Sharing and archiving nucleic acid structure mapping data

Philippe Rocca-Serra, Stanislav Bellaousov, Amanda Birmingham, et al.

RNA 2011 17: 1204-1212 originally published online May 24, 2011
Access the most recent version at doi:10.1261/rna.2753211

Sharing and archiving nucleic acid structure mapping data

PHILIPPE ROCCA-SERRA,[1] STANISLAV BELLAOUSOV,[2] AMANDA BIRMINGHAM,[3] CHUNXIA CHEN,[4] PABLO CORDERO,[5] RHIJU DAS,[5] LAUREN DAVIS-NEULANDER,[4] CAIA D.S. DUNCAN,[6] MATTHEW HALVORSEN,[4] ROB KNIGHT,[7,8] NEOCLES B. LEONTIS,[9] DAVID H. MATHEWS,[2] JUSTIN RITZ,[4] JESSE STOMBAUGH,[7] KEVIN M. WEEKS,[6] CRAIG L. ZIRBEL,[10] and ALAIN LAEDERACH[4,11]

[1]Oxford e-Research Center, University of Oxford, OX1 3QG, Oxford, United Kingdom
[2]Department of Biochemistry and Biophysics and Center for RNA Biology, University of Rochester, Rochester, New York 14642, USA
[3]Thermo Fisher Scientific, Lafayette, Colorado 80026, USA
[4]Biology Department, University of North Carolina, Chapel Hill, North Carolina 27599-3290, USA
[5]Biochemistry Department, Stanford University, Stanford, California 94305, USA
[6]Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290, USA
[7]Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA
[8]Howard Hughes Medical Institute, Boulder, Colorado 80309, USA
[9]Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, USA
[10]Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio 43403, USA

ABSTRACT

Nucleic acids are particularly amenable to structural characterization using chemical and enzymatic probes. Each individual structure mapping experiment reveals specific information about the structure and/or dynamics of the nucleic acid. Currently, there is no simple approach for making these data publically available in a standardized format. We therefore developed a standard for reporting the results of single nucleotide resolution nucleic acid structure mapping experiments, or SNRNASMs. We propose a schema for sharing nucleic acid chemical probing data that uses generic public servers for storing, retrieving, and searching the data. We have also developed a consistent nomenclature (ontology) within the Ontology of Biomedical Investigations (OBI), which provides unique identifiers (termed persistent URLs, or PURLs) for classifying the data. Links to standardized data sets shared using our proposed format along with a tutorial and links to templates can be found at http://snrnasm.bio.unc.edu.

Keywords: RNA structure; chemical mapping; secondary structure

INTRODUCTION

Fields in which data standardization has allowed sharing among many researchers, including sequence data in GenBank (Benson et al. 2008; Wheeler et al. 2008) and structural data in the Protein Data Bank (Bernstein et al. 1977), have benefited enormously from the ability of investigators to draw insights from the work of thousands of people dispersed across the globe (Cannone et al. 2002; Griffiths-Jones et al. 2003; Noy et al. 2003; Zhang et al. 2006; Elnitski et al. 2007; Musen et al. 2008; Brown et al. 2009). At present, there is currently no standard database for archiving and sharing nucleic acid structure mapping data, despite the compelling opportunities to incorporate such data in studies with direct relevance to human health and to a wide range of scientific challenges (Russell and Herschlag 2001; Tullius 2002; Schroeder et al. 2004; Takamoto et al. 2004; Thirumalai and Hyeon 2005; Mortimer and Weeks 2007; Tijerina et al. 2007; Shcherbakova and Brenowitz 2008; Woodson 2008; Deigan et al. 2009). Chemical and enzymatic structure mapping techniques are useful in the field of nucleic acids and are commonly used to experimentally validate and/or constrain structural predictions, "footprint" protein-binding sites, and characterize folding reactions both kinetically and thermodynamically (Mathews et al. 2004; Deigan et al. 2009; Quarrier et al. 2010; Weeks 2010). Recent developments allowing the analysis of chemical mapping reactions in a quantitative and high-throughput manner yield large amounts of high-quality data that require automated processing and annotation (Das et al. 2005; Laederach et al. 2008; Mitra et al. 2008; Vasa et al. 2008; Wilkinson et al. 2008; Deigan et al. 2009; Watts et al. 2009; Underwood et al. 2010).

A standardized approach for making such data available upon publication is needed to facilitate sharing and wider

[11]**Corresponding author.**
E-mail alain@unc.edu.
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.2753211.

dissemination of these results. Figure 1A illustrates the unfortunately common scenario in our laboratories when structure-mapping data are collected. A laboratory colleague carefully collects data and meticulously records this work in a laboratory notebook. The data are then analyzed and published in a thesis and a scientific journal as an detailed multicolored diagram. Upon graduation, the thesis and data are often misplaced (Fogarty 2002). As a result, the primary data are lost and any attempt to reanalyze the data in a new context requires manually extracting data from a manuscript figure or from a PDF file in a manuscript supplement. In this work, we seek to advocate for an alternative scenario that greatly diminishes the risk of data loss and provides the data in a computer readable format (Fig. 1B).

We consider here the distinct types of structure mapping data and organize them into an ontology that reveals the relationships among various techniques. We then describe a system that both allows diverse users to integrate their nucleic acid probing data and facilitates the description of new techniques as they are developed. This systematization of knowledge and data will thus facilitate comparisons among methods, meta-analyses combining many independent lines of evidence about nucleic acid structure, and automated retrieval of nucleic acids for which good structural data are available.

APPROACH

Classification of SNRNASM assays

An important first step in sharing data efficiently is accurately defining the vocabulary used to describe an experiment. This is particularly important if one of the goals of sharing data is to facilitate meta-analyses using automated tools. Ontologies are commonly used to define terms and the relations between them in a precise way (Noy et al. 2003; Leontis et al. 2006; Brown et al. 2009). We therefore describe single nucleotide resolution nucleic acid structure mapping (SNRNASM) experiments in terms of an ontological framework. We note that the use of the idiosyncratic term SNRNASM is intentional. This term is unique to our approach for archiving nucleic acid probing data and will make it readily Internet searchable.

We have added terms to the Ontology of Biomedical Investigations (OBI) for 23 types of SNRNASM assays (Brinkman et al. 2010). We chose to include terms describing SNRNASMs into OBI, which focuses specifically on describing assays like structure mapping. We define two types of SNRNASMs, chemical and enzymatic mapping (Fig. 2). These two terms have corresponding OBI identifiers, OBI:0001017 and OBI:0001014, respectively (Fig. 2). The lines in Figure 2 represent "is a" relationships between terms. One can therefore infer from our ontological classification that, for example, RNase T1 structure mapping is an enzymatic-mapping assay, which is also a SNRNASM. Although this may seem obvious to those familiar with the field of RNA structure mapping, in the larger context of integrating multiple data sets for meta-analyses, it is essential to identify these elementary relationships explicitly. This strategy greatly facilitates the implementation of automated data meta-analyses algorithms (Leontis et al. 2006; Whetzel et al. 2006; Moreira and Musen 2007).

Our ontological classification of SNRNASMs also captures the fact the chemical and enzymatic structure mapping experiments almost always use a specific probe, which is generally an enzyme or chemical compound. For this reason, we have defined the "input" of structure-mapping assays as the chemical or enzyme reagent used to probe the nucleic acid (Table 1, Specific Input column). Furthermore, we explicitly identify these chemicals and enzymes in their respective ontologies, Chemical Entities of Biological Interest (CHEBI) (Degtyarenko et al. 2008, 2009), and Protein Ontology (PRO) (Natale et al. 2007, 2011). Additionally, for each entry, we have provided alternative names (for example, NMIA structure mapping is commonly known as SHAPE), and corresponding primary references. The SNRNASM classification is thus integrated into the larger ontological framework being developed for genomic annotations (Natale et al. 2007, 2011; Degtyarenko et al. 2008, 2009).

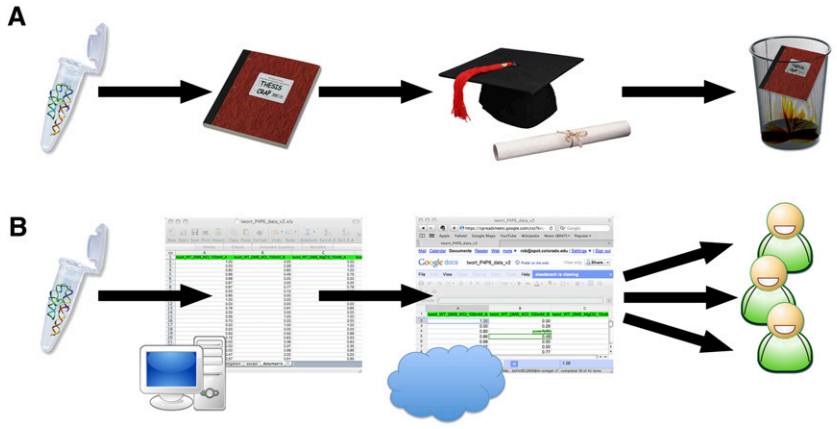

FIGURE 1. Different possible scenarios for SNRNASM (single nucleotide resolution nucleic acid structure mapping) data. (*A*) RNA chemical probing data is collected, recorded in a laboratory notebook, and published in a manuscript as an elaborate, colorful figure. This allows the graduate student who collected the data to graduate. Unfortunately, the raw data, meticulously recorded in the laboratory notebook, becomes lost (Fogarty 2002). (*B*) Proposed alternative in which the data is stored in a computer, uploaded to a publicly available server (in the cloud), and made downloadable upon publication of the manuscript, allowing other investigators access for new analyses.

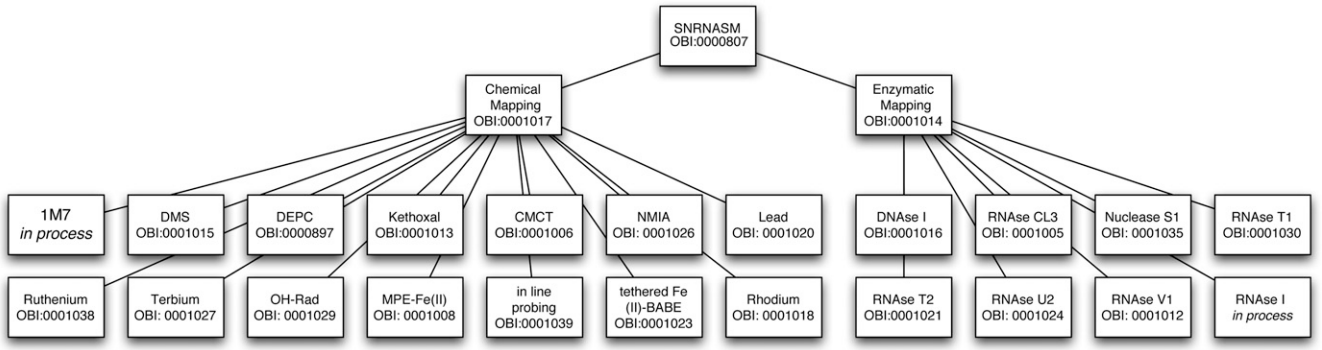

FIGURE 2. Graphical representation of the different terms we added to OBI (the Ontology of Biomedical Investigations) used to describe single nucleotide resolution nucleic acid structure mapping (SNRNASM) experiments. Each term is assigned a unique identifier (e.g., OBI:0001014) and organized by a series of hierarchical relationships. We used "is a" relationships in this case. For example DNAse I structure mapping (OBI:0001016) "is an" enzymatic mapping (OBI:0001014) experiment, which "is an" SNRNASM (OBI:0000807). In organizing our description of structure mapping experiments in this way, it becomes possible to design algorithms that will automatically identify relationships between different data sets.

Accessing SNRNASM classifications

The field of nucleic acid structure mapping is continuously evolving as new chemical and enzymatic probes are developed (Wilkinson et al. 2006; Mortimer and Weeks 2007; Regulski and Breaker 2008; Shcherbakova and Brenowitz 2008). It is therefore important that any effort to classify these experimental techniques also evolve to reflect the field accurately. All SNRNASM terms to date have been submitted to OBI, and are therefore accessible in OWL and OBO format (Moreira and Musen 2007) from http://obi-ontology.org. Practically, the annotations are easily visualized and edited in an ontology editor such as Protégé (Noy et al. 2003; Supplemental Fig. 1). New annotations from the community can be readily added and will appear in new OBI releases. For example, some annotations (e.g., RNase I) are "in process" and will therefore be added to OBI after the publication of this manuscript (Fig. 2).

To facilitate access to our SNRNASM classification we have developed a series of spreadsheets that provide a straightforward framework for annotating a chemical or enzymatic mapping experiment. Links to these spreadsheets can be found at http://snrnasm.bio.unc.edu, which are hosted in the "cloud," currently Google Docs servers. "Cloud" servers can be any publically available computer designed to store and disseminate data. By placing these documents in the cloud, no single lab is responsible for hosting these files on their servers, and anyone can edit an archived file if necessary. Our goal is to facilitate community involvement in the annotation process and to enable the groups developing new structure-mapping techniques to specify the terms that best describe their techniques.

Data sharing using the ISA-Tab format

The ontological framework we described above allows us to define structure-mapping experiments precisely. From a practical perspective, by associating an OBI term with each type of structure-mapping experiment (Fig. 2), it is possible to specify uniquely the type of experiment that was carried out on a nucleic acid. Although this represents a significant advantage in terms of being able to search for specific data sets, additional experimental information is required to be able to compare data sets effectively. For example, experimental conditions such as monovalent and divalent ion concentrations significantly affect RNA folding; it is essential to specify these conditions when undertaking comparative data analysis (Deras et al. 2000; Heilman-Miller et al. 2001a; Uchida et al. 2003; Das et al. 2008; Quarrier et al. 2010). Furthermore, SNRNASM data can be collected in different ways (using direct labeling of the RNA and gel electrophoresis, or reverse transcription followed by cDNA fragment analysis on a capillary sequencer (Mitra et al. 2008; Vasa et al. 2008). It is therefore important to capture, at minimum, the defining characteristics of experimental details in an annotation.

Defining best practices for experimental annotation of data is a nuanced challenge (Griffiths-Jones et al. 2005; Whetzel et al. 2006; Brown et al. 2009). On one hand, capturing as much detail as possible is ideal from a future analysis perspective. However, excessive annotation requirements are burdensome for the individual trying to share data, and can significantly decrease the overall amount of data shared. We therefore chose to require minimal annotations and developed a flexible format for sharing data that allows the user to decide which annotations to provide. Furthermore, we use a simple format that is easily edited in a spreadsheet program (including Excel and OpenOffice). We base our standard on the ISA-Tab (Investigation/Study/Assay) format, which is sufficiently extensible to allow easy SNRNASM annotation, is well established, and is widely used for biomedical data sets (Rocca-Serra et al. 2010).

The SNRNASM ISA-Tab format is based on the concept of a multi-tabular spreadsheet. It includes three

TABLE 1. SNRNASM assays currently in OBI along with their corresponding inputs (probes)

Assay name	Alternative name	Specific input (probe)	Input identifier	Reference
DMS structure mapping assay		Dimethyl sulfate	CHEBI:59050	(Peattie and Gilbert 1980; Ehresmann et al. 1987)
DEPC structure mapping assay		Diethylpyrocarbonate	CHEBI:59051	(Peattie and Gilbert 1980; Johnston and Rich 1985; Ehresmann et al. 1987),
Kethoxal structure mapping assay		Kethoxal (1,1-Dihydroxy-3-ethoxy-2-butanone)	CHEBI:59052	(Danesh et al. 1986; Ehresmann et al. 1987)
CMCT structure mapping assay		1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate	CHEBI:59053	(Danesh et al. 1986; Ehresmann et al. 1987)
NMIA RNA structure mapping assay	SHAPE mapping assay	N-methylisatoic anhydride	CHEBI:59054	(Merino et al. 2005)
Fe-BABE RNA structure mapping assay		Fe(II)-BABE (iron(S)-1-(p-bromoacetamidobenzyl) ethylenediaminetetraacetate)	CHEBI:59055	(Heilek et al. 1995)
MPE-Fe(II) structure mapping assay		Methidiumpropyl-EDTA.Fe(II)	CHEBI:59056	(Vary and Vournakis 1984a)
ENU structure mapping assay		Ethylnitrosourea	CHEBI:23995	(Vlassov et al. 1980; Ehresmann et al. 1987)
Lead structure mapping assay		Lead	CHEBI:27889	(Gornicki et al. 1989)
Rhodium DNA structure mapping assay		Rhodium	CHEBI:33359	(Kirshenbaum et al. 1988)
Ruthenium DNA structure mapping assay		Ruthenium	CHEBI:30682	(Barton 1986)
Terbium RNA structure mapping assay		Terbium	CHEBI:33376	(Walter et al. 2000)
DNAse I structure mapping assay	DNAse footprinting	DNAse I	PRO:000006592	(Galas and Schmitz 1978; Brenowitz et al. 1986)
RNAse CL3 structure mapping assay		RNAse CL3	PRO:000025478	(Florentz et al. 1982)
Nuclease S1 structure mapping assay		Nuclease S1	PRO:000025471	(Wurst et al. 1978)
RNAse T1 structure mapping assay		RNAse T1	PRO:000025467	(Wrede et al. 1979)
RNAse T2 structure mapping assay		RNAse T2	PRO:000014060	(Vary and Vournakis 1984b)
RNAse U2 structure mapping assay		RNAse U2	PRO:000025475	(Mougel et al. 1987)
RNAse V1 structure mapping assay		RNASE V1	PRO:000025477	(Lockard and Kumar 1981)
OH-radical structure mapping assay	OH footprinting assay; MOHCA	OH-radical	CHEBI:29191	(Latham and Cech 1989)
Inline Probing		Inline Probing	No added chemical probe	(Soukup and Breaker 1999; Regulski and Breaker 2008)
1M7 RNA structure mapping assay	SHAPE mapping assay	1-methyl-7-nitroisatoic anhydride (1M7)	CHEBI:60343	(Mortimer and Weeks 2007)
RNAse I		RNAse I	PRO:000014042	(Tranguch et al. 1994)

tabs: "Investigation," "Study-Assay," and "Data Matrix." The Investigation tab contains bibliographical references, authorship, dates, and protocol-related information. In general, a single ISA-Tab file will communicate all data presented in one manuscript. An assay is defined as a mapping experiment using one probe on one nucleic acid, and each row in the Study-Assay tab (Supplemental Fig. 2) corresponds to one such experiment. The actual data is stored in the third Data Matrix tab where each column corresponds to one assay (Supplemental Fig. 2). There is therefore an implicit one-to-one correspondence between rows in the Assays tab and the columns in the Data Matrix tabs. This correspondence is explicitly coded in the Study-Assay tab by a column with Assay Names that correspond to the first row of the Data-Matrix tab.

The Assays tab is where the ontological classification outlined above is used. The Term Accession Number column corresponds to the OBI accession number specifying the type of chemical or enzymatic mapping experiment. Furthermore, other variables (such as monovalent and divalent salt concentration and type) are specified in additional columns in the Assays tab. In principle, any number of experimental conditions can be specified in this way; in practice, only those experimental variables that change (for example, $MgCl_2$ concentration) are recorded. In this way, the most important variables in the experiment are captured systematically.

Creating and sharing an ISA-Tab file

Given that an ISA-Tab file is simply a spreadsheet, specialized software is not required. To simplify the process of creating the appropriate file, we have developed a tutorial document, provided in the supplement of this manuscript and also at the SNRNASM website (http://snrnasm.bio.unc.edu/SNRNASM_Tutorial.pdf). Additionally, links to template ISA-Tab files and example data sets are also available online. Column and row headers are colored in green and yellow, indicating fields that require user input or not, respectively. In practice, most users will simply download an example ISA-Tab file and modify it according to their needs. In most cases, data can be simply pasted into the template to produce a new ISA-Tab file, greatly reducing the burden of data sharing. Alternatively, ISAcreator (http://isatab.sourceforge.net/isacreator.html) used in combination with dedicated configurations (http://tinyurl.com/69r7au3) can provide the necessary support for managing structure mapping data locally prior to release. It is an easy to use tool that automatically helps create and populate ISA-Tab files as well as organize data in the lab. Controlled vocabularies ease data reporting while reducing annotation ambiguity. The capability to save ISA-Tab reports as Google spreadsheets directly from ISAcreator tool is currently being developed and will facilitate sharing.

As mentioned above, we propose a distributed approach to storing ISA-Tab files. Therefore, we have not created a central server where such data are to be uploaded. Instead users may choose to upload their ISA-Tab files to their own servers, or alternatively make them publically available through a free cloud service like Google Docs. Instructions on how to make data public are provided in the tutorial. One advantage of making data available in the cloud is that it allows us to leverage web search engines to find SNRNASM data. Links to the SNRNASM data from pages that are already indexed will facilitate discovery by automated Internet crawling engines. We therefore encourage users to link to their data from their homepages, as well as from the primary publication. Additionally, we link to any SNRNASM data submitted to http://snrnasm.bio.unc.edu. We have also created an automated ISA-Tab validation tool for SNRNASM data at http://rmdb.stanford.edu/repository/tools/validate/ that will identify inconsistencies in a file.

APPLICATIONS

Example use cases

The most likely SNRNASM use case is also the most straightforward in terms of implementation. An investigator reads a paper in which structure-mapping data was collected and wishes to reanalyze these data in a new context. Rather than having to extract the data from a pdf in the supplement, the original SNRNASM data can be obtained in a format that is easily parsed (Fig. 1B). Alternatively, the user can search for SNRNASM data and the names of the authors.

As mentioned above, solution conditions (especially monovalent and divalent cation concentration) significantly alter the three-dimensional (3D) conformation of RNA (Heilman-Miller et al. 2001a; Takamoto et al. 2002, 2004; Koculi et al. 2007). Chemical and enzymatic probes are often used to study the effects of solution conditions on the structure of RNA (Vary and Vournakis 1984b; Celander and Cech 1991; Mathews et al. 1997; Uchida et al. 2003; Takamoto et al. 2004).

To illustrate the value of sharing chemical and enzymatic mapping data, we performed a simple meta-analysis of the effects of solution conditions on the DMS (OBI:0001015) accessibilities of functional RNA residues. Specifically, we wanted to find DMS structure mapping data that were collected under similar divalent solution conditions for different RNAs. We therefore searched for SNRNASM files containing the terms OBI:0001015 and CHEBI:6636 ($MgCl_2$), and identified two studies where DMS chemical mapping data were collected on RNA in the absence and in the presence of 10 mM $MgCl_2$. In the first study, DMS chemical mapping data were collected on sequence variants of the SRP (Signal Recognition Particle) domain IV motif (Das et al. 2010), while in the second study, data was collected on the P4–P6 subdomain of the L-21 *Tetrahymena thermophila* group I intron (Quarrier et al. 2010). SNRNASM classification therefore facilitated identification of similar data sets for meta-analysis.

Because SNRNASM data files provide easy access to the data, rapidly generating new visualizations is greatly simplified. We used a tool to project structure mapping data on an RNA secondary structure diagram provided with the SAFA software (Das et al. 2005) to visualize the DMS data from these experiments on two-dimensional (2D) representations of the RNA (Fig. 3). What is immediately apparent from our visualization of the DMS mapping data is that the addition of 10 mM $MgCl_2$ results in significant changes in the overall DMS reactivity for P4–P6 (Fig. 3A) and, to a lesser degree, for the SRP domain IV motif (Fig. 3B). The effects of this structural change are visible when comparing the no-Mg^{2+} and plus-Mg^{2+} data sets for the P4P6 domain (Fig. 3A), which includes significant tertiary contact formation upon folding (Deras et al. 2000; Doherty and Doudna 2001; Russell et al. 2002, 2006).

Interestingly, subtle effects in DMS reactivity are also observed upon Mg^{2+} addition to the SRP domain IV hairpin (Fig. 3B). No tertiary contacts are present in this RNA, so one might expect the DMS reactivity to be identical in both solution conditions. This domain was chosen for study because it is composed of a series of noncanonical base pairs (indicated using the Leontis-Westhof nomenclature in Figure 3B (Leontis and Westhof 2003). The relationships between 3D structure and chemical reactivity are not simple,

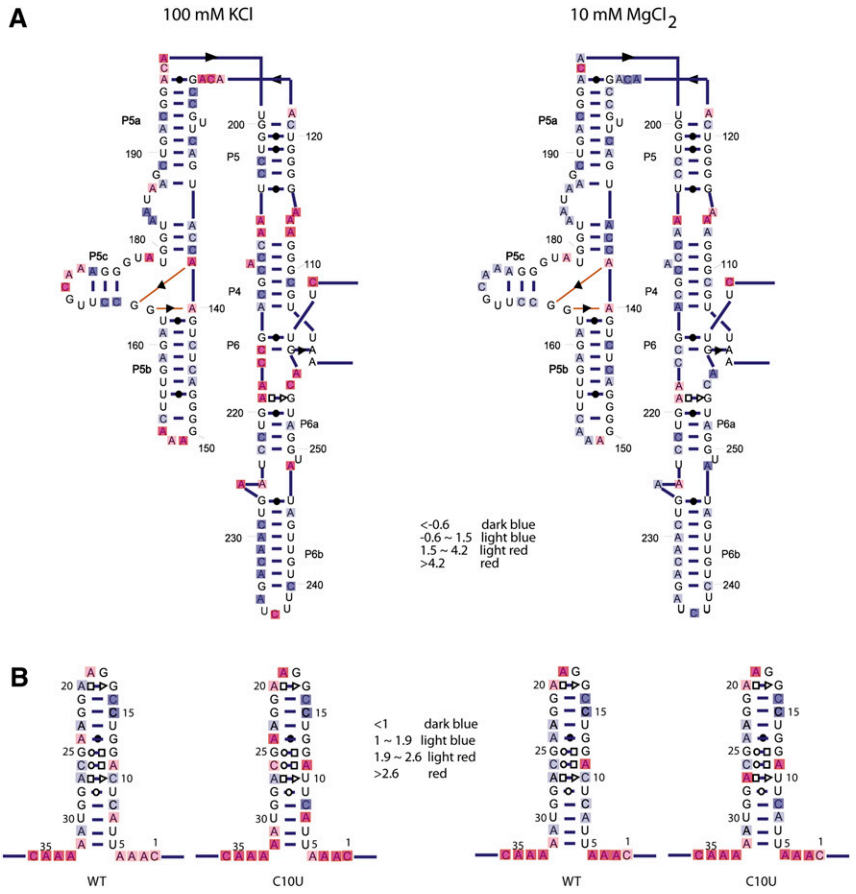

FIGURE 3. Example meta-analysis of DMS (OBI:0001015) chemical mapping data from two separate studies on RNA. For example, to visualize the effects of Mg^{2+} on the DMS reactivity of nucleic acids, we searched for OBI:0001015 (DMS) and CHEBI:6636 ($MgCl_2$) in ISA-Tab (Rocca-Serra et al. 2010) files and identified two studies where RNA was probed with DMS in the absence and presence of Mg^{2+}. We then downloaded the two ISA-Tab files (https://spreadsheets.google.com/ccc?key=0As58Pw6ZT3UtdGFveExsek9tdUJNS0xXbUFmRE1ZR0E&hl=en#gid=1 and https://spreadsheets.google.com/ccc?key=0AvCayBYdTclldEJoQ3otbWE5RGx0VzdobmVjX2Q5b3c&hl=en#gid=0) and used a tool included in the SAFA software (Simmons et al. 2009) to visualize the data on the RNA. (*A*) Secondary structure diagram of the L-21 *T. thermophila* group I intron with DMS data mapped to its secondary structure with (*right*) and without (*left*) Mg^{2+} present. (*B*) Secondary structure diagram of domain IV of SRP with and without Mg^{2+} present.

but the availability of large numbers of quantitative data sets like the two we analyzed here will allow us to better analyze these relationships in a quantitative and predictive manner (Woodson 2000; Heilman-Miller et al. 2001b; Koculi et al. 2006; Laederach et al. 2007).

DISCUSSION

Our objective in this work is to establish a simple and robust standard that facilitates sharing of single nucleotide resolution nucleic acid structure mapping (SNRNASM) data. To achieve this objective, we:

1. Describe and classify common SNRNASM experiments using a standardized (ontological) vocabulary.
2. Develop a standard format for reporting probing data that is easily read both by humans and computers.
3. Provide a means by which to make these data widely available.

Our SNRNASM classification depends on several ontologies, including the Ontology of Biomedical Investigations (OBI) (Whetzel et al. 2006), the Chemical Entities for Biomedical Investigations (CHEBI) (Degtyarenko et al. 2008, 2009), the Protein Ontology (PRO) (Natale et al. 2007, 2011), and the RNA Ontology (RNAO) (Leontis et al. 2006; Brown et al. 2009; Hoehndorf et al. 2011). SNRNASM experiments are described as assays in OBI, with the input being the nucleic acid and the chemical or enzymatic probe, while the output is a measurement of reactivity. We have added the chemical probes that were not already in CHEBI (Table 1) to uniquely identify the OBI inputs. Similarly, for the nucleases used for enzymatic probing, we obtained unique Protein Ontology identifiers (Table 1). This allows us to uniquely identify each SNRNASM type and assign it an OBI identifier (Fig. 2). The RNA Ontology (RNAO) annotates crucial structural features of RNA molecules extracted from atomic-resolution 3D structures, including all non-Watson–Crick base pairs (Leontis et al. 2006; Brown et al. 2009; Hoehndorf et al. 2011).

We sought to be as inclusive as possible, and any omissions from the SNRNASM techniques (described in Fig. 2 and Table 1) are inadvertent. These are publically available and can easily be updated by community input (available at http://bit.ly/d51yNY); thus, expanding the SNRNASM classification is straightforward. Our criteria for including an assay into our classification require: a primary publication, that the assay either modifies or cleaves a nucleic acid, that the data can be interpreted structurally, and that the modification or cleavage is localized to a specific nucleotide. The list of SNRNASM assays reported in Table 1 therefore represents a starting point for the classification of these experimental techniques and will evolve as new methods are developed. We defined two broad classes of SNRNASMs, chemical and enzymatic (Fig. 2). It is likely that new categories of SNRNASM will be required in the future. Advances in deep sequencing and other genome-wide techniques will lead to

whole- or large-scale transcriptome analysis in a single experiment (Kertesz et al. 2010; Mauger and Weeks 2010; Underwood et al. 2010). These experiments generate large amounts of data and will require a systematic approach for documenting and distributing results accurately and efficiently.

This standardization effort represents the beginning of a community effort to make SNRNASM data widely accessible, to facilitate quantitative comparative analysis, to establish predictive relationships between nucleic acid structure and chemical or enzymatic reactivity, and to provide the software and algorithm development communities with essential data for training and validation. By enabling large-scale meta-analysis, it may become possible to discover new approaches for interpreting the results of SNRNASM assays. We therefore strongly encourage laboratories carrying out these assays to make their data available upon publication.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

Workshops allowing the co-authors to develop the SNRNASM standards were made possible through the RNA Ontology Consortium US National Science Foundation Grant #0443508. This work was also funded in part through NIH grants R21MH087336, R00GM079953, R01GM07648, R01AI068462, and the Howard Hughes Medical Institute (HHMI), a Burroughs-Wellcome CASI award (to R.D.), and a Stanford Graduate Fellowship (to P.C.)

Received March 28, 2011; accepted April 23, 2011.

REFERENCES

Barton JK. 1986. Metals and DNA: molecular left-handed complements. *Science* **233:** 727–734.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2008. GenBank. *Nucleic Acids Res* **36:** D25–D30.

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur J Biochem* **80:** 319–324.

Brenowitz M, Senear DF, Shea MA, Ackers GK. 1986. Quantitative DNase footprint titration: a method for studying protein-DNA interactions. *Methods Enzymol* **130:** 132–181.

Brinkman RR, Courtot M, Derom D, Fostel JM, He Y, Lord P, Malone J, Parkinson H, Peters B, Rocca-Serra P, et al. 2010. Modeling biomedical experimental processes with OBI. *J Biomed Semantics* (Suppl 1) **1:** S7. doi: 10.1186/2041-1480-1-S1-S7.

Brown JW, Birmingham A, Griffiths PE, Jossinet F, Kachouri-Lafond R, Knight R, Lang BF, Leontis N, Steger G, Stombaugh J, et al. 2009. The RNA structure alignment ontology. *RNA* **15:** 1623–1631.

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, et al. 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3:** 2. doi: 10.1186/1471-2105-3-2.

Celander DW, Cech TR. 1991. Visualizing the higher order folding of a catalytic RNA molecule. *Science* **251:** 401–407.

Danesh M, Seth S, Noller HF. 1986. Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J Mol Biol* **187:** 399–416.

Das R, Laederach A, Pearlman SM, Herschlag D, Altman RB. 2005. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* **11:** 344–354.

Das R, Kudaravalli M, Jonikas M, Laederach A, Fong R, Schwans JP, Baker D, Piccirilli JA, Altman RB, Herschlag D. 2008. Structural inference of native and partially folded RNA by high-throughput contact mapping. *Proc Natl Acad Sci* **105:** 4144–4149.

Das R, Karanicolas J, Baker D. 2010. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods* **7:** 291–294.

Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcantara R, Darsow M, Guedj M, Ashburner M. 2008. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res* **36:** D344–D350.

Degtyarenko K, Hastings J, de Matos P, Ennis M. 2009. ChEBI: an open bioinformatics and cheminformatics resource. *Curr Protoc Bioinformatics* doi: 10.1002/0471250953.bi1409s26.

Deigan KE, Li TW, Mathews DH, Weeks KM. 2009. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci* **106:** 97–102.

Deras ML, Brenowitz M, Ralston CY, Chance MR, Woodson SA. 2000. Folding mechanism of the *Tetrahymena* ribozyme P4-P6 domain. *Biochemistry* **39:** 10975–10985.

Doherty EA, Doudna JA. 2001. Ribozyme structures and mechanisms. *Annu Rev Biophys Biomol Struct* **30:** 457–475.

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res* **15:** 9109–9128.

Elnitski LL, Shah P, Moreland RT, Umayam L, Wolfsberg TG, Baxevanis AD. 2007. The ENCODEdb portal: simplified access to ENCODE Consortium data. *Genome Res* **17:** 954–959.

Florentz C, Briand JP, Romby P, Hirth L, Ebel JP, Glege R. 1982. The tRNA-like structure of turnip yellow mosaic virus RNA: structural organization of the last 159 nucleotides from the 3′ OH terminus. *EMBO J* **1:** 269–276.

Fogarty M. 2002. Fire hits UC-Santa Cruz labs. *Scientist* **16:** 18.

Galas DJ, Schmitz A. 1978. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res* **5:** 3157–3170.

Gornicki P, Baudin F, Romby P, Wiewiorowski M, Kryzosiak W, Ebel JP, Ehresmann C, Ehresmann B. 1989. Use of lead(II) to probe the structure of large RNA's. Conformation of the 3′ terminal domain of E. coli 16S rRNA and its involvement in building the tRNA binding sites. *J Biomol Struct Dyn* **6:** 971–984.

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res* **31:** 439–441.

Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33:** D121–D124.

Heilek GM, Marusak R, Meares CF, Noller HF. 1995. Directed hydroxyl radical probing of 16S rRNA using Fe(II) tethered to ribosomal protein S4. *Proc Natl Acad Sci* **92:** 1113–1116.

Heilman-Miller SL, Pan J, Thirumalai D, Woodson SA. 2001a. Role of counterion condensation in folding of the *Tetrahymena* ribozyme. II. Counterion-dependence of folding kinetics. *J Mol Biol* **309:** 57–68.

Heilman-Miller SL, Thirumalai D, Woodson SA. 2001b. Role of counterion condensation in folding of the *Tetrahymena* ribozyme. I. Equilibrium stabilization by cations. *J Mol Biol* **306:** 1157–1166.

Hoehndorf R, Batchelor C, Bittner T, Dumontier M, Eilbeck K, Knight R, Mungall CJ, Richardson JS, Stombaugh J, Westhof E, et al. 2011. The RNA Ontology (RNAO): an ontology for integrating RNA sequence and structure data. *Applied Ontology* **6:** 53–89.

Johnston BH, Rich A. 1985. Chemical probes of DNA conformation: detection of Z-DNA at nucleotide resolution. *Cell* **42:** 713–724.

Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467:** 103–107.

Kirshenbaum MR, Tribolet R, Barton JK. 1988. Rh(DIP)3(3+): a shape-selective metal complex which targets cruciforms. *Nucleic Acids Res* **16:** 7943–7960.

Koculi E, Thirumalai D, Woodson SA. 2006. Counterion charge density determines the position and plasticity of RNA folding transition states. *J Mol Biol* **359:** 446–454.

Koculi E, Hyeon C, Thirumalai D, Woodson SA. 2007. Charge density of divalent metal cations determines RNA stability. *J Am Chem Soc* **129:** 2676–2682.

Laederach A, Shcherbakova I, Jonikas MA, Altman RB, Brenowitz M. 2007. Distinct contribution of electrostatics, initial conformational ensemble, and macromolecular stability in RNA folding. *Proc Natl Acad Sci* **104:** 7045–7050.

Laederach A, Das R, Vicens Q, Pearlman SM, Brenowitz M, Herschlag D, Altman RB. 2008. Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nat Protoc* **3:** 1395–1401.

Latham JA, Cech TR. 1989. Defining the inside and outside of a catalytic RNA molecule. *Science* **245:** 276–282.

Leontis NB, Westhof E. 2003. Analysis of RNA motifs. *Curr Opin Struct Biol* **13:** 300–308.

Leontis NB, Altman RB, Berman HM, Brenner SE, Brown JW, Engelke DR, Harvey SC, Holbrook SR, Jossinet F, Lewis SE, et al. 2006. The RNA Ontology Consortium: an open invitation to the RNA community. *RNA* **12:** 533–541.

Lockard RE, Kumar A. 1981. Mapping tRNA structure in solution using double-strand-specific ribonuclease V1 from cobra venom. *Nucleic Acids Res* **9:** 5125–5140.

Mathews DH, Banerjee AR, Luan DD, Eickbush TH, Turner DH. 1997. Secondary structure model of the RNA recognized by the reverse transcriptase from the R2 retrotransposable element. *RNA* **3:** 1–16.

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci* **101:** 7287–7292.

Mauger DM, Weeks KM. 2010. Toward global RNA structure analysis. *Nat Biotechnol* **28:** 1178–1179.

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127:** 4223–4231.

Mitra S, Shcherbakova IV, Altman RB, Brenowitz M, Laederach A. 2008. High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res* **36:** e63. doi: 10.1093/nar/gkn267.

Moreira DA, Musen MA. 2007. OBO to OWL: a protege OWL tab to read/save OBO ontologies. *Bioinformatics* **23:** 1868–1870.

Mortimer SA, Weeks KM. 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129:** 4144–4145.

Mougel M, Eyermann F, Westhof E, Romby P, Expert-Bezancon A, Ebel JP, Ehresmann B, Ehresmann C. 1987. Binding of *Escherichia coli* ribosomal protein S8 to 16 S rRNA. A model for the interaction and the tertiary structure of the RNA binding site. *J Mol Biol* **198:** 91–107.

Musen MA, Shah NH, Noy NF, Dai BY, Dorf M, Griffith N, Buntrok J, Jonquet C, Montegut MJ, Rubin DL. 2008. BioPortal: ontologies and data resources with the click of a mouse. *AMIA Annu Symp Proc* **6:** 1223–1224.

Natale DA, Arighi CN, Barker WC, Blake J, Chang TC, Hu Z, Liu H, Smith B, Wu CH. 2007. Framework for a protein ontology. *BMC Bioinformatics* (Suppl 9) **8:** S1. doi: 10.1186/1471-2105-8-S9-S1.

Natale DA, Arighi CN, Barker WC, Blake JA, Bult CJ, Caudy M, Drabkin HJ, D'Eustachio P, Evsikov AV, Huang H, et al. 2011. The Protein Ontology: a structured representation of protein forms and complexes. *Nucleic Acids Res* **39:** D539–D545.

Noy NF, Crubezy M, Fergerson RW, Knublauch H, Tu SW, Vendetti J, Musen MA. 2003. Protégé-2000: an open-source ontology-development and knowledge-acquisition environment. *AMIA Annu Symp Proc* **2003:** 953.

Peattie DA, Gilbert W. 1980. Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci* **77:** 4679–4682.

Quarrier S, Martin JS, Davis-Neulander L, Beauregard A, Laederach A. 2010. Evaluation of the information content of RNA structure mapping data for secondary structure prediction. *RNA* **16:** 1108–1117.

Regulski EE, Breaker RR. 2008. In-line probing analysis of riboswitches. *Methods Mol Biol* **419:** 53–67.

Rocca-Serra P, Brandizi M, Maguire E, Sklyar N, Taylor C, Begley K, Field D, Harris S, Hide W, Hofmann O, et al. 2010. ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics* **26:** 2354–2356.

Russell R, Herschlag D. 2001. Probing the folding landscape of the *Tetrahymena* ribozyme: commitment to form the native conformation is late in the folding pathway. *J Mol Biol* **308:** 839–851.

Russell R, Zhuang X, Babcock HP, Millett IS, Doniach S, Chu S, Herschlag D. 2002. Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci* **99:** 155–160.

Russell R, Das R, Suh H, Travers KJ, Laederach A, Engelhardt MA, Herschlag D. 2006. The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J Mol Biol* **363:** 531–544.

Schroeder R, Barta A, Semrad K. 2004. Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol* **5:** 908–919.

Shcherbakova I, Brenowitz M. 2008. Monitoring structural changes in nucleic acids with single residue spatial and millisecond time resolution by quantitative hydroxyl radical footprinting. *Nat Protoc* **3:** 288–302.

Simmons K, Martin JS, Shcherbakova I, Laederach A. 2009. Rapid quantification and analysis of kinetic •OH radical footprinting data using SAFA. *Methods Enzymol* **468:** 47–66.

Soukup GA, Breaker RR. 1999. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5:** 1308–1325.

Takamoto K, He Q, Morris S, Chance MR, Brenowitz M. 2002. Monovalent cations mediate formation of native tertiary structure of the *Tetrahymena thermophila* ribozyme. *Nat Struct Biol* **9:** 928–933.

Takamoto K, Das R, He Q, Doniach S, Brenowitz M, Herschlag D, Chance MR. 2004. Principles of RNA compaction: insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations. *J Mol Biol* **343:** 1195–1206.

Thirumalai D, Hyeon C. 2005. RNA and protein folding: common themes and variations. *Biochemistry* **44:** 4957–4970.

Tijerina P, Mohr S, Russell R. 2007. DMS footprinting of structured RNAs and RNA-protein complexes. *Nat Protoc* **2:** 2608–2623.

Tranguch AJ, Kindelberger DW, Rohlman CE, Lee JY, Engelke DR. 1994. Structure-sensitive RNA footprinting of yeast nuclear ribonuclease P. *Biochemistry* **33:** 1778–1787.

Tullius TD. 2002. Probing DNA structure with hydroxyl radicals. *Curr Protoc Nucleic Acid Chem* **6:** 6–7.

Uchida T, Takamoto K, He Q, Chance MR, Brenowitz M. 2003. Multiple monovalent ion-dependent pathways for the folding of the L-21 *Tetrahymena thermophila* ribozyme. *J Mol Biol* **328:** 463–478.

Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, Lowe TM, Salama SR, Haussler D. 2010. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Methods* **7:** 995–1001.

Vary CP, Vournakis JN. 1984a. RNA structure analysis using methidiumpropyl-EDTA.Fe(II): a base-pair-specific RNA structure probe. *Proc Natl Acad Sci* **81:** 6978–6982.

Vary CP, Vournakis JN. 1984b. RNA structure analysis using T2 ribonuclease: detection of pH and metal ion induced conformational changes in yeast tRNAPhe. *Nucleic Acids Res* **12:** 6763–6778.

Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. 2008. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14:** 1979–1990.

Vlassov VV, Giege R, Ebel JP. 1980. The tertiary structure of yeast tRNAPhe in solution studied by phosphodiester bond modification with ethylnitrosourea. *FEBS Lett* **120:** 12–16.

Walter NG, Yang N, Burke JM. 2000. Probing non-selective cation binding in the hairpin ribozyme with Tb(III). *J Mol Biol* **298:** 539–555.

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460:** 711–716.

Weeks KM. 2010. Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol* **20:** 295–304.

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, et al. 2008. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **36:** D13–D21.

Whetzel PL, Brinkman RR, Causton HC, Fan L, Field D, Fostel J, Fragoso G, Gray T, Heiskanen M, Hernandez-Boussard T, et al. 2006. Development of FuGO: an ontology for functional genomics investigations. *OMICS* **10:** 199–204.

Wilkinson KA, Merino EJ, Weeks KM. 2006. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1:** 1610–1616.

Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6:** e96. doi: 10.1371/journal.pbio.0060096.

Woodson SA. 2000. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell Mol Life Sci* **57:** 796–808.

Woodson SA. 2008. RNA folding and ribosome assembly. *Curr Opin Chem Biol* **12:** 667–673.

Wrede P, Wurst R, Vournakis J, Rich A. 1979. Conformational changes of yeast tRNAPhe and *E. coli* tRNA2Glu as indicated by different nuclease digestion patterns. *J Biol Chem* **254:** 9608–9616.

Wurst RM, Vournakis JN, Maxam AM. 1978. Structure mapping of 5′-32P-labeled RNA with S1 nuclease. *Biochemistry* **17:** 4493–4499.

Zhang B, Pan X, Wang Q, Cobb GP, Anderson TA. 2006. Computational identification of microRNAs and their targets. *Comput Biol Chem* **30:** 395–407.