Cell PRESS
Cell Stem Cell
Forum

A Call to Standardize Teratoma Assays Used to Define Human Pluripotent Cell Lines

Franz-Josef Müller,[1,2,*] Johanna Goldmann,[1] Peter Löser,[3] and Jeanne F. Loring[1,*]
[1]Center for Regenerative Medicine, Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA 92037, USA
[2]Zentrum für Integrative Psychiatrie, 24105 Kiel, Germany
[3]Robert Koch Institute, 13353 Berlin, Germany
*Correspondence: fj.mueller@zip.kiel.de (F.-J.M.), jloring@scripps.edu (J.F.L.)
DOI 10.1016/j.stem.2010.04.009

The teratoma assay is the gold standard for documenting pluripotency of human stem cells. However, reports of new human ESC and iPSC lines vary widely in both methods and analysis of teratoma data. We call for consensus standards to be established to make this assay worthy of its "golden" status.

In this journal, researchers have recently participated in a lively discussion about standards that could be used in characterization of human pluripotent stem cells (Daley et al., 2009; Ellis et al., 2009; Maherali and Hochedlinger, 2008).

For mouse cells, the gold standard for proving pluripotency is germline transmission (Bradley et al., 1984), which demonstrates the ability to make all cell types, including germ cells. For human cells, the closest surrogate for the germline transmission assay is the generation in immunodeficient mice of human cell-derived teratomas, solid tumors that contain a mixture of differentiated tissues such a neurons, heart muscle, and secretory epithelia (Damjanov, 2005). Because of the rich variety of mature histologically distinct tissue types that develop, the teratoma assay is currently regarded as the most rigorous assay to prove pluripotency of human stem cells.

Unfortunately, although the end point of a germline transmission assay is simple (a mouse derived from cultured cells), the teratoma assay necessarily lacks this level of absolute clarity. To determine what methods and outcome measures are used for teratoma analysis, we systematically screened the literature from 1998 to 2009 (Table 1). In spite of the gold standard status of the teratoma assay for determining pluripotency, we found that there was little consistency in either methods or reporting of results. The protocols differed widely in key parameters, such as preparation of cells, site of injection, and number of cells injected per animal. We found that the majority of studies do not report key factors, such as a systematic histopathological evaluation or even the number of mice transplanted.

The variation in reporting of teratoma results brings up the question of this assay's value as a standard for proving pluripotency of human stem cells. Given the accelerating rate of derivation of new human stem cell lines, we would like to propose that the stem cell research community work toward improving the rigor of assays used for defining cell pluripotency, including the teratoma assay.

In many ways this situation is analogous to the challenges faced by other groups of scientists whenever novel or little-used technologies capture the attention of researchers in prominent fields. For example, the rapid adoption of microarray technology in the late 1990s led to an explosion of data in the literature, and scientists became concerned about the inconsistency of methods and reporting of microarray data. The lack of standards led to inaccurate reports and the broad scientific community was justifiably suspicious of the methods (Loring, 2006). In response to these concerns, users and developers of microarray technology proposed that researchers agree on a set of standards for reporting of microarray data (Brazma et al., 2001). These criteria, called "Minimum Information About a Microarray Experiment" (MIAME), have transformed the reporting of microarray data in the literature, simplifying the review of manuscripts and building confidence in the validity of the data.

In this Forum, we would like to initiate a similar conversation about establishing standards for the analysis of human pluripotent stem cells. Focusing on the most highly regarded assay for pluripotency, we propose that stem cell researchers explore a systematic reporting system for teratoma experiments and data. Such standards may aid pluripotent stem cell researchers to better evaluate and compare results across different reprogramming strategies and differentiation protocols and to contribute valuable information about human development.

We surveyed more than 1200 original research manuscripts that were published from 1998 to 2009 in English language journals, indexed in NCBI Medline, and describing research on human embryonic stem cells (hESCs). We also identified 124 original research manuscripts that were published from 2007 to 2009 that report research on human induced pluripotent stem cells (iPSCs).

For an initial survey of trends and main findings, we analyzed in-depth 95 papers describing the successful derivation of 639 novel hESC lines and 81 papers describing the establishment of 777 novel human iPSC lines (Table 1). Detailed descriptions of reports and teratoma results on the single manuscript level with citations can be accessed in Table S1 available online.

To our surprise, we found that for more than half of the 639 published hESC lines (355 lines, 56%) and 64% of the published iPSC lines (501 lines), no teratoma data were included in the manuscripts that reported the establishment of these cell lines. These numbers indicate that the gold standard for demonstration of pluripotency in human cells has been used for only half of novel hESC and human iPSC lines reported in the scientific literature. Although we are unable to analyze the review process of manuscripts, our

Cell PRESS

Table 1. Summary of Our Systematic Literature Survey

	Papers Examined	Lines Reported	Lines with Teratoma Data	Avg. AIS of Journals with Teratoma Data	Avg. AIS of Journals without Teratoma Data	Spearman Rank Correlation
hESC	95	639	44%	3.442	1.177	positive: 0.253 (p = 0.0134)
hiPSC	81	777	36%	7.339	8.165	weak positive: 0.084 (p = 0.4564)

Publications identified in a Pubmed/Medline search (see Table S1 for detailed screen and inclusion strategies) for novel hESC and hiPSC lines were reviewed for inclusion of teratoma data. The Eigenfactor Article Influence Score (AIS) was retrieved for each article (see Table S1 for details and examples of relatively high and low AIS scores). The Spearman rank correlation, rho, was used assay for a correlation between the journal's AIS and the number of cell lines tested with the teratoma assay in a given study. Note: Papers reporting teratoma data of previously established cell lines (follow-up studies) were not included in this analysis.

anecdotal experience is that reviewers for higher-impact journals are more likely to require teratoma data in order to pass peer review. Interestingly, when we analyzed the data, we discovered that our impression was correct for hESCs, but not for iPSCs; for novel hESCs, there was significant positive correlation between the impact (Eigenfactor Article Influence Score) of a journal and inclusion of teratoma results, but for iPSC lines, although there was a trend toward positive correlation, there was no significant distinction among journals (Table 1).

A second issue is the reporting of teratoma results. We found that even when a teratoma assay is included, the methodology for inducing teratomas is often poorly reported. The descriptions in the methods sections varied so widely that it was impossible to classify them into groups. To illustrate the problem, we describe several types of reporting that are representative for the group of papers, without identifying the specific papers. The number of injected cells is often reported with vague quantifications, for example "clumps of 200–300 cells" without mentioning how many of such clumps were transplanted. Moreover, cell numbers for induction of teratomas vary widely; even transplantations into the same site are performed with cell numbers varying more than three orders of magnitude (e.g., 3000 to 5 million cells for testicular injections). The time "in vivo" varies from 4 to 15 weeks, and the size of the resected tumors is mentioned in very few cases. Only 16 of the hESC and 3 of the iPSC papers mention the passage number of the cells used for teratoma experiments. Just 16 of the hESC and only 10 of the iPSC papers note the number of animals used for the experiments. In six hESC reports and five papers on iPSCs that report teratomas, no single detail on how the teratomas were induced and analyzed is given.

In addition to the variability in information reported for generating teratomas, there is also a large inconsistency in reporting of analysis of the tumors. The teratomas are examined in most cases by classic histological methods, via hematoxylin/eosin (H&E) stains, and usually only isolated examples of tissue types are provided, with no quantification or basis for comparison across studies. In general, a histopathologist is consulted to identify tissue types within the tumor, and to classify them as derivatives from ectoderm, mesoderm, or endoderm. Only a small fraction of the studies (17 of 75 hESC papers and 16 of 67 iPSC papers that provided data on teratoma assays in first reports of novel cell lines) included immunohistochemical staining of teratomas for differentiation markers. It is not possible to quantify the types of tissues reported because of the heterogeneity of the reported results, but we have the impression that the histological analysis is generally subjective and dependent on the experience of the pathologist.

We were also surprised to find that five hESC lines were reported to fail to form teratomas when injected into immunodeficient mice, and several lines formed only small tumors consisting mainly of fluid-filled cysts. Also, individual reports indicated that it was difficult to obtain well-differentiated teratomas from certain iPSC lines. This trend may be relevant to our understanding of pluripotency, because the "definitive" pluripotent stem cell phenotype remains ill defined and it has been reported by several independent groups that even partially reprogrammed iPSC lines possess the ability to form teratomas (Chan et al., 2009).

The clear conclusion from our review of the literature is that accepted and practiced standards for both conducting and reporting teratoma assays currently do not exist. We would like to argue that if teratomas are to be regarded as the gold standard for demonstrating pluripotency, then the stem cell community would benefit from development of consensus standards for performing and reporting teratoma results.

Another important reason for improving reporting of teratoma results is the tremendous amount of information that they can provide beyond their use as proof of pluripotency. The cellular structure of teratomas has the potential to offer insights into the development of human embryonic tissues and differences among cell lines. There are numerous questions that could be addressed by following consensus criteria for reporting teratoma results. For example, does the site of injection and number of cells affect the types of tissues that develop and/or the timing of their development? How mature do tissues become in teratomas—do they more resemble early stages of human development or fully differentiated adult tissues? Does the source of cell type for reprogramming to iPSCs bias the differentiation of the cells: for example, if cartilage precursors are reprogrammed, are they better at making cartilage than cells reprogrammed from dermal fibroblasts? Can we use teratoma analysis to determine whether clinical transplants of derivatives of particular stem cell lines are more likely to form dangerous tumors than others?

Of the two distinct issues regarding teratomas, agreement about the mechanics of teratoma generation (dissociation methods, number of cells, site of injection) may be easier to achieve than coming to a consensus about standards for interpretation of teratoma results (tumor descriptions, identification of tissues). Because histological analysis of embryonic tissues is a skill that had its dominance in the 1950s and 1960s, very few of the current generation of stem

cell researchers has any experience at all in histology. As a result, the task of assessing the tissue content of teratomas is usually outsourced to professional pathologists, and the stem cell researchers who generated the teratomas typically learn very little beyond a "checklist" of isolated examples of derivatives of the three germ layers: ectoderm, mesoderm, and endoderm. This type of reporting adds little to our knowledge about the variety of tissues and cell types that develop in teratomas; there is not even a standard basis for the identification of a structure as a derivative of a particular germline. As an anecdotal example, we recently sought input from a local pathologist colleague during our analysis of a group of teratomas generated from human iPSCs. She noted the frequent appearance of yolk sac cells in the sections. The yolk sac is an extraembryonic endodermal tissue, and yolk sac tumors, which are aggressively metastatic, are most often found in testicular tissue. Thus, it is reasonable for this cell type to appear in stem cell-derived human teratomas. But surprisingly, the literature contains no reports of yolk sac tissue in tumors generated for the purpose of proving pluripotency of human stem cells. With our preoccupation with identifying tissues from the three germ layers, most of us have lost sight of the fact that extraembryonic tissues play an essential role during embryogenesis, and there is no reason not to expect to find them in teratomas. This observation raises an interesting point that deserves further discussion: in filling out the checklist of germ layers at the request of reviewers of a publication, is it important whether the histological identification of endodermal derivatives in teratomas correctly distinguishes embryonic from extraembryonic tissues?

We acknowledge that this Forum article raises many questions and addresses only a few. In particular, we have highlighted that reports of human pluripotent lines are often devoid of teratoma data, and when the assays are conducted, they are performed, analyzed, and reported inconsistently across the literature. A recent methods publication authored by the ISSCR standards committee (Gertow et al., 2007) indicates that members of the field can come together to begin to establish consistency in methodology and reporting. To extend this discussion, we have compiled a list of criteria (Table S2) that could potentially be incorporated into a standard teratoma reporting system. We would like to promote the idea that having standards for reporting methods and results of teratoma assays will benefit the stem cell community, not only by making the assay more reproducible, but also by providing deeper knowledge about human development.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two tables and can be found with this article online at doi:10.1016/j.stem.2010.04.009.

ACKNOWLEDGMENTS

F.-J.M. is supported by an Else Kröner-Fresenius-Stiftung fellowship, and J.F.L. by grants from the California Institute of Regenerative Medicine, the NIH, The Bill and Melinda Gates Foundation, and the O'Keeffe Foundation. We are in debt to Jacqueline Schirm and Steven Posselt for their help with database searches. Dr. Mana Parast helped us with expert advice on the histopathological assessment of teratoma tissue. We would also like to thank Anja Fritz and David Barker for helpful discussions. The authors are grateful to Louise Laurent, Suzanne Peterson, and Ha Tran for their intellectual input in the early planning of this manuscript.

REFERENCES

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Nature *309*, 255–256.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., et al. (2001). Nat. Genet. *29*, 365–371.

Chan, E.M., Ratanasirintrawoot, S., Park, I.H., Manos, P.D., Loh, Y.H., Huo, H., Miller, J.D., Hartung, O., Rho, J., Ince, T.A., et al. (2009). Nat. Biotechnol. *27*, 1033–1037.

Daley, G.Q., Lensch, M.W., Jaenisch, R., Meissner, A., Plath, K., and Yamanaka, S. (2009). Cell Stem Cell *4*, 200–201, author reply 202.

Damjanov, I. (2005). Stem Cell Rev. *1*, 273–276.

Ellis, J., Bruneau, B.G., Keller, G., Lemischka, I.R., Nagy, A., Rossant, J., Srivastava, D., Zandstra, P.W., and Stanford, W.L. (2009). Cell Stem Cell *4*, 198–199, author reply 202.

Gertow, K., Przyborski, S., Loring, J.F., Auerbach, J.M., Epifano, O., Otonkoski, T., Damjanov, I., and Ahrlund-Richter, L. (2007). Curr. Protoc. Stem Cell Biol., *Chapter 1*, Unit1B 4.

Loring, J.F. (2006). Neurobiol. Aging *27*, 1084–1086.

Maherali, N., and Hochedlinger, K. (2008). Cell Stem Cell *3*, 595–605.