Minimum information specification for *in situ* hybridization and immunohistochemistry experiments (MISFISHIE)

Eric W Deutsch*[1], Catherine A Ball[2], Jules J Berman[3], G Steven Bova[4], Alvis Brazma[5], Roger E Bumgarner[6], David Campbell[1], Helen C Causton[7], Jeffrey H Christiansen[8], Fabrice Daian[9], Delphine Dauga[9], Duncan R Davidson[8], Gregory Gimenez[9], Young Ah Goo[1,10], Sean Grimmond[11], Thorsten Henrich[12], Bernhard G Herrmann[13], Michael H Johnson[1], Martin Korb[1], Jason C Mills[14], Asa J Oudes[1,10], Helen E Parkinson[5], Laura E Pascal[1,10], Nicolas Pollet[15], John Quackenbush[16], Mirana Ramialison[12], Martin Ringwald[17], David Salgado[9], Susanna-Assunta Sansone[5], Gavin Sherlock[18], Christian J Stoeckert Jr[19], Jason Swedlow[20], Ronald C Taylor[21], Laura Walashek[1,10], Anthony Warford[22], David G Wilkinson[23], Yi Zhou[24], Leonard I Zon[24], Alvin Y Liu[1,10] & Lawrence D True[25]

One purpose of the biomedical literature is to report results in sufficient detail that the methods of data collection and analysis can be independently replicated and verified. Here we present reporting guidelines for gene expression localization experiments: the minimum information specification for *in situ* hybridization and immunohistochemistry experiments (MISFISHIE). MISFISHIE is modeled after the Minimum Information About a Microarray Experiment (MIAME) specification for microarray experiments. Both guidelines define what information should be reported without dictating a format for encoding that information. MISFISHIE describes six types of information to be provided for each experiment: experimental design, biomaterials and treatments, reporters, staining, imaging data and image characterizations. This specification has benefited the consortium within which it was developed and is expected to benefit the wider research community. We welcome feedback from the scientific community to help improve our proposal.

High-throughput analyses of gene expression in biological samples (for example, using microarrays or proteomics) often do not provide information about the cell types or spatial domains within tissues in which genes are expressed and may not reveal dynamic or transient gene expression. Therefore, such analyses are often followed by experiments to determine the location and degree of gene expression in specific cell types within the tissue by probing with reporters for the genes of interest. In addition, high-throughput analyses of fresh samples can be supplemented with a wealth of clinical information associated with tissue samples in large collections worldwide. However, studies that use

in situ hybridization (ISH) and immunohistochemistry (IHC) staining and/or their resulting images are often presented without the information needed to understand the images or the methods that produced them. Furthermore, neither the reagents and methods nor the results are easily searchable through current biomedical literature databases such as PubMed. As the interpretation of ISH and IHC stains may differ between observers, between different image analysis platforms and programs, and even between different sessions using the same image analysis platform and program[1], communication of the methods used is critically important for evaluating published work.

Data annotation specifications for microarray experiments[1–4] have begun to benefit the biomedical research community. Many researchers participated in the debate around MIAME and contributed to its development. The accessibility of data increased significantly, aided by common exchange formats; open-source software and ontologies were developed by many groups; and discussion forums promoted interaction between manufacturers and experimenters. Similar guidelines are under development for other high-throughput technologies[5–10].

In the area of microscopy images, data formats that facilitate the exchange of data have been proposed. The XML (Extensible Markup Language) data format for tissue microarrays does not include minimum-information reporting guidelines[11]. Also available is Open Microscopy Environment (OME), which provides a flexible XML data format for storing and transmitting metadata for microscopy image datasets (see **Table 1** for URL). However, there is no comprehensive specification for facilitating the exchange of data from visual interpretation–based experiments that seek to determine the abundance and/or localization of proteins or mRNA in tissues (hereafter referred to as 'gene expression localization experiments'), such as ISH and IHC.

The goals of MISFISHIE

MISFISHIE describes the minimum information to be provided when publishing, making public or exchanging results from visual interpretation–based tissue gene expression localization experiments, such as

*A list of affiliations appears at the end of the paper. Correspondence should be addressed to E.W.D. (edeutsch@systemsbiology.org).

ISH, IHC, lectin affinity histochemistry, and experiments that involve reporter gene constructs (for example, green fluorescent protein (GFP) and β-galactosidase). Compliance with this specification should enable researchers at different laboratories to fully evaluate data and reproduce experiments. Although MISFISHIE facilitates the identification of specific sources of variability, it cannot, and does not aim to, reduce this variability. However, if complete information, including raw image data, is always provided, the original interpretations may be reevaluated by other researchers. Like MIAME[12], MISFISHIE prescribes that the most relevant details within each of the sets of broad categories of information be provided, relying on data producers and reviewers to ensure that each category contains the information deemed necessary to allow readers to fully assess and reproduce experiments.

MISFISHIE does not dictate a specific format for reporting information. We intend to develop a data model based on the concepts of MAGE-OM (MicroArray Gene Expression Object Model) and software based on the MAGEstk (MicroArray Gene Expression software tool kit)[3]. This model and the associated XML-based mark-up language will provide a data format for archiving or transferring data. Because a major revision of MAGE-OM, the FuGE-OM (Functional Genomics Experiment Object Model)[13], is at present being developed to accommodate data from other functional genomics experiments, the MISFISHIE-derived object model will probably be an extension of FuGE-OM rather than a separate construct. A simpler, non-XML format following the concepts of MAGE-TAB[14] may also facilitate data sharing in cases where simplicity is most important[15].

MISFISHIE was designed to function together with other technology-related specifications such as MIAME and MIAPE (Minimum Information About a Proteomics Experiment)[16] to support functional genomics investigations. We anticipate that MISFISHIE will be integrated with other MGED (Microarray and Gene Expression Data) Society standards[17], in particular through the Reporting Structure for Biological Investigation (RSBI) working group[18] and the Minimum Information for Biological and Biomedical Investigations (MIBBI) project[19]. Clearly, the goal of integrating different data types will be best served by a common reporting structure. Separation of the minimum information specification and the data format is important because in the data format there should be scope to provide unlimited further information beyond the minimum specification and there should be the ability to encode incomplete information for optimal flexibility. Furthermore, broad acceptance of a minimum information standard would greatly aid the design of a data model.

To facilitate data transfer between some existing expression databases, a MISFISHIE-compliant XML data format has been developed. A document type definition (DTD) was developed for three expression databases: ANISEED (**Table 1**), COMPARE[20] and 4DXpress[21]. Its format follows MISFISHIE. This DTD and an associated example are available at ANISEED (http://crfb.univ-mrs.fr/aniseed/exchange_format.php) and at COMPARE (http://compare.ibdml.univ-mrs.fr/exchange_format.php).

It has long been appreciated that improved standards for IHC are needed. However, the focus has been on developing standardized technical protocols that would produce more uniform staining[22] or on reducing subjectivity in interpreting histological sections[23]. MISFISHIE does not endorse standardized methodologies or data interpretation but rather seeks to promote complete disclosure of the methodologies used.

Guidelines for tumor-marker prognostic studies, known as REMARK[24], were recently established. REMARK encompasses outcome studies based on tumor markers of any kind, not just those of

Table 1 Organizations, standards and resources

ANISEED	Ascidian Network for *In situ* Expression and Embryological Data	http://crfb.univ-mrs.fr/aniseed/
	ANISEED MISFISHIE DTD	http://crfb.univ-mrs.fr/aniseed/exchange_format.php
BioImage	Image repository	http://www.bioimage.org
COMPARE	Multi organism information system: MISFISHIE DTD	http://compare.ibdml.univ-mrs.fr/exchange_format.php
EMAGE	Edinburgh Mouse Atlas Gene Expression database	http://genex.hgu.mrc.ac.uk
FuGE-OM	Functional Genomics Experiment Object Model	http://fuge.sourceforge.net
FuGO	Functional Genomics Ontology (renamed OBI in October 2006)	
GXD	Gene Expression Database	http://www.informatics.jax.org
MAGE-OM/ML	MicroArray Gene Expression Object Model/ Markup Language	http://www.mged.org/Workgroups/MAGE/mage.html
MGED	Microarray and Gene Expression Data Society	http://www.mged.org
MIAME	Minimum Information About a Microarray Experiment	http://www.mged.org/Workgroups/MIAME/miame.html
MIAPE	Minimum Information About a Proteomics Experiment	http://www.psidev.info/index.php?q=node/91
MISFISHIE	Minimum Information Specification For *In situ* Hybridization and Immunohistochemistry Experiments	http://www.mged.org/Workgroups/MISFISHIE
	Printable brief checklist	http://scgap.systemsbiology.net/standards/misfishie/MISFISHIE_Checklist_2007-10-28.xls
	Complete checklist	http://scgap.systemsbiology.net/standards/misfishie
MO	MGED Ontology	http://mged.sourceforge.net/ontologies/MGEDontology.php
MorphBank	Image repository	http://www.morphbank.net
OBI	Ontology for Biomedical Investigations (formerly FuGO)	http://obi.sourceforge.net
OBO	Ontologies for Biology Organization	http://obo.sourceforge.net
OME	Open Microscopy Environment	http://www.openmicroscopy.org
PEDRo	Proteomics Experiment Data Repository	http://pedro.man.ac.uk
RSBI	Reporting Structure for Biological Investigation	http://www.mged.org/Workgroups/rsbi/index.html
UMLS	Unified Medical Language System	http://www.nlm.gov/research/ulms/
ZFIN	Zebrafish Information Network	http://www.zfin.org

IHC. MISFISHIE encompasses nearly any study employing IHC or ISH regardless of context, such as a tumor-marker study or a zebrafish embryo study. We expect that requirements pertaining to specialized subdomains (for example, clinical prognostic studies) will be added to MISFISHIE in the future.

Existing databases for gene expression localization data provide a useful framework from which to build a specification. Two databases for the mouse research community, the Mouse Gene Expression Database (GXD)[25] and the Edinburgh Mouse Atlas Gene Expression (EMAGE) database[26], influenced the design of MISFISHIE. We replaced mouse-specific fields with more organism-neutral ones and eliminated fields deemed unnecessary. In these databases, many experiments entered by curators using information in journal articles have empty fields because the papers lacked sufficient detail. MISFISHIE-compliant publications will result in more complete database descriptions. Although MISFISHIE is primarily designed for peer-reviewed journal articles, it will guide database development as well. For example, the release of ANISEED version 3.0 is based on MISFISHIE rules, and the new schema of this database is MISFISHIE compliant. The inclusion of specific experimental details, such as tissue type, reagents and methods, will allow investigators to more efficiently find precedents for their experiments. For example, an investigator might rapidly search all publications that reported immunoperoxidase localization of membrane metallo-endopeptidase (CD10, MME) in the human prostate using a database and retrieve information on how the gene localization experiments were conducted.

Proposed guidelines

An abbreviated MISFISHIE checklist is provided in **Box 1**. A printable version is available at http://scgap.systemsbiology.net/standards/misfishie/MISFISHIE_Checklist_2007-10-28.xls. The complete checklist is available at http://scgap.systemsbiology.net/standards/misfishie. One example of real experimental data annotated according to MISFISHIE is given in **Supplementary Note 1** online; more examples accompany the complete checklist at the preceding URL.

The checklist covers six types of information; for each, the guiding principle is to supply enough information to allow the experiment to be reproduced.

1. Experimental design
2. Biomaterials (specimens) and treatments (section or whole-mount preparation)
3. Reporters (probes or antibodies)
4. Staining
5. Imaging data
6. Image characterizations

Ontologies, such as the MGED Ontology (MO)[4] or Ontology for Biomedical Investigations (OBI; formerly named FuGO)[27] are extremely advantageous as a source of descriptors because they facilitate computational searches of data. For terms outside the scope of OBI, such as those in anatomy, another appropriate ontology may be used. A good list of ontologies is maintained at the Ontologies for Biology Organization (OBO)

website (**Table 1**). Use of OBI and other ontologies will be especially important as MISFISHIE-supporting applications and databases are developed. Many of the terms used in this specification are already defined in OBI.

Experimental design. The experiment as a whole is described by the following:

A. Experiment description: the aims of the experiment.

B. Assay type(s): for example, IHC, ISH, lectin affinity histochemistry, cell lineage– or tissue-specific reporter expression.

C. Experiment design type: for example, comparisons of normal versus diseased tissue, of multiple tissue or embryo specimens of similar type, or of multiple probes or antibodies applied to the same tissue; a localization screen; etc. The MGED Ontology ExperimentDesignType has many entries categorizing design type.

D. Experimental factors: the parameters or conditions that are tested, such as probe or antibody, disease state, genetic variation, structural unit, age, etc. Again, the MGED Ontology is a rich source of terms for describing the factors being tested.

E. Total number of assays performed in the experiment: an assay is defined as one instance of a hybridization/stain of a single specimen with a single reporter. Thus, the result of a tissue microarray

---

Box 1 Checklist

This checklist should be used in conjunction with the full specification, not instead of it.

- Experiment design
  - Experiment description
  - Assay type(s) (IHC, ISH, GFP, etc.)
  - Experimental design (multiple reporter survey, specimen variation)
  - Experimental factors (variables in assays such as reporter or specimen, etc.)
  - Total number of assays performed
  - (optional) URL for more information
  - Contact information
- Biomaterials and treatments
  - Attributes of the individual (organism, sex, strain, line, developmental stage, age, etc.)
  - Physiologic state (e.g., normal versus disease)
  - Relevant exogenous factors (treatment, special diet, etc.)
  - Anatomic source of specimens
  - Provider of the specimens
  - Assay preparation protocol (enough to reproduce?)
- Reporter (probe or antibody) information
  - Unambiguous reporter identification, ideally genomic
  - Full sequence or clone ID of the reporters
  - Protocol for obtaining exact reporter (purchase from_____, create, etc.)
  - Other important attributes (mono- or polyclonal, generating organism, etc.)
- Staining protocols and parameters
  - Detection method (number of reporters, detection reagent and systems)
  - Staining protocol (enough to reproduce?)
  - Details about positive and negative controls
- Imaging data and parameters
  - The digital images for each assay (can download to your computer and explore?)
  - Detection method
  - (optional) Control images and imaging acquisition protocol
- Image characterizations
  - Definition of structural units (from ontology or manual definition)
  - Definition of intensity scale
  - Characterization of results in tabular form (digital or printed)
  - (optional) Characterization protocol

consisting of a 10×10 array of tissues would be counted as 100 assays. If replicates or reruns are a component of the experimental design, provide details that should include number of replicates per tissue, per reporter, etc.

F. URL of any websites or database accession numbers (if available) pertinent to the experiment.

G. Contact information for communicating with the experimenters.

Biomaterials (specimens) and treatments (section or whole-mount preparation). Describing specimens comprehensively is challenging, as they may have dozens or hundreds of characteristics. This is especially true for material from human subjects when clinical information is available. Characteristics that are known to differ among specimens should be provided with each specimen, whereas common attributes of all the specimens may be provided only once. The biological sample is described by the following:

A. Origin of the specimens.
   i. Attributes of the individual(s). The organism species must be named, preferably using the US National Center for Biotechnology Information (NCBI) taxonomy, and for non-human organisms the strain and mutant alleles should be named according to the accepted standards for that organism. Other attributes may include, but are not limited to, sex, age, developmental stage, genotype, and phenotype.
   ii. Physiologic state of the individual(s) (normal versus diseased).
   iii. Relevant exogenous factors (for example, treatment, special diet).
   iv. Anatomic source of the tissue or cell sample.
   v. Provider of the specimens.

The information necessary to reproduce the biomaterials is not limited to the above examples. Use of an ontology or controlled vocabulary is highly encouraged, although a standardized set of terms and a single, widely accepted ontology is not yet available. The rationale for providing specific structural detail is that the location of an object, such as a cell type that is being studied, may correlate with expression of a specific gene by that cell type. Structural detail may be important not only for cases where gene expression depends on tissue handling (for example, there is stronger labeling at the specimen edges) but also in cases where there is heterogeneity even within a single microanatomical unit (for example, in lung tumors, cell cycle regulatory genes are highest at the periphery)[28].

B. Manner of preparation of the specimens.
   i. Nature of the specimens (for example, whole tissue, whole mounts of tissue, tissue sections, thickness of sections, whole cells, or sections of cells).
   ii. Manner in which the specimens were prepared for the experiments (for example, fixed specimens, with type of fixative and duration of fixation; fresh, non-fixed, non-frozen specimens; or non-fixed, frozen specimens; sections mounted on slides versus sections floating in reagents).
   iii. Protocols used. Referencing previously published protocols is permissible if the protocols are appropriately detailed and were strictly followed.

Sensitivity of the immunoreaction of some gene products to fixation is exemplified by the observation that cyclin-dependent kinase inhibitor p27Kip1 (CDKN1B) was least frequent and least intense in prostate cancer cells that were farthest from the cut surface of a fixed tissue. These were the cells that were least rapidly fixed[29].

Reporters (probes or antibodies). Reporters (probes, lectins or antibodies) can differ in reactivity from lot to lot and from manufacturer to manufacturer. A manufacturer's literature usually provides most of the

needed information but may not be permanent. For privately produced reporters, enough information should be provided to enable another lab to generate them. Validation of reporters in the current literature is often poor. MISFISHIE does not at present require that researchers validate reporters, but such validation is encouraged and should be reported when performed.

A. Unambiguous genomic identification of each reporter:
   i. For *in situ* hybridizations, provide the corresponding GenBank/EMBL/DDBJ accession number and, if applicable, the start and end nucleotide positions of the probe within that sequence. Also, provide the accession number version or database release version.
   ii. For antibodies, provide the protein identifier, including specific version information for the accession number or database release.

B. Full sequence of each probe or clone number of each antibody. For fluorescent protein experiments, the promoter sequence should be specified. In each case, provide the method by which the reporter was characterized.
   i. If the sequence or clone number is not known, the template or clone must be made publicly available. Provide specific details on how the template or clone may be obtained.
   ii. In tissue localization experiments based on expression of a fluorescent protein reporter gene fused to the promoter of a gene of interest, what is important is not the sequence of the reporter but the sequence of the promoter, which confers cell and tissue specificity on the reporter.

C. Protocol(s) for how the reporters were designed and produced or the source from which they were obtained.
   i. For reporters purchased from a company, the company name, address, catalog number and lot number should be provided.
   ii. For a custom-made antibody, the putative antigen and references to studies that characterize the sensitivity and specificity of the antibody in tissue immunostains should be given.

D. Additional attributes of the reporter:
   i. For antibodies, the type of primary antibody (monoclonal or polyclonal), the immunoglobulin isotype and the organism in which the antibody was generated.
   ii. For lectins, the full name (for example, *Dolichos biflorus*), the source of the lectin (for example, which company produced it), how it was detected (for example, whether it was fluorescently labeled or biotinylated, with follow-up histochemical analysis), and how it was labeled (if, for example, the investigators labeled the lectin themselves, they should give source of the reagents, the method and/or the labeling kit).

Staining. Staining protocols vary considerably, and the merits of standardizing them have been discussed extensively in the literature.

A. Number of detectable reporters in the hybridization or stain (for example, more than one for multiple-dye fluorescence microscopy) and details about the detection method:
   i. Detection reagent (for example, fluors, enzyme substrates, gold particles).
   ii. Source of the detection system and description of the reaction.

B. Protocol to produce the hybridization or immunostain, including a description of how the tissue (organism, organ or section) was mounted onto the slide or substrate and treatments of the section (for example, IHC protocol inclusive of parameters such as buffer, temperature, post-wash conditions, etc.). Referencing previously published protocols is permissible if the protocols are appropriately detailed and were strictly followed. Also:
   i. What steps, if any, were taken to decrease nonspecific reaction product. For example, in immunoperoxidase experiments, the specimen

preparation may be preincubated with albumin solution to block nonspecific binding or with peroxide solution to block signal due to endogenous peroxidase.

   ii. Use of an antigen or gene product retrieval method.

C. Information about assay controls: the nature of both positive and negative tissue and reporter controls (or state if controls were not performed). The same level of detail for the tissue controls should be reported as for the cells or tissues that are being studied. Optionally, provide specificity reporter controls, such as competitive inhibition with either purified protein or peptide in IHC.

Imaging data. Although the MIAME specification stops short of requiring microarray image data, we propose that representative IHC or ISH images be provided, as interpretation of these images varies among observers. Images are not needed to reproduce an experiment, but they aid in the analysis. Both positive and negative results should be reported.

Repositories for images from gene expression localization experiments exist for several model organisms, such as GXD[25] and EMAGE[26] for mouse and ZFIN[30] for zebrafish, but not for human. A general, organism-independent database would be very valuable, as it could provide examples of tissue localization studies, serve as a reference site for verifying the tissue localizations of reporter reagents and provide accession numbers for publications. There are two projects that aim to provide these features. MorphBank is an available general purpose image repository for biological research. BioImage is an image repository under construction (ref. 31).

The information on imaging data should include:

A. Digital images for each assay in the study, digitally available for download without charge. Images should be of sufficient resolution to allow independent characterization and provided in a standard file format (for example, JPEG, PNG, GIF, TIFF). Images should be named or tagged with the reporter and specimen that they represent.

B. Detection method by which hybridization or staining is observed (for example, for each channel, fluorescence excitation and emission wavelengths if more than one reporter is used). If the detection method is the same for all images, it need only be mentioned once.

C. Images for the controls are optional.

Image characterizations. Interpretations of the results should be reported in categorized tabular format so that they can be easily stored in a database, queried and compared with other expression data. The following minimum requirements can be supplemented with further characterizations as needed.

A. Ontology entries, including reference to the ontology terms, accession numbers or terms and definitions if sufficient detail cannot be found in an existing ontology[32–35] for individual structural units used for classification. (Note that some ontologies, such as the College of American Pathologists' Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) and the Unified Medical Language System (UMLS) of the US National Institutes of Health's National Library of Medicine, may contain licensing restrictions that make them unavailable to some or limit the use of the terms; a MISFISHIE-compliant document that contains SNOMED CT entries or some UMLS entries may not be legally redistributable[36]). Structural units include organ, tissue, cell and subcellular component. List and characterize only the relevant structural units and not those that are visible in assays or slides but irrelevant.

B. Intensity scale, ideally one from the MGED Ontology. For example, a three-level scale of present, absent or equivocal might be appropriate for evaluating IHC stains. However, any appropriate scale may be used as long as each gradation of intensity in the scale is defined.

C. For each relevant structural unit in each assay or image:

   i. Staining intensity or the fraction of the structural unit's population showing each intensity (see example below).

   ii. Other optional annotations or characterizations of the structural unit: for example, feature density, qualitative characteristics or spatial distribution of the structural unit or staining. The use of referenced ontology terms is encouraged.

Both positive and negative calls of staining relevant to the experiment should be reported. A negative result is an upper limit to the expression level, where the limit is usually not well known. If some structural units cannot be characterized for some reporters, corresponding calls may be null. For example:

Luminal epithelial cell: present
Basal epithelial cell: absent

is sufficient; or, when appropriate, more detail:

Luminal epithelial cell: 90% present, 10% equivocal, 0% absent
Basal epithelial cell: 10% present, 10% equivocal, 80% absent
Unless only a few expression calls are presented, it is clearest if the calls are presented in tabular form.

D. Optionally, the protocol for the characterization and information about the basic characterization technique. For example, how many observers performed the characterizations, whether the characterizations were performed from the images themselves or visually through the instrument and any exceptions or assumptions made in characterizing the data. One example of a well described characterization protocol may be found in ref. 37.

N	%	Statistic
		Table 2 Summary of statistics from the MISFISHIE assessment survey of a cohort of selected current literature
32	100%	Number of articles assessed for compliance
4	13%	Number of articles considered to be fully MISFISHIE compliant
9	28%	Number of articles for which MISFISHIE information is missing for one section
10	31%	Number of articles for which MISFISHIE information is missing for two sections
6	19%	Number of articles for which MISFISHIE information is missing for more than two sections
31	97%	Number of articles that meet the data content requirements for section 1 (experimental design)
29	91%	Number of articles that meet the data content requirements for section 2 (biomaterials and treatments)
24	75%	Number of articles that meet the data content requirements for section 3 (reporters)
24	75%	Number of articles that meet the data content requirements for section 4 (staining)
5	16%	Number of articles that meet the data content requirements for section 5 (imaging data)
15	47%	Number of articles that meet the data content requirements for section 6 (image characterizations)

We also note that the use of digital images may have advantages in terms of replication and decreased intra- and interobserver variability[38].

Survey of the recent literature

To compare MISFISHIE with current publication practice, we examined articles reporting on IHC or ISH in the last 7 years. Three articles[39–41] were assessed and discussed by all ten *ad hoc* reviewers to minimize inter-reviewer variability. Another 29 articles[42–69] were assigned to individual reviewers. Each reviewer assessed articles as if reviewing a manuscript submitted to a journal that required MISFISHIE compliance. Compliance for each MISFISHIE subsection was rated on a scale of 0 to 10, where a 10 indicates inclusion of all information needed to understand or reproduce the experiment without making any assumptions. Scores of 8 and 9 were considered a low pass; the reviewer could reproduce the experiment with a few assumptions. One example of a paper deemed MISFISHIE compliant is ref. 64.
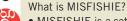
Of the 32 papers assessed, four (13%) were deemed MISFISHIE compliant in all six sections (**Table 2**). Another 28% were out of compliance with only one section, and 31% did not comply in two sections. More than 90% complied with sections 1 and 2 (experimental design; biomaterials and treatments). Compliance for sections 3 and 4 (reporters; staining) was ~75%. Section 5 (imaging data) proved to be the most troublesome, with only 16% of the articles compliant. Finally, ~47% complied with section 6 (image characterizations). The reviewers felt that the majority of noncompliant papers would require only modest additions to become compliant, with the possible exception of section 5. This section requires that at least one representative image of each assay be made electronically available in a model organism database, a generic image database, a journal's supplemental data web site or even the author's web site (the least preferable option).

Conclusion

MISFISHIE was developed by the Stem Cell Genome Anatomy Projects consortium of the US National Institutes of Health National Institute of Diabetes & Digestive & Kidney Diseases to facilitate data sharing within the consortium and was discussed with members of the larger research community. The history of the creation of MISFISHIE and the lessons learned from it[70] may be helpful to others aiming to create similar guidelines for other data types. We expect that MISFISHIE will be updated as other localization methods, such as DNA *in situ* hybridization to chromosomes, are implemented. There is still considerable room for researching the scientific best practice for performing and reporting these types of studies, and the eventual accepted specification will be achieved through discussion and consensus. Suggestions from the community are actively encouraged and will be collected and incorporated into an eventual second release, published at the MISFISHIE domain of the MGED web site: http://www.mged.org/Workgroups/MISFISHIE. Comments may be addressed to the email distribution list dedicated to discussion about MISFISHIE: mged-misfishie@lists.sourceforge.net. Some frequently asked questions and answers are listed in **Box 2**. After a suitable period of dialog and revision by the community, and if there is widespread acceptance by the community, we would encourage reviewers, journal editors and funding agencies to promote compliance with MISFISHIE.

Our survey of recent articles indicated that only ~15% of published works are fully compliant and that most fail by not making images of assays used in the study digitally accessible to the research community. Most of the surveyed papers could be brought into compliance by uploading the images into a repository and adding fewer than a dozen more sentences of description. If article length constraints hinder MISFISHIE compliance, the required information could be provided in supplementary information. Several of the model organism databases, including GXD[25] and EMAGE[26] for mouse and ZFIN[30] for zebrafish, are already able to accept and archive the results from a publication

---

Box 2 Frequently asked questions about MISFISHIE

What is MISFISHIE?
- MISFISHIE is a set of guidelines for reporting the relevant materials, methods and results of a gene expression localization experiment in a way that allows a researcher from a different lab to validate the findings.

When should the MISFISHIE guidelines be applied?
- Authors of a manuscript should include the information requested in MISFISHIE.
- Reviewers should use MISFISHIE to ensure that a manuscript describes the materials, methods and results adequately.

Does MISFISHIE provide guidelines for how I should perform my experiment?
- No, MISFISHIE by design does not promote any particular technique or best practice for carrying out or analyzing experiments. It merely provides guidelines for disclosing what was done.

What data format should be used to encode MISFISHIE-compliant information?
- At present, it is sufficient to encode this information in free text in a manuscript. Data formats that allow the encoding of MISFISHIE-compliant information in a machine-readable format are becoming available and might be required for submission to a database.

Why are there so many detailed requirements?
- The experiments can be complex and sensitive to the materials and methods used. These variables must be explicitly described to validate or reproduce the results.

Why does MISFISHIE not have more detailed requirements?
- Only those requirements deemed truly minimum for all gene expression localization experiments are included. If some critical aspect of the experiment at hand is not listed in MISFISHIE, it should nonetheless be reported; MISFISHIE should not be construed as a maximum information specification.

that provides all information that MISFISHIE specifies. We encourage authors to submit their data to these databases upon submission of the manuscript for publication.

Note: Supplementary information is available on the Nature Biotechnology website.

1. True, L.D. Quantitative immunohistochemistry: a new tool for surgical pathology? *Am. J. Clin. Pathol.* **90**, 324–325 (1988).
2. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).
3. Spellman, P.T. *et al.* Design and implementation of Microarray Gene Expression Markup Language (MAGE-ML). *Genome Biol.* **3**, RESEARCH0046 (2002).
4. Stoeckert, C.J. & Parkinson, H. The MGED ontology: a framework for describing functional genomics experiments. *Comp. Funct. Genomics* **4**, 127–132 (2003).
5. Taylor, C.F. *et al.* A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nat. Biotechnol.* **21**, 247–254 (2003).
6. Garwood, K. *et al.* PEDRo: a database for storing, searching and disseminating experimental proteomics data. *BMC Genomics* **5**, 68 (2004).
7. Jones, A., Hunt, E., Wastling, J.M., Pizarro, A. & Stoeckert, C.J. Jr. An object model and database for functional genomics. *Bioinformatics* **20**, 1583–1590 (2004).
8. Xirasagar, S. *et al.* CEBS object model for systems biology data, SysBio-OM. *Bioinformatics* **20**, 2004–2015 (2004).
9. Jenkins, H. *et al.* A proposed framework for the description of plant metabolomics experiments and their results. *Nat. Biotechnol.* **22**, 1601–1606 (2004).
10. Lindon, J.C. *et al.* Summary recommendations for standardization and reporting of metabolic analyses. *Nat. Biotechnol.* **23**, 833–838 (2005).
11. Berman, J.J., Edgerton, M.E. & Friedman, B.A. The tissue microarray data exchange specification: a community-based, open source tool for sharing tissue microarray data. *BMC Med. Inform. Decis. Mak.* **3**, 5 (2003).
12. Stoeckert, C.J., Quackenbush, J., Brazma, A. & Ball, C.A. Minimum information about a functional genomics experiment: the state of microarray standards and their extension to other technologies. *Drug Discov. Today Targets* **3**, 159–164 (2004).
13. Jones, A.R. *et al.* The Functional Genomics Experiment model (FuGE): an extensible framework for standards in functional genomics. *Nat. Biotechnol.* **25**, 1127–1133 (2007).
14. Rayner, T.F. *et al.* A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC Bioinformatics* **7**, 489 (2006).
15. Brazma, A., Krestyaninova, M. & Sarkans, U. Standards for systems biology. *Nat. Rev. Genet.* **7**, 593–605 (2006).
16. Taylor, C.F. *et al.* HUPO — Proteomics Standards Initiative (PSI). *OMICS* **10**, 145–151 (2006).
17. Ball, C.A. & Brazma, A. MGED standards. *OMICS* **10**, 138–144 (2006).
18. Sansone, S.-A. *et al.* A strategy capitalizing on synergies: the Reporting Structure for Biological Investigation (RSBI) working group. *OMICS* **10**, 164–171 (2006).
19. Taylor, C.F. *et al.* Promoting coherent minimum reporting requirements for biological and biomedical investigations: the MIBBI Project. *Nat. Biotechnol.* (in the press).
20. Salgado, D., Gimenez, G., Coulier, F. & Marcelle, C. COMPARE, a multi-organism system for cross-species data comparison and transfer of information. *Bioinformatics*, published online 1 December 2007 (doi:10.1093/bioinformatics/btm599).
21. Haudry, Y. *et al.* 4DXpress: a database for cross-species expression pattern comparisons. *Nucleic Acids Res.* **36** (database issue), D847-D853 (2007).
22. Swanson, P.E. Methodologic standardization in immunohistochemistry: a doorway opens. *Appl. Immunohistochem.* **1**, 229–231 (1993).
23. Taylor, C.R. An exaltation of experts: concerted efforts in the standardization of immunohistochemistry. *Hum. Pathol.* **25**, 2–11 (1994).
24. McShane, L.M. *et al.* Reporting recommendations for tumor marker prognostic studies (REMARK). *J. Natl. Cancer Inst.* **97**, 1180–1184 (2005).
25. Smith, C.M. *et al.* The mouse Gene Expression Database (GXD): 2007 update. *Nucleic Acids Res.* **35** (database issue), D618–D623 (2007).
26. Baldock, R.A. *et al.* EMAP and EMAGE: a framework for understanding spatially organized data. *Neuroinformatics* **1**, 309–325 (2003).
27. Whetzel, P.L. *et al.* Development of FuGO – an ontology for functional genomics experiments. *OMICS* **10**, 199–204 (2006).
28. Dobashi, Y. *et al.* Active cyclin A-CDK2 complex, a possible critical factor for cell proliferation in human primary lung carcinomas. *Am. J. Pathol.* **153**, 963–972 (1998).
29. De Marzo, A.M., Fedor, H.H., Gage, W.R. & Rubin, M.A. Inadequate formalin fixation decreases reliability of p27 immunohistochemical staining: probing optimal fixation time using high-density tissue microarrays. *Hum. Pathol.* **33**, 756–760 (2002).
30. Sprague, J. *et al.* The Zebrafish Information Network (ZFIN): the zebrafish model organism database. *Nucleic Acids Res.* **31**, 241–243 (2003).
31. Carazo, J.M. & Stelzer, E.H. The BioImage Database Project: organizing multidimensional biological images in an object-relational database. *J. Struct. Biol.* **125**, 97–102 (1999).
32. Rosse, C. & Mejino, J.L. Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J. Biomed. Inform.* **36**, 478–500 (2003).
33. Bard, J., Rhee, S.Y. & Ashburner, M. An ontology for cell types. *Genome Biol.* **6**, R21 (2005).
34. Bard, J.L. *et al.* An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature. *Mech. Dev.* **74**, 111–120 (1998).
35. Hayamizu, T.F., Mangan, M., Corradi, J.P., Kadin, J.A. & Ringwald, M. The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biol.* **6**, R29 (2005).
36. Berman, J.J. A tool for sharing annotated research data: the "Category 0" UMLS (Unified Medical Language System) vocabularies. *BMC Med. Inform. Decis. Mak.* **3**, 6 (2003).
37. Abd El-Rehim, D.M. *et al.* Expression of luminal and basal cytokeratins in human breast carcinoma. *J. Pathol.* **203**, 661–671 (2004).
38. Bova, G.S. *et al.* Web-based tissue microarray image data analysis: initial validation testing through prostate cancer Gleason grading. *Hum. Pathol.* **32**, 417–427 (2001).
39. Liu, A.Y. & True, L.D. Characterization of prostate cell types by CD cell surface molecules. *Am. J. Pathol.* **160**, 37–43 (2002).
40. Kernek, K.M. *et al.* Fluorescence *in situ* hybridization analysis of chromosome 12p in paraffin-embedded tissue is useful for establishing germ cell origin of metastatic tumors. *Mod. Pathol.* **17**, 1309–1313 (2004).
41. McKenney, J.K. *et al.* Basal cell proliferations of the prostate other than usual basal cell hyperplasia: a clinicopathologic study of 23 cases, including four carcinomas, with a proposed classification. *Am. J. Surg. Pathol.* **28**, 1289–1298 (2004).
42. Amara, N. *et al.* Prostate stem cell antigen is overexpressed in human transitional cell carcinoma. *Cancer Res.* **61**, 4660–4665 (2001).
43. Ayala, G. *et al.* High levels of phosphorylated form of Akt-1 in prostate cancer and non-neoplastic prostate tissues are strong predictors of biochemical recurrence. *Clin. Cancer Res.* **10**, 6572–6578 (2004).
44. Bart, J. *et al.* The distribution of drug-efflux pumps, P-gp, BCRP, MRP1 and MRP2, in the normal blood-testis barrier and in primary testicular tumours. *Eur. J. Cancer* **40**, 2064–2070 (2004).
45. Browne, T.J. *et al.* Prospective evaluation of AMACR (P504S) and basal cell markers in the assessment of routine prostate needle biopsy specimens. *Hum. Pathol.* **35**, 1462–1468 (2004).
46. Chen, D. *et al.* Syndecan-1 expression in locally invasive and metastatic prostate cancer. *Urology* **63**, 402–407 (2004).
47. Clayton, H., Titley, I. & Vivanco, M. Growth and differentiation of progenitor/stem cells derived from the human mammary gland. *Exp. Cell Res.* **297**, 444–460 (2004).
48. Cooray, H.C., Blackmore, C.G., Maskell, L. & Barrand, M.A. Localisation of breast cancer resistance protein in microvessel endothelium of human brain. *Neuroreport* **13**, 2059–2063 (2002).
49. Giangreco, A., Shen, H., Reynolds, S.D. & Stripp, B.R. Molecular phenotype of airway side population cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **286**, L624–L630 (2004).
50. Gmyrek, G.A. *et al.* Normal and malignant prostate epithelial cells differ in their response to hepatocyte growth factor/scatter factor. *Am. J. Pathol.* **159**, 579–590 (2001).
51. Hwang, J.H. *et al.* Isolation of muscle derived stem cells from rat and its smooth muscle differentiation. *Mol. Cells* **17**, 57–61 (2004); erratum **17**, 381 (2004).
52. Jonker, J.W. *et al.* The breast cancer resistance protein BCRP (ABCG2) concentrates drugs and carcinogenic xenotoxins into milk. *Nat. Med.* **11**, 127–129 (2005).
53. Knudsen, B.S. *et al.* High expression of the Met receptor in prostate cancer metastasis to bone. *Urology* **60**, 1113–1117 (2002).
54. Larkin, A. *et al.* Investigation of MRP-1 protein and MDR-1 P-glycoprotein expression in invasive breast cancer: a prognostic study. *Int. J. Cancer* **112**, 286–294 (2004).
55. Lee, K., Klein-Szanto, A.J. & Kruh, G.D. Analysis of the MRP4 drug resistance profile in transfected NIH3T3 cells. *J. Natl. Cancer Inst.* **92**, 1934–1940 (2000).
56. Li, R. *et al.* High level of androgen receptor is associated with aggressive clinicopathologic features and decreased biochemical recurrence-free survival in prostate: cancer patients treated with radical prostatectomy. *Am. J. Surg. Pathol.* **28**, 928–934 (2004).
57. Martin, C.M. *et al.* Persistent expression of the ATP-binding cassette transporter, Abcg2, identifies cardiac SP cells in the developing and adult heart. *Dev. Biol.* **265**, 262–275 (2004).
58. Martin, M.J., Muotri, A., Gage, F. & Varki, A. Human embryonic stem cells express an immunogenic nonhuman sialic acid. *Nat. Med.* **11**, 228–232 (2005).
59. Master, V.A., Wei, G., Liu, W. & Baskin, L.S. Urothelium facilitates the recruitment and trans-differentiation of fibroblasts into smooth muscle in acellular matrix. *J. Urol.* **170**, 1628–1632 (2003).
60. Piotrowska, A.P. *et al.* Alterations in smooth muscle contractile and cytoskeleton proteins and interstitial cells of Cajal in megacystis microcolon intestinal hypoperistalsis syndrome. *J. Pediatr. Surg.* **38**, 749–755 (2003).
61. Ricciardelli, C. *et al.* Androgen receptor levels in prostate cancer epithelial and peritumoral stromal cells identify non-organ confined disease. *Prostate* **63**, 19–28 (2005).

62. Roudier, M.P. *et al.* Phenotypic heterogeneity of end-stage prostate carcinoma metastatic to bone. *Hum. Pathol.* **34**, 646–653 (2003).

63. Rubin, M.A. *et al.* Quantitative determination of expression of the prostate cancer protein alpha-methylacyl-CoA racemase using automated quantitative analysis (AQUA): a novel paradigm for automated and continuous biomarker measurements. *Am. J. Pathol.* **164**, 831–840 (2004).

64. Santagata, S. *et al.* JAGGED1 expression is associated with prostate cancer metastasis and recurrence. *Cancer Res.* **64**, 6854–6857 (2004).

65. Scotlandi, K. *et al.* C-kit receptor expression in Ewing's sarcoma: lack of prognostic value but therapeutic targeting opportunities in appropriate conditions. *J. Clin. Oncol.* **21**, 1952–1960 (2003).

66. Shah, R.B. *et al.* Androgen-independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program. *Cancer Res.* **64**, 9209–9216 (2004).

67. St Croix, B. *et al.* Genes expressed in human tumor endothelium. *Science* **289**, 1197–1202 (2000).

68. Wang, Z. *et al.* Expression of the human cachexia-associated protein (HCAP) in prostate cancer and in a prostate cancer animal model of cachexia. *Int. J. Cancer* **105**, 123–129 (2003).

69. Zhigang, Z. & Wenlv, S. Prostate stem cell antigen (PSCA) expression in human prostate cancer tissues: implications for prostate carcinogenesis and progression of prostate cancer. *Jpn. J. Clin. Oncol.* **34**, 414–419 (2004).

70. Deutsch, E.W. *et al.* Minimum Information Specification For *In Situ* Hybridization and Immunohistochemistry Experiments (MISFISHIE). *OMICS* **10**, 205–208 (2006).

Affiliations:

[1]Institute for Systems Biology, 1441 N 34th Street, Seattle, Washington 98103, USA. [2]Department of Biochemistry, Stanford University School of Medicine, 279 Campus Drive West, Stanford, California 94305, USA. [3]Association for Pathology Informatics, 9650 Rockville Pike, Bethesda, Maryland 20814, USA. [4]Johns Hopkins University School of Medicine, PELICAN Laboratory, Departments of Pathology, Health Information Sciences, Genetic Medicine, Oncology, and Urology, Baltimore, Maryland, 21287 USA .[5]European Molecular Biology Laboratory (EMBL)–European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [6]Department of Microbiology, University of Washington, Seattle, Washington 98195, USA. [7]Medical Research Council (MRC) Clinical Sciences Centre Microarray Centre, Imperial College School of Medicine, London W12 ONN, UK. [8]MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK. [9]Institut de Biologie du Développement de Marseille, UMR 6216, Centre National de la Recherche Scientifique (CNRS)/Université de la Méditerranée, Parc Scientifique de Luminy, Case 907, F-13288 Marseille Cedex 9, France. [10]Department of Urology, University of Washington, Seattle, Washington 98195, USA. [11]Institute for Molecular Bioscience, University of Queensland, St. Lucia, Queensland 4027, Australia. [12]EMBL Heidelberg, Meyerhofstrasse 1, D-69117 Heidelberg, Germany. [13]Institute for Medical Genetics, Charité-Campus Benjamin Franklin, and Department of Developmental Genetics, Max-Planck-Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany. [14]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, USA. [15]CNRS UMR 8080, Université Paris-Sud, 91405 Orsay, France. [16]Dana-Farber Cancer Institute, 44 Binney Street, M232, Boston, Massachusetts 02115, USA. [17]The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA. [18]Department of Genetics, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305-5120, USA. [19]Center for Bioinformatics and Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [20]Division of Gene Regulation and Expression, Wellcome Trust Biocentre, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland, UK. [21]Computational Biology & Bioinformatics Group, Pacific Northwest National Laboratory, PO Box 999, MS K7-90, Richland, Washington 99352, USA. [22]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 IHH, UK. [23]Division of Developmental Neurobiology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK. [24]Stem Cell Program, Division of Hematology/Oncology, Children's Hospital Boston and Harvard Medical School, 300 Longwood Ave., Boston, Massachusetts 02115, USA. [25]Department of Pathology, University of Washington, Seattle, Washington 98195-6100, USA.