI, VII, VIII, IX, X, XI, XII-XIII, XIV and “Stage Not Analyzed” based on normal testicular tissue structure.
In addition, the solution for identifying and quantifying testicular and epididymal lesions in rat toxicology studies is also shown.

Materials & Methods

■ H&E stained sections of testis from young Sprague-Dawley (SD) rats (males, 8-week-old) treated/non-treated with various compounds in 2-week toxicity studies were digitised at 40x magnification. The training dataset for model development consisted of 512x512 sized image tiles extracted at 10X and 40X magnification from different WSIs of testis and epididymis tissue sections from both normal (non-treated) and treated animals. Multiple variants of U-Net based convolutional neural networks were trained on this dataset for semantic segmentation of tubules, germ cells and various lesions. The details of the parameters, network architectures and training datasets used are given in Table 1. A network architecture for the HistNet [1] model is shown in Figure 1.

Table 1: Parameters and Training Datasets

Parameters	Training Dataset
➢ **Testis**	
Tubule and Lumen	30 WSIs (2000 tiles of size 512x512 dimensions at 10x magnification)
Germ Cells for Staging (Round Spermatid, Elongated Spermatid, Spermatogonia, Spermatocyte, Meiotic Body, Residual body)	30 WSIs (1200 tiles at 40x magnification)
Degeneration/Necrosis of Germ Cell	10 WSIs (750 tiles at 40x magnification)
Tubular Atrophy	25 WSIs (1500 tiles at 10x magnification)
Tubular Dilatation	25 WSI (1250 tiles at 10x magnification)
Vacuolation of Sertoli Cell	10 WSIs (800 tiles at 40x magnification)
Multinucleated Giant Cell	10 WSIs (600 tiles at 20x magnification)

Parameters	Training Dataset
➢ **Epididymis**	
Tubule and Lumen	20 WSIs (1200 tiles of size 512x512 dimensions at 10x magnification)
Cell Debris	20 WSIs (1300 tiles of size 512x512 at 40x magnification)
Sperm	20 WSIs (850 tiles of size 512x512 at 40x magnification)
Head and Tail Classification	390 WSIs (4 sections of Head/Tail per slide)

■ Spermatogenic Staging
- First, individual tubules, lumens and various germ cells were segmented out as mentioned in Table 1. After segmentation of the germ cells, relevant features for staging based on the presence, count and location of these cells were extracted to correspond to each tubule [5]. Then, a decision-tree based classifier was used to classify these features for each tubule into stage groups. A stage group frequency map was then generated, showing the percentage of tubules belonging to each stage group.
- The performance of spermatogenic staging on WSIs from non-treated animals was validated by JSTP-certified pathologists (Table 3). A comparative analysis of spermatogenic stage groups in non-treated, very slight, slight and moderate groups of testicular toxicity is shown in Figure 8.

■ Testicular Lesion Detection
- DL-based models were trained for the detection of various lesions in the testis namely “Degeneration/Necrosis of germ cell”, “Tubular atrophy”, ”Tubular dilatation”, “Vacuolation of Sertoli cell”, and “Multinucleated giant cell”. Then the number of abnormal tubules and the number of foci corresponding to each lesion were calculated for each testis section (Figure 9).

■ Epididymal Lesion Detection
- In order to assess the early changes associated with testicular toxicity, epididymis sections were also assessed. The models for identifying sperm and cell debris were developed. The absolute count of sperm (and cell debris) and the percentage area occupied by sperm (and cell debris) with respect to lumen area were calculated for each tubule (Figure 9). Head and tail sections of the epididymis were also identified to assess the objective findings for each area.

■ Validation
- The performance of this model was verified using WSIs of 6 toxicity studies (in total 119 WSIs) (Table 2) that were not used in the learning phase. In the 6 studies, 6-week-old male SD rats were orally administered certain testicular toxic compounds (A to F) for 2 weeks, and 119 WSIs were prepared as well as the WSIs used in the learning phase.
- First, histopathological data (“no findings (−)” or “findings (+)”) diagnosed by the pathologists were concatenated with the quantitative values obtained from the algorithm for each WSI.
- The most reliable thresholds were calculated for each finding based on a receiver operating characteristic (ROC) curve using Python scikit-learn (ver.1.0.2) module. The best threshold value was calculated by maximizing Youden’s index in the ROC curve. The discriminative performance was evaluated based on the area under the ROC curve (AUC-ROC). Based on the threshold value from the ROC curve, binary diagnostic results by the pathologists were classified into four classes: true positive, false positive, false negative, or true negative for each finding. The statistical parameters, including the F1-score, were calculated (Table 4).
- For the 119 WSIs, the quantitative values of each finding in the testis or epididymis generated from the model were compared with the histological diagnostic grade diagnosed by the pathologists (Figure 8).
- Quantitative values for 3 major parameters (Abnormal tubules in the testis, Sperm and Cell Debris in the epidymis) to evaluate testicular toxicity were generated for each dose group (vehicle, low, mid and high dose levels) in each study. Testicular toxicity was then diagnosed by the pathologists and the quantitative values were compared for each dosing group (Figure 9).

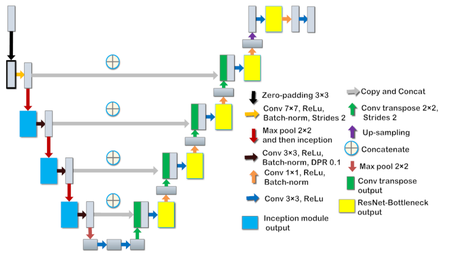


Figure 1: HistNet Model Architecture Used for the Germ Cell Segmentation Tasks

Table 2: Study design for validation

Study No.	Compound	Dose	Animal No./group	Total No. of Animals (WSIs)
No. 1	**A**	**Vehicle/Low/Mid/High**	**N= 5/group**	**20**
No. 2	**B**	**Vehicle/Low/Mid/High**	**N= 5/group (Vehicle)**	**35**
No. 3	**C**	**Vehicle/Low/Mid/High**	**N= 5/group**	**20**
No. 4	**D**	**Vehicle/Low/Mid/High**	**N= 5/group**	**20**
No. 5	**E**	**Vehicle/Low/High**	**N= 5/group**	**15**
No. 6	**F**	**Vehicle/Low/High**	**N= 3/group**	**9**
				Total 119

Discussion & Conclusions

✓ This model had a high performance in classifying and quantifying the 11 stage groups of spermatogenesis in H&E slides accurately and faster than manual staging in accordance with the pathologists’ results.
✓ The solution provided an automated, objective and accurate method for toxicological assessment in the rat testis and epididymis. Moreover, the model was able to accurately detect and quantify even the very slight changes observed in the testis sections. Results correlated well with the pathologist’s grading, with good F1 scores.

➢ **This study suggests that this deep learning-based model is able to classify the stage groups of spermatogenesis and detect various findings simultaneously on a WSI scanned at 40X of rat testis/epididymis with high statistical performance even though the findings are very slight. The model could be a useful and supportive tool for histopathological evaluation, especially for primary testicular screening in early toxicity studies in rats.**

Results

Results of Normal (Non-treated) Testis and Epididymis

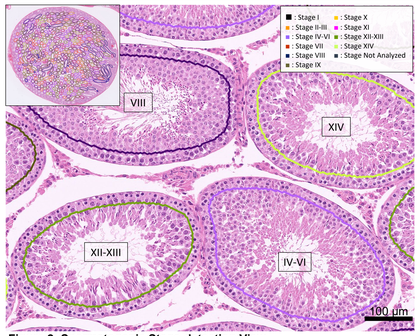


Figure 2: Spermatogenic Stage-detection View
The model was able to recognize spermatogenic stage groups on a WSI (H&E stained)

Table 3: Stage-Wise Performance Metrics of the Algorithm

Stage Group	Recall	Precision	F1 score *	No. of tubules
I	**0.98**	**0.98**	**0.98**	**766**
II-III	**0.96**	**0.93**	**0.95**	**363**
IV-VI	**0.98**	**0.99**	**0.98**	**984**
VII	**0.99**	**0.98**	**0.99**	**968**
VIII	**0.97**	**0.97**	**0.97**	**531**
IX	**0.97**	**0.97**	**0.97**	**211**
X	**0.99**	**0.99**	**0.99**	**130**
XI	**0.98**	**0.98**	**0.98**	**215**
XII-XIII	**0.99**	**0.98**	**0.99**	**797**
XIV	**0.99**	**0.99**	**0.99**	**288**
Accuracy	**0.98**			

The performance by the model in detecting spermatogenic stage groups on H&E-stained WSI was in accordance with the expert pathologist’s results on H&E and PAS-stained sections (Precision, Recall and F1 score).
*: A high value of F1 score indicates a high value for both Recall and Precision.

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{Recall} = \frac{TP}{TP+FN} \quad \text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

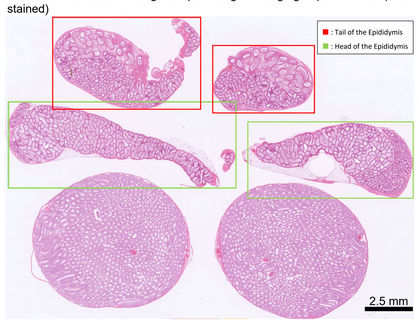


Figure 3: Identification of the Head and Tail Parts of the Epididymis
The model was able to recognize head (green) and tail (red) parts of the epididymis at low magnification.

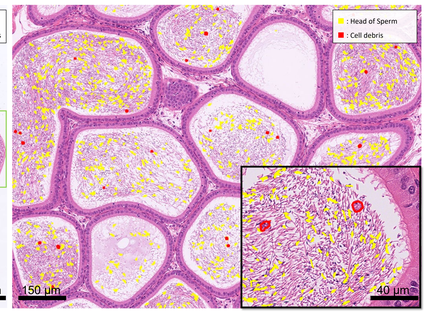
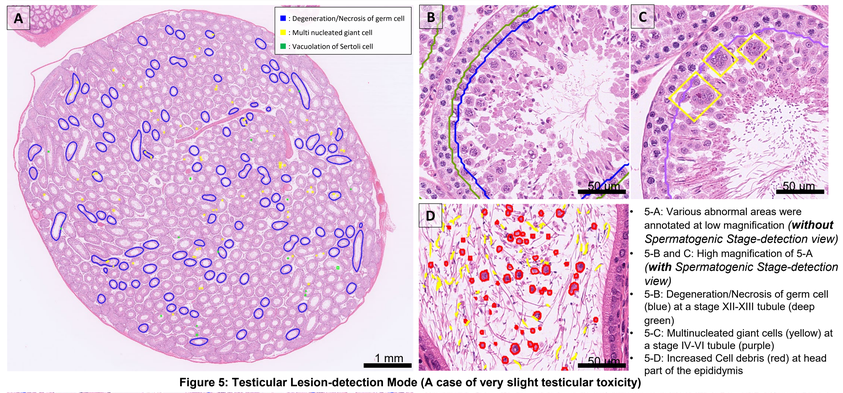


Figure 4: Recognition of Sperm and Cell Debris in the Epididymis
The model was able to recognize and quantify the head of the sperm (yellow) and cell debris (red) in the epididymal duct of the head part at high magnification.

Results of Drug-Induced Lesions of Testis and Epididymis



- 5-A: Various abnormal areas were annotated at low magnification (***without** Spermatogenic Stage-detection view)*
- 5-B and C: High magnification of 5-A (***with** Spermatogenic Stage-detection view)*
- 5-B: Degeneration/Necrosis of germ cell (blue) at a stage XII-XIII tubule (deep green)
- 5-C: Multinucleated giant cells (yellow) at a stage IV-VI tubule (purple)
- 5-D: Increased Cell debris (red) at head part of the epididymis

Figure 5: Testicular Lesion-detection Mode (A case of very slight testicular toxicity)

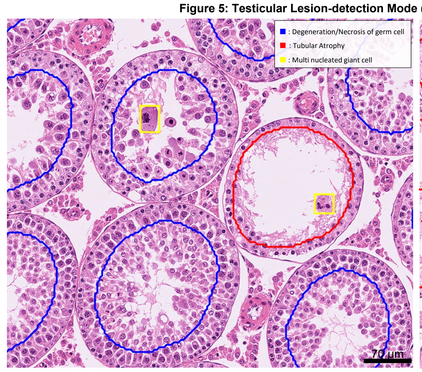


Figure 6: Detection of Drug-induced Lesions (A case of slight testicular toxicity)
The model was able to clearly recognize and quantify Degeneration/ Necrosis of germ cell (blue), Tubular Atrophy (red) and Multinucleated giant cell (yellow) in the testis. ***(without** Spermatogenic Stage-detection view)*

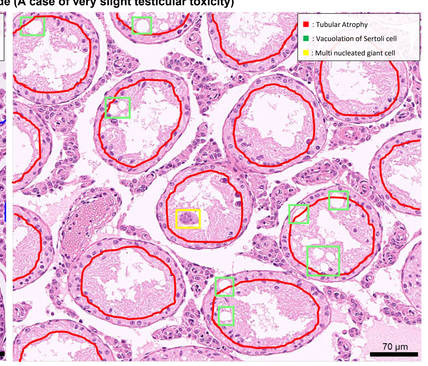


Figure 7: Detection of Drug-induced Lesions (A case of moderate testicular toxicity)
The model was able to recognize and quantify Tubular Atrophy (red), Vacuolation of Sertoli cell (green) and Multinucleated giant cell (yellow) in the testis. ***(without** Spermatogenic Stage-detection view)*

Table 4: Setting thresholds and evaluating statistical parameters by ROC analysis

	AUC	Cut off value	No. of samples	Recall	Specificity	Precision	Balanced accuracy	F1 score
Abnormal Tubules	**0.93**	**9.60**	**119**	**0.89**	**1.00**	**1.00**	**0.94**	**0.94**
Degeneration/Necrosis of Germ cell	**0.93**	**9.60**	**119**	**0.83**	**1.00**	**1.00**	**0.91**	**0.91**
Tubular Atrophy	**1.00**	**0.20**	**119**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Tubular Dilatation	**1.00**	**2.80**	**119**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Vacuolation of Sertoli cell	**1.00**	**6.10**	**119**	**1.00**	**0.99**	**0.97**	**0.99**	**0.98**
Multinucleated Giant Cell	**1.00**	**1.00**	**119**	**1.00**	**0.98**	**0.91**	**0.99**	**0.95**
Sperm	**0.99**	**3.62**	**119**	**0.98**	**0.89**	**0.85**	**0.94**	**0.91**
Cell Debris	**0.99**	**0.93**	**119**	**0.91**	**1.00**	**1.00**	**0.95**	**0.95**

The AUCs and various statistical parameters based on ROC analysis showed high scores, suggesting that the performance of determining the presence or absence of findings by the thresholds was very high.

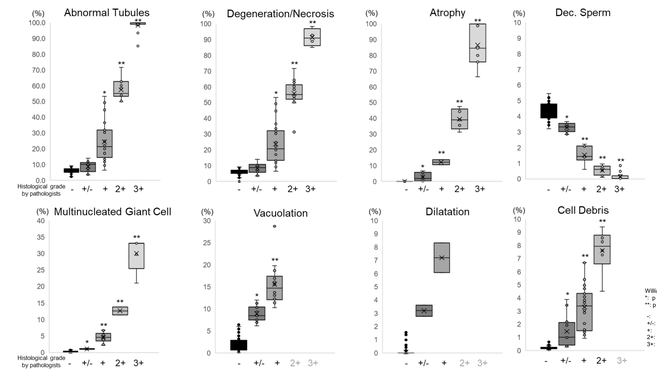


Figure 8: Comparison of quantitative values of the detected findings in the testis or epididymis with the histological diagnostic grade by pathologists
Vertical axis shows % of seminiferous tubules with each finding in the testis or % of sperm or debris in the epididymal duct lumen of the epididymis. The horizontal axis indicates the grade (-, +/-, +, 2+ or 3+) of each finding as diagnosed by pathologists. The grades marked in gray (ex. Vacuolation and dilatation) are groups with no positive specimens.

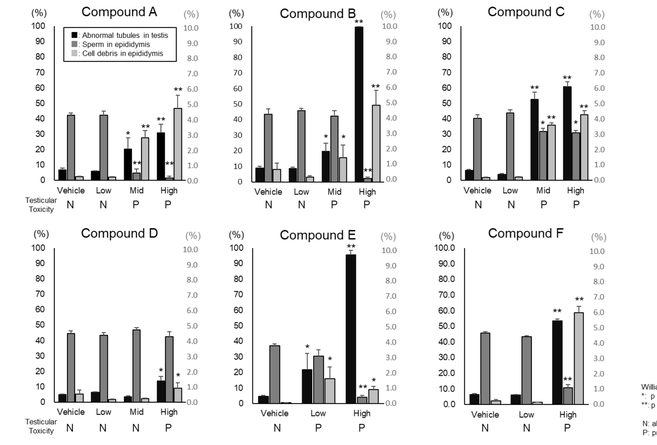


Figure 9: Comparison of quantitative values of findings per group (vehicle to high dose level) with presence of testicular toxicity diagnosed by the pathologists in each study
Vertical axis on the left side shows % (mean value for each group) of abnormal tubules in the Testis, and Vertical axis on the right side shows % of sperm and debris in epididymal duct lumen of the epididymis. The horizontal axis indicates presence/absence of testicular toxicity diagnosed by pathologists. (N: absent, P: present). Black bar: Abnormal tubules in testis, Dark gray bar: Sperm in epididymis, Gray bar: Cell debris in epididymis.
(N: absent, P: present) (Williams’ test; *: p < 0.05, **: p < 0.01)

Compounds A to C: There were significant differences in Abnormal tubules, Sperm or Debris values from Mid dose level.
Compound D: There were significant differences in Abnormal tubules and Debris values at High dose level.
Compound E: There were significant differences in Abnormal tubules and Debris values from Low dose level and Sperm value at High dose level.
Compound F: There were significant differences in all the findings-values at High dose level.

The groups with significantly different quantitative values and those with testicular toxicity diagnosed by pathologists were all consistent.
Quantitative values perfectly explained testicular toxicity diagnosed by the pathologists.

References

1. P. Samanta, G. Raipuria and N. Singhal, Context Aggregation Network For Semantic Labeling In Histopathology Images, IEEE 18th International Symposium on Biomedical Imaging (ISBI), 673-676, 2021
2. P. Samanta, N. Singhal, YAMU-NET: Yet Another Modified U-Net for semantic segmentation, PMLR 172:1019-1033, 2022
3. M. Tan and Q. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, Proceeding of Machine Learning Research 2019
4. M. Ester, H.P. Kriegel, J. Sander and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise. KDD 96: 226-231, 1996
5. D. Creasy, S. Panchal, R. Garg and P. Samanta, Deep Learning-Based Spermatogenic Staging Assessment for Hematoxylin and Eosin-Stained Sections of Rat Testes, Tox Path 49:872-887, 2021