OMICS A Journal of Integrative Biology
Volume 10, Number 2, 2006
© Mary Ann Liebert, Inc.

Development of the Minimum Information Specification for *In Situ* Hybridization and Immunohistochemistry Experiments (MISFISHIE)

ERIC W. DEUTSCH,[1] CATHERINE A. BALL,[2] G. STEVEN BOVA,[3] ALVIS BRAZMA,[4] ROGER E. BUMGARNER,[5] DAVID CAMPBELL,[1] HELEN C. CAUSTON,[6] JEFF CHRISTIANSEN,[7] DUNCAN DAVIDSON,[7] LILLIAN J. EICHNER,[1,8] YOUNG AH GOO,[1,8] SEAN GRIMMOND,[9] THORSTEN HENRICH,[10] MICHAEL H. JOHNSON,[1] MARTIN KORB,[1] JASON C. MILLS,[11] ASA OUDES,[1,8] HELEN E. PARKINSON,[4] LAURA E. PASCAL,[1,8] JOHN QUACKENBUSH,[12] MIRANA RAMIALISON,[10] MARTIN RINGWALD,[13] SUSANNA-A. SANSONE,[4] GAVIN SHERLOCK,[14] CHRISTIAN J. STOECKERT, JR.,[15] JASON SWEDLOW,[16] RONALD C. TAYLOR,[17] LAURA WALASHEK,[1,8] YI ZHOU,[18] ALVIN Y. LIU,[1,8] and LAWRENCE D. TRUE[19]

ABSTRACT

We describe the creation process of the Minimum Information Specification for *In Situ* Hybridization and Immunohistochemistry Experiments (MISFISHIE). Modeled after the existing minimum information specification for microarray data, we created a new specification for gene expression localization experiments, initially to facilitate data sharing within a

[1]Institute for Systems Biology, Seattle, Washington.
[2]Department of Biochemistry, Stanford University School of Medicine, Stanford, California.
[3]Departments of Pathology, Health Information Sciences, Genetic Medicine, Oncology, and Urology, Johns Hopkins University School of Medicine, Baltimore, Maryland.
[4]EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom.
[5]Department of Microbiology, University of Washington, Seattle, Washington.
[6]MRC Human Genetics Unit, Western General Hospital, Edinburgh, United Kingdom.
[7]CSC/Imperial College School of Medicine Microarray Centre, London, United Kingdom.
[8]Department of Urology, University of Washington, Seattle, Washington.
[9]Institute for Molecular Bioscience, University of Queensland, St. Lucia, Australia.
[10]EMBL Heidelberg, Heidelberg, Germany.
[11]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri.
[12]Dana-Farber Cancer Institute, Boston, Massachusetts.
[13]The Jackson Laboratory, Bar Harbor, Maine.
[14]Department of Genetics, Stanford University School of Medicine, Stanford, California.
[15]Center for Bioinformatics and Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania.
[16]Division of Gene Regulation and Expression, Wellcome Trust Biocentre, University of Dundee, Dundee, Scotland.
[17]Pacific Northwest National Laboratory, Richland, Washington.
[18]Children's Hospital Boston and Harvard Medical School, Boston, Massachusetts.
[19]Department of Pathology, University of Washington, Seattle, Washington.

consortium. After successful use within the consortium, the specification was circulated to members of the wider biomedical research community for comment and refinement. After a period of acquiring many new suggested requirements, it was necessary to enter a final phase of excluding those requirements that were deemed inappropriate as a minimum requirement for all experiments. The full specification will soon be published as a version 1.0 proposal to the community, upon which a more full discussion must take place so that the final specification may be achieved with the involvement of the whole community.

This paper is part of the special issue of OMICS on data standards.

INTRODUCTION

THE NIH/NIDDK Stem Cell Genome Anatomy Projects (SCGAP) funded seven groups with a common goal of studying stem/progenitor cells in blood, bone, kidney, liver, gastrointestinal tract, prostate, and bladder, and with a mandate to make all data generated rapidly available to each other and to the research community via a public database (⟨www.scgap.org/⟩). This sharing and collaborative analysis of data will, it is hoped, lead to greater insight into cell differentiation and organ development.

To share these data effectively, we agreed that a common set of standards for all large data types is needed. Where such data standards already exist, such as those developed by Microarray Gene Expression Data (MGED) Society (Ball and Brazma, *this issue*), we adopted them. MIAME (Brazma et al., 2001) defines a specification of minimum information that must be provided for others to analyze independently and to reproduce, if necessary, the results of microarray-based experiments. MAGE-OM/ML (Spellman et al., 2002) defines an object model and markup language for encoding the information about a microarray experiment. Following this example, we have defined MISFISHIE—the Minimum Information Specification for *In Situ* Hybridization and Immunohistochemistry Experiments—for data generated from visual interpretation-based studies, such as *in situ* hybridization and immunohistochemistry, in which protein and gene transcript localizations (hereafter referred to as "gene expression localization experiments") are determined.

While no official minimum specification for such data yet exists, there have been several efforts at organizing gene expression localization data in databases and such database designs provide a useful framework from which to build MISFISHIE. Two notable databases specialized for the mouse community—the Mouse Gene Expression Database (GXD) (Hill et al., 2004) and the Edinburgh Mouse Atlas Gene Expression (EMAGE) database (Christiansen et al., 2006)—have influenced the design of MISFISHIE.

Following its successful use and refinement within the SCGAP consortium, we believe that MISFISHIE may well be of benefit to the wider community of biomedical researchers. Therefore, we have been circulating a draft manuscript describing MISFISHIE in detail to researchers and journal editors in order to build a consensus on which pieces of information should be provided at minimum about such experiments.

CREATING A STANDARD

The MISFISHIE specification describes the types of information that should be provided for each gene expression localization experiment organized in six sections: (1) Experimental Design, (2) Biomaterials and Treatments, (3) Reporters, (4) Staining, (5) Imaging Data, and (6) Image Characterizations. A checklist is provided for quick and easy reference, and to promote adherence to the specification. An article describing MISFISHIE in detail is forthcoming (Deutsch et al., 2006).

The initial specification used in the consortium as we began generating data was fairly simple and somewhat variably interpreted. As it became more widely circulated, we entered a period of inflation, where the specification acquired many new requirements suggested by researchers. However, there was soon considerable alarm that MISFISHIE had become bloated with too many requirements. Thus, the final phase of its development involved excluding those that are deemed inappropriate as a *minimum* requirement for *all* ex-

periments. Many of these experimental parameters may well have some impact on the results of some experiments, but we came to understand that they are irrelevant for many other experiments. Including them in a *minimum* information specification would only serve to hinder acceptance of MISFISHIE by the community. This initial expansion and later contraction in response to criticism seemed to be a healthy process in the formulation of a community-based specification.

As with any minimum information specification, there will inevitably be different views on what needs to be listed in MISFISHIE. The more contentious requirements are how precisely the reporter should be defined (e.g., must the exact probe sequence be provided), if an image of each assay should be provided, and if and how interpretations of the images should be provided. Consensus requirements that pleased everyone were not reached for these items, but after long discussions, we came to a point where most of those who had voiced their opinions on these issues accepted the resultant effort as a workable specification.

Briefly, MISFISHIE describes the minimum information as (1) experimental design attributes such as a brief description, the assay types, experimental factors (variables among the assays), and contact information; (2) biomaterial attributes such as physical attributes, physiologic state, and relevant exogenous factors associated with the individuals from which the specimens were obtained plus treatment protocols applied to the specimen prior to staining; (3) reporter (probe or antibody) information, such as unambiguous reporter identification or sequence and protocol for obtaining the reporter; (4) staining protocol and parameters including detection methods, reagents, and details about positive and negative controls; (5) imaging data, at minimum one representative image per assay, that can be downloaded to a computer and explored; and (6) image characterizations in a tabular form per assay using properly defined intensity scale and structural units (i.e., organs, glands, cell types).

Once we developed a complete draft of the specification and a simplified checklist, we tested MISFISHIE by first having a dozen or so MISFISHIE authors review several recently published papers and compare notes on their ratings for compliance. Next, a pool of a few dozen articles was assigned among these reviewers, each one tasked with assessing the articles for MISFISHIE compliance. The survey results were compiled and are reported along with the specification (Deutsch et al., 2006). This exercise proved to be a useful tool in refining the specification and identifying the more notable shortfalls in compliance in the current literature.

We believe that it has been beneficial to separate the tasks of defining a minimum information specification and a full-blown object model for this data type, as was done with MIAME and MAGE-OM. This separation allowed contributors first to focus discussions on what information is necessary while deferring the more difficult question of how to model the data to a later stage. Now that the minimum information specification is well underway, we will develop an object model and ontology, building on the Functional Genomics Experiment Object Model (FuGE-OM) (Jones et al., *this issue*) and Functional Genomics Ontology (FuGO) (Whetzel et al., *this issue*)

In order to give our specification a permanent home in the larger community, a MISFISHIE working group has been created within the MGED Society. This provides MISFISHIE with a permanent web presence (⟨www.mged.org⟩) and opportunities for discussion at MGED meetings, and involvement by researchers familiar with the crafting of community-based standards for data sharing.

CONCLUSION

The full MISFISHIE specification is now ready for publication. If the manuscript is accepted, it will become a version 1.0 proposal to the community. However, further discussions must take place as the eventual accepted standard cannot be dictated but must rather be reached by consensus. Suggestions from the community are encouraged and will be collected and folded into a second release, to be promulgated at the MISFISHIE area of the MGED website (⟨http://mged.sourceforge.net/misfishie/⟩). Comments may be addressed to the email distribution list dedicated to MISFISHIE (⟨mged-misfishie@lists.sourceforge.net⟩). If a consensus on the minimum requirements in MISFISHIE can be achieved, and if the specification becomes a requirement for publication, we think this will significantly improve the completeness of data reporting and reusability of the data.

ACKNOWLEDGMENTS

We thank Jules Berman, Rachel Drysdale, Mervi Heiskanen, and Monte Westerfield for comments and discussions during the preparation of MISFISHIE. This work was funded in part with support from the National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health, to members of the Stem Cell Genome Anatomy Project Consortium (including 1U01 DK63483 to Jeff Gordon, Washington University in St. Louis; 1U01 DK63481 to Ihor Lemischka, Princeton University; 1R01 DK63400 to Melissa Little, University of Queensland; 1U01 DK63630 to Alvin Liu, University of Washington; and 1U01 DK63328 to Leonard Zon, Children's Hospital, Boston.

REFERENCES

BALL, C.A., and BRAZMA, A. (2006). MGED standards: work in progress. OMICS (*this issue*).

BRAZMA, A., HINGAMP, P., QUACKENBUSH, J., et al. (2001). Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat Genet **29,** 365–371.

CHRISTIANSEN, J.H., YANG, Y., VENKATARAMAN, S., et al. (2006). EMAGE: a spatial database of gene expression patterns during mouse embryo development. Nucleic Acids Res **34,** D637–D641.

DEUTSCH, E.W., BALL, C.A., BRAZMA, A., et al. (2006). Minimum information specification for *in situ* hybridization and immunohistochemistry experiments (MISFISHIE). Nature Biotechnology (submitted).

HILL, D.P., BEGLEY, D.A., FINGER, J.H., et al. (2004). The mouse Gene Expression Database (GXD): updates and enhancements. Nucleic Acids Res **32,** D568–D571.

JONES, A.R., PIZARRO, A., SPELLMAN, P., et al. FuGE: Functional Genomics Experiment object model. OMICS (*this issue*).

SPELLMAN, P.T., MILLER, M., STEWART, J., et al. (2002). Design and implementation of Microarray Gene Expression Markup Language (MAGE-ML). Genome Biol **3,** RESEARCH0046.

WHETZEL, P.L., BRINKMAN, R.R., CAUSTON, H.C., et al. (2006). Development of FuGO: an ontology for functional genomics investigations. OMICS (*this issue*).

Address reprint requests to:
Dr. Eric W. Deutsch
Institute for Systems Biology
1441 N 34th St.
Seattle, WA 98103

E-mail: edeutsch@systemsbiology.org