

# Quantitative identification of teratoma tissues formed by human embryonic stem cells with TeratomEye

Steve Kah Weng Oh · Pauline Chua ·  
Koh Lin Foon · Emily Ng · Angela Chin ·  
Andre Boon Hwa Choo · Rajagopalan Srinivasan

Received: 11 November 2008 / Revised: 15 December 2008 / Accepted: 19 December 2008 / Published online: 25 January 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** An automated vision system, TeratomEye, was developed for the identification of three representative tissue types: muscle, gut and neural epithelia which are commonly found in teratomas formed from human embryonic stem cells. Muscle tissue, a common structure was identified with an accuracy of 90.3% with high specificity and sensitivity greater than 90%. Gut epithelia were identified with an accuracy of 87.5% with specificity and sensitivity greater than 80%. Neural epithelia which were the most difficult structures to distinguish gave an accuracy of 47.6%. TeratomEye is therefore useful for the automated identification of differentiated tissues in teratoma sections.

**Keywords** Human embryonic stem cells · Identification · Stem cells · Teratoma

## Introduction

The generally accepted measure of pluripotency for human embryonic stem cells (hESC) is their ability to form differentiated tissues of three representative germ layers within a teratoma 8–10 weeks after injection of hESC into severe combined immunodeficient disease (SCID) mice (Choo et al. 2005; Przyborski 2005). This assay though abstract, has been vigorously debated by the International Stem Cell Banking Initiative (ISCBI) and for now has been accepted by the hESC research community as the standard for pluripotency; as there are no better methods of measuring early tissues developed from hESC (Healy et al. 2008). Representatives of these three different common tissue types such as muscle, gut and neural epithelia are shown in Fig. 1a. Typically, single images of these tissues are shown in publications to indicate that hESC have the capacity to differentiate into the three germ layers, without any further quantification of each tissue type. To aid in the objective identification and partial quantification of these three types of differentiated tissues formed within teratomas from hESC, we have developed three separate algorithms along with a user-friendly graphical interface, named TeratomEye. This program provides a more objective means of identifying tissues, instead of a qualitative one, and allows quantification of each tissue as a measure of pluripotency. Depending on the sites of injection of hESC, the types of tissues that result can be

---

S. K. W. Oh (✉) · A. Chin · A. B. H. Choo  
Stem Cell Group, Bioprocessing Technology Institute,  
A\*STAR (Agency for Science, Technology  
and Research), 20 Biopolis Way #06 - 01 Centros,  
Singapore 138668, Singapore  
e-mail: steve\_oh@bti.a-star.edu.sg  
URL: www.bti.a-star.edu.sg

P. Chua · K. L. Foon · E. Ng · R. Srinivasan  
Department of Chemical and Biomolecular Engineering,  
National University of Singapore, 4 Engineering Drive 4,  
Singapore 117576, Singapore

haphazard and difficult to identify (Cooke et al. 2006). TeratomEye will thus be useful in identifying differentiated tissues in different environments, and more importantly to compare the pluripotency or differentiation capability between various hESC lines deposited in the international stem cell banks.

## Materials and methods

### Image collection

Human embryonic stem cells were injected into SCID mice and, after 10 weeks, teratomas were harvested and sections were prepared and stained with haematoxylin and eosin (H&E) as described previously (Choo et al. 2005). Images were collected at 10× and 20× objective magnification using a Carl Zeiss AxioVert microscope and examined with the imaging software Axio Vision, Release 4.5. All images were sized at 1300 × 1030 pixels and consolidated into a database of 93 images.

### Muscle segmentation

All software was written using Matlab version 7.3. The GETmuscle algorithm for muscle segmentation comprises of three stages. Firstly, the image is converted from RGB color space to  $L^*a^*b^*$  color space, where  $L^*$ ,  $a^*$  and  $b^*$  refer to the luminosity value, chromaticity value on the red-green axis and chromaticity value on blue-yellow axis respectively. This is followed by  $K$ -means (Duda et al. 2000) clustering to classify the image into four distinct components. Muscle tissues are separated from the background and other tissues since they generally have a distinct red/pink coloration enables. The muscle segment is extracted by calculating the Euclidean distance, with the *mean*  $a^*$  and  $b^*$  for each cluster, from a threshold/color-marker. Finally, the identified muscle segment is converted to a binary image to removing trace elements of the background after which the outline is created.

### Gut segmentation

The GETgut algorithm for gut segmentation is affected by a series of morphological operations (Mathworks Inc 1997) which are divided into two stages. The primary aim of the first stage is to

eliminate the background and remove or mask other small elements in the image; this helps to reduce the possibility of neural structures being included in the segments. Gut tissues can be identified by their lumen, as they would constitute maxima in the images. This is followed by conversion to a binary image via thresholding to obtain gut markers. A threshold level was determined by trial-and-error with the images from the training set. This results in a shortlist of possible gut structures.

Each candidate in the shortlist is then evaluated individually in the second stage using partial least squared determinant analysis (PLSDA) (Wise et al. 2004). Thirteen sub-images depicting gut epithelium and 25 sub-images depicting non-gut epithelium structures were used to build the PLSDA model. Each sub-image was compressed to a standard size of 60 × 64 pixels, then unfolded to form an array  $X$ . Mean centering was used to pre-process  $X$ .  $Y$  then contains the corresponding class memberships of each row in  $X$ , where class 1 denotes gut epithelium and class 2 denotes non-gut. The PLSDA model thus developed is used for classifying new candidates.

### Neural segmentation

The same segmentation process for gut tissues was adopted for neural tissues in the GETneural algorithm with thresholds set at different values. The underlying reason for the differences is the smaller dimensions of the neural structures, thus the structuring element for image reconstruction has to be reduced to create the neural tissue maxima (size 7), and a higher binary threshold in order to isolate the neural tissues (0.8196). The development of the PLSDA model for neural epithelium classification is similar to that for gut. Twenty-one sub-images depicting neural epithelium and 35 sub-images depicting non-neural epithelium structures were used to build the PLSDA model for neural classification. In both the neural and gut identification algorithms where there are two phases of selection and identification, both of these algorithms run automatically and there is no need for user intervention for the second phase.

### TeratomEye graphical user interface

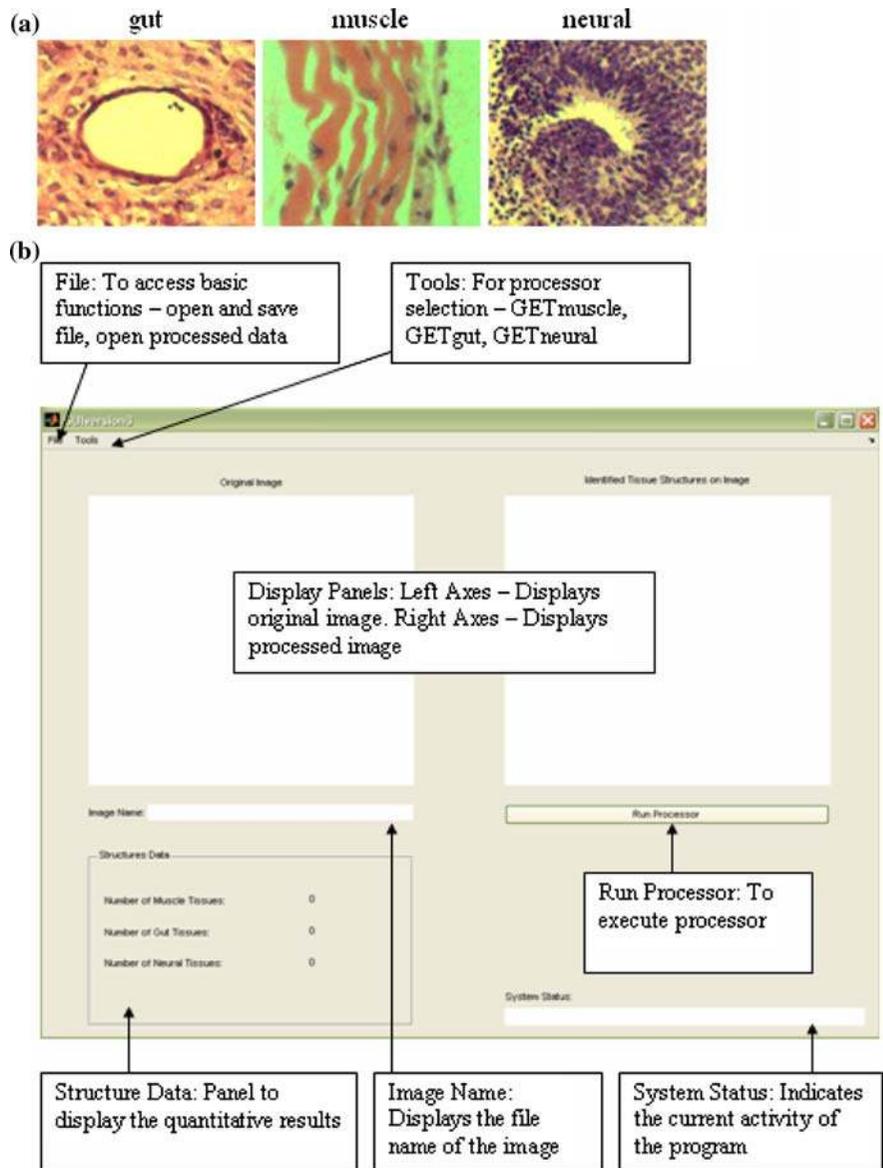
The graphical user interface shown in Fig. 1b was designed for users with limited knowledge of Matlab

and to encourage ease of use. Scanned images can be stored and opened as an original image, then the tools GETmuscle, GETgut, GETneural or GETall can be selected to identify one tissue at a time, or all three types of tissues by running the processor. After which the numbers of each tissues are presented at the bottom left screen. Figure 2 shows an example of eight gut structures identified by TeratomEye. This software is available to other researchers for beta testing from the authors.

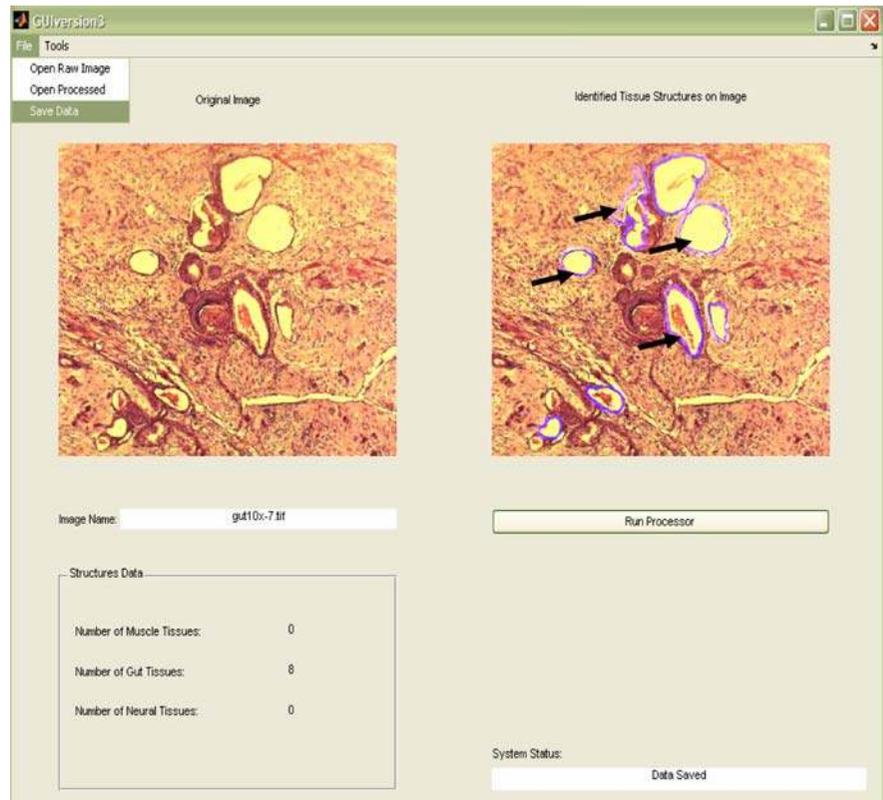
**Results and discussion**

The GETmuscle, GETgut and GETneural algorithms were developed with training images, after which independent images were tested to determine the predictive accuracy, specificity and sensitivity of various tissue identifications, all of which are summarized in Table 1. Muscle was identified by a two-step color-based segmentation process in the L\*a\*b\* color space. By employing K-means clustering as the

**Fig. 1** **a** Typical representative H&E stained pictures of gut structure with a hollow lumen, striated muscle and neural epithelia with a rosette-like structure. **b** Layout of TeratomEye user interface with tools for running GETmuscle, GETgut and GETneural as well as quantification of each tissue



**Fig. 2** An example of eight gut structures identified and displayed on TeratomEye, some of which are indicated by dark arrows



first stage of segmentation, the distinctive red and pink colors of muscle enables clear separation from the background. An example of muscle section clearly identified by GETmuscle is shown in Fig. 3a. The user can clearly perceive the entire outline of the muscle tissues compared to an earlier algorithm where parts of the muscle tissues were excluded. Using a test set of 14 images, muscle was identified with an accuracy of 90.3% with high specificity and sensitivity of greater than 90%. Only on one occasion was the algorithm unable to identify a muscle structure due to poor contrast with the background.

A series of morphological operations was used to isolate gut/neural epithelial structures which were then assigned into grayscale sub-images. PLSDA models were developed to recognize sub-images containing gut/neural epithelia. Fifty test images for gut epithelia gave a predictive accuracy of 87.5% with specificity and sensitivity of greater than 80%. GETgut was able to correctly identify all five gut epithelia as shown in Fig. 3b. Gut structures with large lumen maxima were thus relatively easy to identify.

It was necessary to use 20 training images and 58 test images for identifying neural epithelia with GETneural, which were the most difficult structures to distinguish, giving an accuracy of 47.6%. It is possible that identification using lumen maxima results in some neural epithelium cells, which do not have a distinctive lumen, being excluded during segmentation. Structural diversity of neural structures which are sometime elongated and sometimes more rosette-like further added to the difficulty in prediction. In particular, irregularly shaped neural structures failed to be identified. Examples of the variety of neural structures can be seen in Fig. 3c, in this case only three out of eight neural epithelia with their borders well-defined, were correctly highlighted. Thus we are exploring a wavelet image analysis approach (Misiti et al. 1996) to further improve this program. It may also be necessary to stain with antibodies specifically to highlight neural epithelia prior to identification with TeratomEye to increase the accuracy of this process.

Currently, the classical method of measuring the pluripotency of hESC is by qualitative visualization

**Table 1** Number of images tested, predictive accuracy, specificity and sensitivity of GETmuscle, GETgut and GETneural algorithms for the identification of muscle, gut and neural tissues in TeratomEye

Algorithms	No. of images tested <sup>a</sup>	Predictive accuracy (%)	Specificity (%)	Sensitivity (%)
GETmuscle	14	90.3	92.9	91.9
GETgut	50	87.5	83.8	84
GETneural	58	47.6	56	17.2

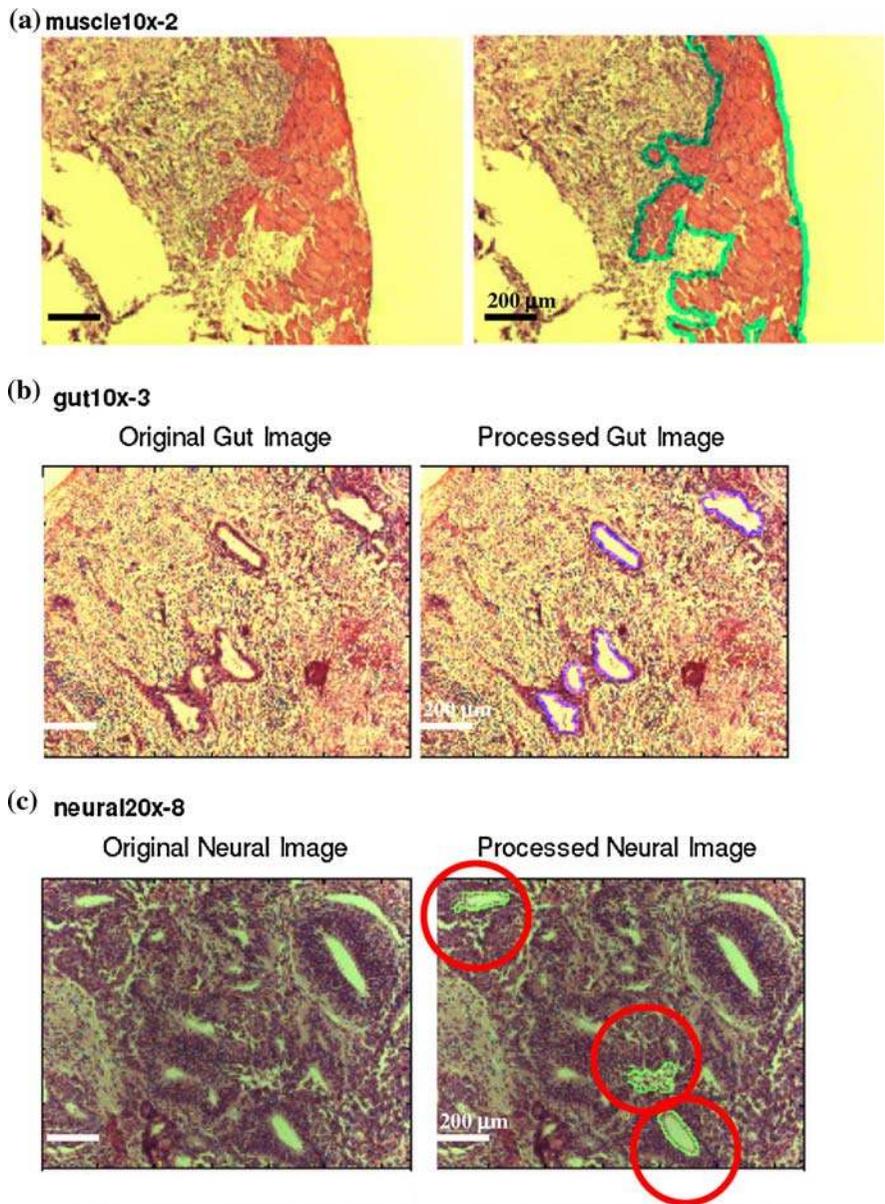
<sup>a</sup> Each image contains several tissue structures for identification

Predictive accuracy = True positives/(true positives + false positives) × 100%

Specificity = True negatives/(false positives + true negatives) × 100%

Sensitivity = True positives/(true positives + false negatives) × 100%

**Fig. 3 a** An example of muscle structures that are identified and highlighted with a green border by GETmuscle. Image was captured at 10× objective magnification. **b** An example of multiple gut structures with several hollow lumens that are identified and highlighted with purple borders by GETgut. Image was captured at 10× objective magnification. **c** An example of GETneural identifying three neural epithelia highlighted by red circles, from multiple other neural epithelia. Image was captured at 20× objective magnification



followed by choosing only one representative image of each tissue from the three germ layers found in teratomas of SCID mice models as shown in Fig. 1a. Examples of such tissues can be found in recent publications characterising hESC pluripotency (International Stem Cell Initiative 2007; Cooke et al. 2006; Przyborski 2005). This traditional method may require expert help from a trained pathologist and provides no quantitative data on the numbers of differentiated tissues found in teratomas. The International Stem Cell Banking Initiative has accepted teratoma formation as a measure of pluripotency ([http://www.stemcellforum.org/forum\\_initiatives/international\\_stem\\_cell\\_banking\\_initiative.cfm](http://www.stemcellforum.org/forum_initiatives/international_stem_cell_banking_initiative.cfm)) and it has been suggested recently that teratoma formation may provide a window to study developmental biology (Aleckovic and Simon 2008).

The creation of the TeratomEye program therefore, is aimed at providing embryonic stem cell researchers with a means to more objectively identify and potentially quantify the number of common structures such as muscle, gut and to a lesser extent, neural epithelia. For many stem cell researchers who are not trained as pathologists, this software could provide an automated and easy means for qualitatively identifying the three common structures found in teratoma tissues. As TeratomEye can also count the numbers of muscle, gut and neural structures in tissue sections, potentially different hESC lines could be compared for their propensity to form the three types of differentiated tissues in teratomas. However, TeratomEye has not yet been tested on different hESC lines, but there is anecdotal evidence that some hESC lines are more likely to form cystic structures which are indicative of poorer teratoma formation. Other structures such as bone and cartilage which are also found in teratomas may be added to this program in the future. Potentially, cartilage may be easy to resolve as it has a distinctive round shape with spotted nuclei. While there are other commercially available software which can identify tissue sections for oncology, ophthalmology and diabetes research applications, for example that provided by Aperio ([www.aperio.com](http://www.aperio.com)) there is none available for the identification of a variety of complex tissues found in teratoma.

**In summary**, we have developed an automated vision system TeratomEye, which identified muscle

with an accuracy of 90.3% with specificity and sensitivity greater than 90%. Gut epithelia were identified with accuracy of 87.5% with specificity and sensitivity greater than 80%. Neural epithelia which were the most difficult structures to distinguish, gave an accuracy of 47.6%. With further refinements, TeratomEye can be a useful tool for the automated identification of tissues in teratoma sections enabling a quantitative measure of pluripotency of human embryonic stem cells injected into SCID mice models.

**Acknowledgements** We thank the Agency for Science Technology & Research (A\*STAR) for generous funding of our research and Dr. Jeremy Crook for critically reviewing this manuscript.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Aleckovic M, Simon C (2008) Is teratoma formation in stem cell research a characterization tool or a window to developmental biology? *Reprod Biomed Online* 17:270–280
- Choo ABH, Padmanabhan J, Chin A, Fong WJ, Oh SKW (2005) Immortalised feeders for the scale-up of human embryonic stem cells in feeder and feeder-free conditions. *J Biotechnol* 122:130–141
- Cooke MJ, Stojkovic M, Przyborski SA (2006) Growth of teratomas derived from human pluripotent stem cells is influenced by the graft site. *Stem Cells Dev* 15:254–259
- Duda RO, Hart PE, Stork DG (2000) *Pattern classification*, 2nd edn. Wiley, London
- Healy LE, Ludwig TE, Choo A (2008) International banking: checks, deposits and withdrawals. *Cell Stem Cells* 2:305–306
- International Stem Cell Initiative (2007) Characterization of human embryonic stem cell lines by the international stem cell initiative. *Nat Biotechnol* 25:803–816 (Epub 2007 June 17)
- Mathworks Inc (1997) *Image processing toolbox—user's guide*. Math Works, Inc
- Misiti M, Misiti Y, Oppenheim G, Poggi JM (1996) *Wavelet toolbox for use with MATLAB™*. MathWorks, Inc
- Przyborski SA (2005) Differentiation of human embryonic stem cells after transplantation in immune-deficient mice. *Stem Cells* 23:1242–1250
- Wise BM, Shaver JM, Gallagher NB, Windig W, Bro R, Koch RS (2004) *PLS\_Toolbox 3.5 for use with MATLAB™*. Eigenvector Research, Inc