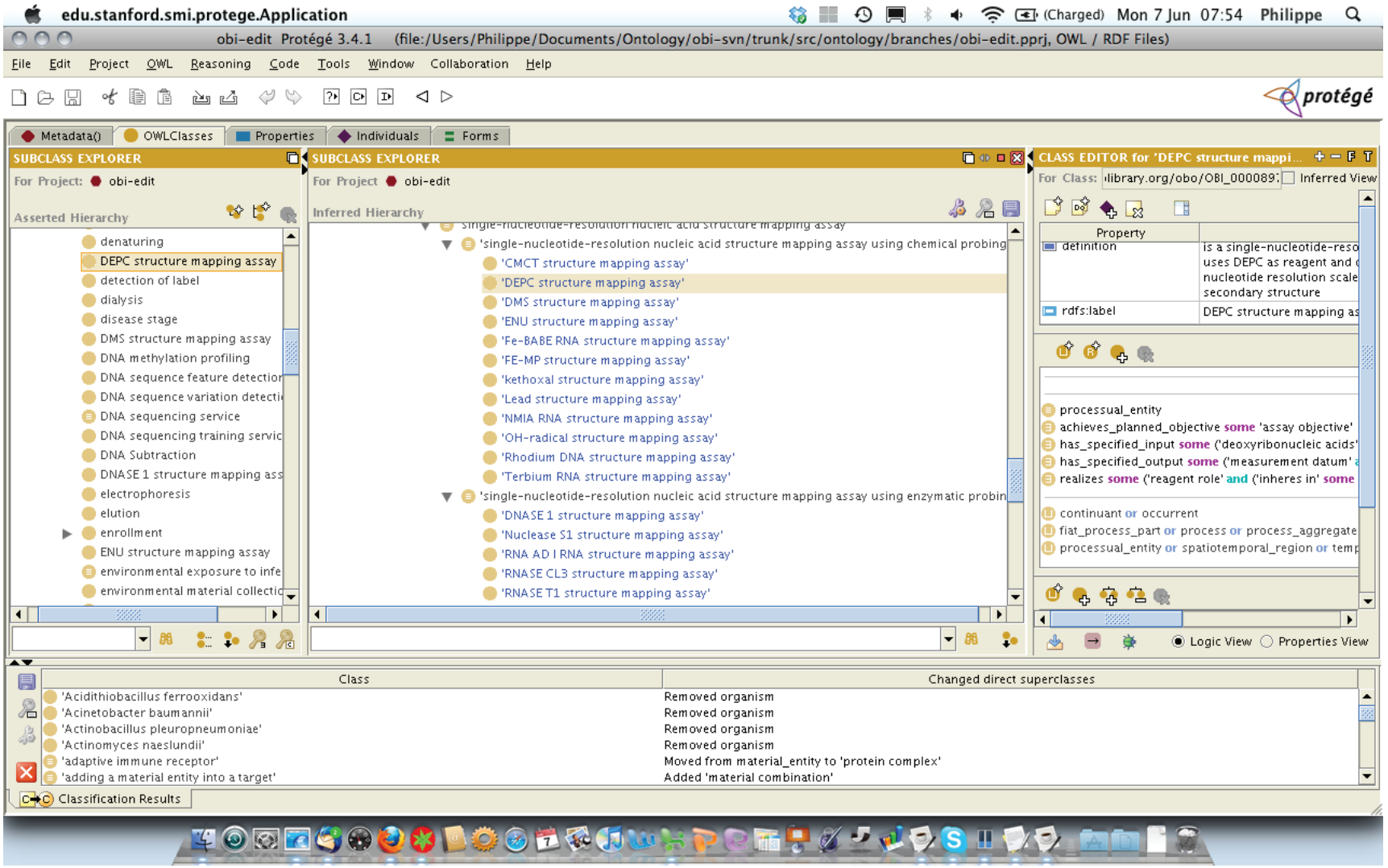
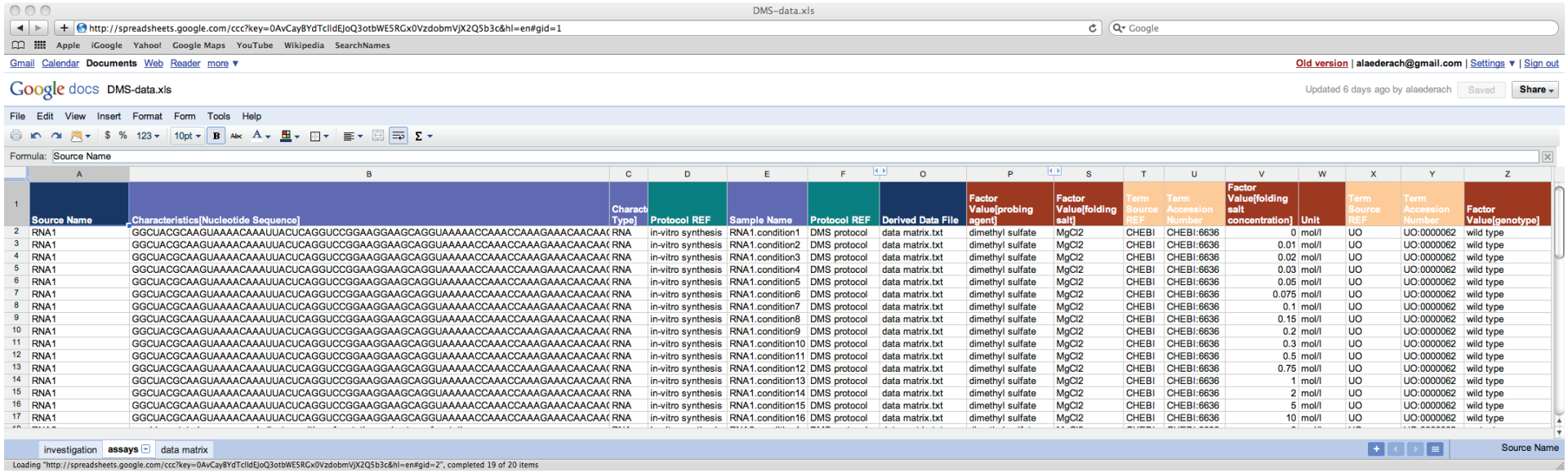
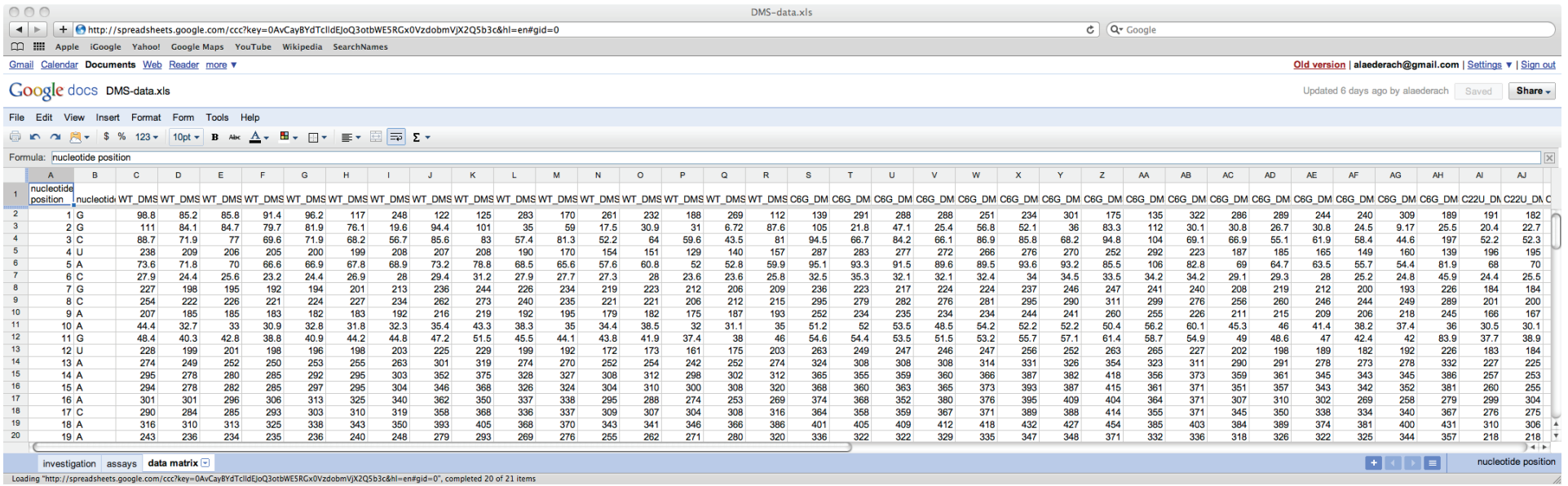


Supplementary Figure 1



Source Name	Characteristics[Nucleotide Sequence]	Charact Type]	Protocol REF	Sample Name	Protocol REF	Derived Data File	Factor Value[probing agent]	Factor Value[folding salt]	Term Source REF	Term Accession Number	Factor Value[folding salt concentration]	Unit	Term Source REF	Term Accession Number	Factor Value[genotype]
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition1	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition2	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.01	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition3	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.02	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition4	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.03	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition5	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.05	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition6	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.075	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition7	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.1	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition8	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.15	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition9	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.2	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition10	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.3	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition11	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.5	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition12	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	0.75	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition13	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	1	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition14	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	2	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition15	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	5	mol/l	UO	UO:0000062	wild type
RNA1	GGCUACGCAAGUAAAACAAAUUACUCAGGUCCGGAAGGAAGCAGGUAAAAACCAAACCAAAGAAACAACAA(	RNA	in-vitro synthesis	RNA1.condition16	DMS protocol	data matrix.txt	dimethyl sulfate	MgCl2	CHEBI	CHEBI:6636	10	mol/l	UO	UO:0000062	wild type



nucleotide position	nucleotid	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	WT_DMS	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C6G_DM	C22U_DM	C22U_DM	C
1	G	98.8	85.2	85.8	91.4	96.2	117	248	122	125	283	170	261	232	188	269	112	139	291	288	288	251	234	301	175	135	322	286	289	244	240	309	189	191	182	
2	G	111	84.1	84.7	79.7	81.9	76.1	19.6	94.4	101	35	59	17.5	30.9	31	6.72	87.6	105	21.8	47.1	25.4	56.8	52.1	36	83.3	112	30.1	30.8	26.7	30.8	24.5	9.17	25.5	20.4	22.7	
3	C	88.7	71.9	77	69.6	71.9	68.2	56.7	85.6	83	57.4	81.3	52.2	64	59.6	43.5	81	94.5	66.7	84.2	66.1	86.9	85.8	68.2	94.8	104	69.1	66.9	55.1	61.9	58.4	44.6	197	52.2	52.3	
4	U	238	209	206	205	200	199	208	207	208	190	170	154	151	129	140	157	287	283	277	272	266	276	270	252	292	223	187	185	165	149	160	139	196	195	
5	A	73.6	71.8	70	66.6	66.9	67.8	68.9	73.2	78.8	68.5	65.6	57.6	60.8	52	52.8	59.9	95.1	93.3	91.5	89.6	89.5	93.6	93.2	85.5	106	82.8	69	64.7	63.5	55.7	54.4	81.9	68	70	
6	C	27.9	24.4	25.6	23.2	24.4	26.9	28	29.4	31.2	27.9	27.7	27.3	28	23.6	23.6	25.8	32.5	35.3	32.1	32.1	32.4	34	34.5	33.5	34.2	34.2	29.1	29.3	28	25.2	24.8	45.9	24.4	25.5	
7	G	227	198	195	192	194	201	213	236	244	226	234	219	223	212	206	209	236	223	217	224	224	237	246	247	241	240	208	219	212	200	193	226	184	184	
8	C	254	222	226	221	224	227	234	262	273	240	235	221	221	206	212	215	295	279	282	276	281	295	290	311	299	276	256	260	246	244	249	289	201	200	
9	A	207	185	185	183	182	183	192	216	219	192	195	179	182	175	187	193	252	234	235	234	234	244	241	260	255	226	211	215	209	206	218	245	166	167	
10	A	44.4	32.7	33	30.9	32.8	31.8	32.3	35.4	43.3	38.3	35	34.4	38.5	32	31.1	35	51.2	52	53.5	48.5	54.2	52.2	52.2	50.4	56.2	60.1	45.3	46	41.4	38.2	37.4	36	30.5	30.1	
11	G	48.4	40.3	42.8	38.8	40.9	44.2	44.8	47.2	51.5	45.5	44.1	43.8	41.9	37.4	38	46	54.6	54.4	53.5	51.5	53.2	55.7	57.1	61.4	58.7	54.9	49	48.6	47	42.4	42	83.9	37.7	38.9	
12	U	228	199	201	198	196	198	203	225	229	199	192	172	173	161	175	203	263	249	247	246	247	256	252	263	265	227	202	198	189	182	192	226	183	184	
13	A	274	249	252	250	253	255	263	301	319	274	270	252	254	242	252	274	324	308	308	308	314	331	326	354	323	311	290	291	278	273	278	332	227	225	
14	A	295	278	280	285	292	295	303	352	375	328	327	308	312	298	302	312	365	355	359	360	366	387	382	418	356	373	359	361	345	343	345	386	257	253	
15	A	294	278	282	285	297	295	304	346	368	326	324	304	310	300	308	320	368	360	363	365	373	393	387	415	361	371	351	357	343	342	352	381	260	255	
16	A	301	301	296	306	313	325	340	362	350	337	338	295	288	274	253	269	374	368	352	380	376	395	409	404	364	371	307	310	302	269	258	279	299	304	
17	C	290	284	285	293	303	310	319	358	368	336	337	309	307	304	308	316	364	358	359	367	371	389	388	414	355	371	345	350	338	334	340	367	276	275	
18	A	316	310	313	325	338	343	350	393	405	368	370	343	341	346	366	386	401	405	409	412	418	432	427	454	385	403	384	389	374	381	400	431	310	306	
19	A	243	236	234	235	236	240	248	279	293	269	276	255	262	271	280	320	336	322	322	329	335	347	348	371	332	336	318	326	322	325	344	357	218	218	

Supplementary Figure 2

SNRNASM ISATab

Objective

The objective of this tutorial is to get you acquainted with the SNRNASM (Single Nucleotide Resolution Nucleic Acid Structure Mapping) ISATab standard for reporting chemical and enzymatic mapping data. We have developed this standard to facilitate the sharing of these data between labs and also to make this data available to computational labs developing new tools to predict Nucleic Acid structure. Our primary objective is to make the process of sharing the data as easy as possible. Nonetheless, it is important to capture experimental details that will allow a computer to automatically find and classify the data. The format we propose is a compromise between simplicity and the need to make the data computer readable. These two objectives are almost completely mutually exclusive, and the solution we present is therefore a compromise.

In all cases, however some data is better than no data, and it is more important to make the data available to others rather than try and get every single possible detail correct in reporting it. Furthermore, the process we propose for making the data available, uploading it to the "cloud" (i.e. Google Docs), makes it possible to correct any mistakes in the data even after release. This also means that you ultimately control access to your data and can maintain it, in much the same way a developer of a computer program releases new versions over time. As such, the indications below should be considered guidelines and not strict rules.This tutorial, along with several example ISATab files can be found at http://snrnasm.bio.unc.edu/ which is a good resource for the latest links and contacts related to SNRNASM data sharing. If you have any questions, feel free to click on the Contact link to obtain the e-mail of the current SNRNASM coordinator.

Computational Requirements

The computational requirements for reporting your data are minimal

- A computer connected to the internet
- A spreadsheet editor, for example:
  - Excel or OpenOffice (http://www.openoffice.org/) which is free.
  - Google Spreadsheets (http://www.google.com/google-d-s/spreadsheets/), also free.
- A Google account for uploading your spreadsheet (sign up at http://docs.google.com). If you already have a gmail account, you can simply use your gmail login.

Procedure

1. The first step in preparing an ISATab file with your data is to obtain a example file (or template) with similar data. Ideally, you will be able to find a file that reports on a similar experiment. A good place to look for example files is at the SNRNASM home page (http://

snrnasm.bio.unc.edu/) which indexes some of these files. However, this index is not meant to be comprehensive. Our idea for finding these data is in fact to be able to "Google" them. You may also want to Google the term "SNRNASM" with some keywords describing the technique you used and see if you find any other data of interest.

2. To edit an ISATab file you have several options. If you are most comfortable using excel, the easiest thing to do is to download an example file (under the Google docs image, Click on File->Download As->Excel) and use a copy of the file on your own computer. Alternatively, you can log into Google Docs and "Make a Copy" of the document to edit directly in your browser. You will be able to re-upload the file once you are ready to share it. A blank ISATab cal also be found at http://snrnasm.bio.unc.edu/browse.html.

3. For the purposes of this tutorial, we will look at two different ISATab files reporting on chemical and enzymatic mapping of different RNAs. These can be found at the following links: https://spread...RE1ZR0E&hl=en and http://bit.ly/9QKNKA. When you open these files in Google spreadsheets (by clicking on the above links), you will notice several important features of ISATab files:

   1. The files consist of multiple (in this case) three tabs, that are independent spreadsheets and, which can be saved as tab-delimited text files. Figure 1 illustrates the "investigation" tab, which provides information on the experiment, when it was carried out, and who carried it out.

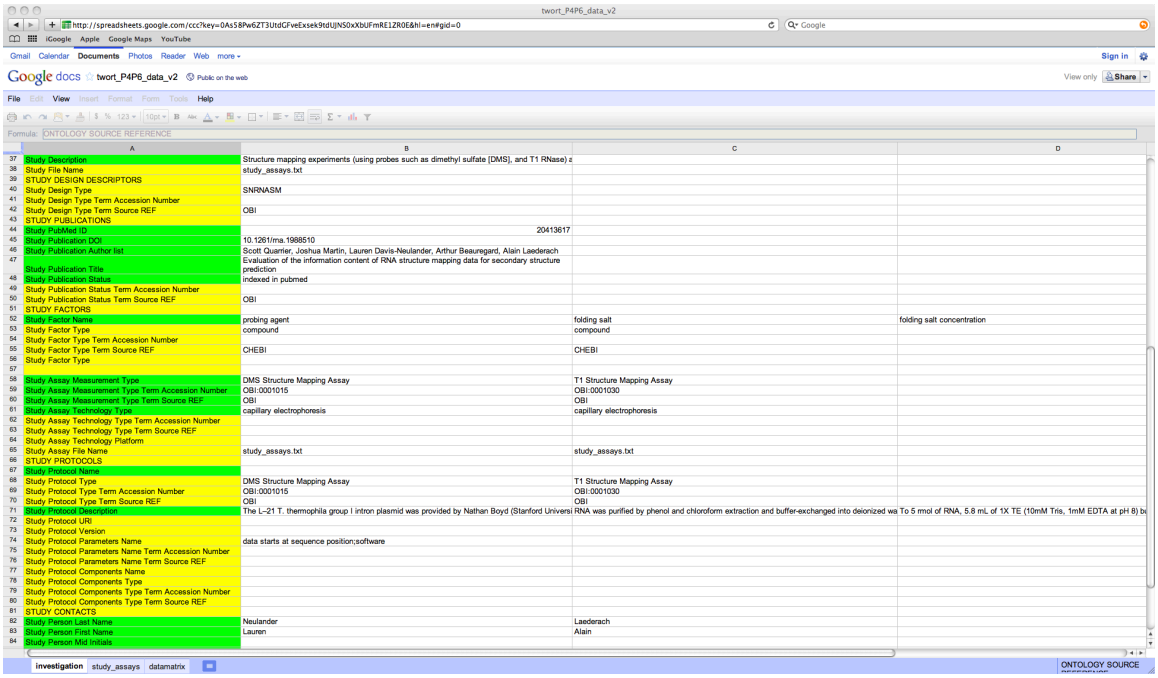


Figure 1: Screenshot of the investigation tab of an ISATab file reporting on DMS chemical mapping and which can be downloaded from http://bit.ly/9QKNKA.

   2. For these specific template files, fields that require your input are highlighted in green, while those that do not require user input are highlighted in yellow. This is done consistently throughout the files to facilitate user input. The yellow fields all describe how this experiment fits into the context of larger biomedical investigations, specifically the

Ontology of Biomedical Investigations (OBI). These will allow databases to understand what the data is and how to classify it in comprehensive data repositories. For the purposes of reporting any chemical or enzymatic chemical mapping experiment, you do not need to alter these fields.

3. All the fields are free text allowing the user much flexibility in what they choose to report. When filling out these fields, try and think about what information you might want others to know about your experiment.

4. The investigation tab: We will now go through the different fields that need to be filled out that are highlighted in green in the investigation Tab.
    1. Study Identifier: This is the last field that will need to be filled out, so you can leave it blank for now. It will be the link (or URL) to your data once you make it public.
    2. Study Title: Usually the title of the manuscript that describes this data, or a descriptive title of the data
    3. Study Submission and Release Dates: These dates can represent the dates the manuscript was submitted and then accepted. If no manuscript is associated, these can be left blank too.
    4. Study Protocol Description: This is a free text field where data describing the experimental conditions of the experiment are provided. Usually, the methods section from the manuscript can be pasted here.
    5. Study PubMed ID: When your manuscript gets published and indexed in pubmed (search for your name at http://www.ncbi.nlm.nih.gov/pubmed/) they issue a PubMed ID for your manuscript which should be reported here. If the manuscript is not yet published this can be left blank.
    6. Study Publication DOI: This is the Digital Object Identifier, which will often appear on the publisher's site of your paper (or on the pdf reprint). It can also be left blank if the paper is not published.
    7. Study Publication Author List and Title: The names of contributors and the title of the manuscript. In general, it is expected that one ISATab corresponds to one paper, however if you have multiple papers you can duplicate a section if needed.
    8. Study Publication Status: can be "in preparation," "submitted for publication," "accepted for publication," "indexed on pubmed," or "published."
    9. Study Factor Name: This is a field that may require some reflection. We declare here the factors that are varied in the experiment. In the case of the data presented in Figure 1 (http://bit.ly/9QKNKA) the experiment involved varying folding salt and measuring DMS. We have therefore declared "folding salt" and "folding salt concentration" as factors. For the second data set (https://spread...RE1ZRoE&hl=en), we performed DMS and T1 RNAse mapping at multiple salt concentrations, so we vary "folding salt," folding salt concentration," and "probing agent." If you are unsure what to put in here, filling out the assays tab will help you decide and you can come back and fill these fields out later. If you are simply reporting one chemical probe under one condition for multiple RNAs this field is blank. Other study factors can include "time" for time-resolved

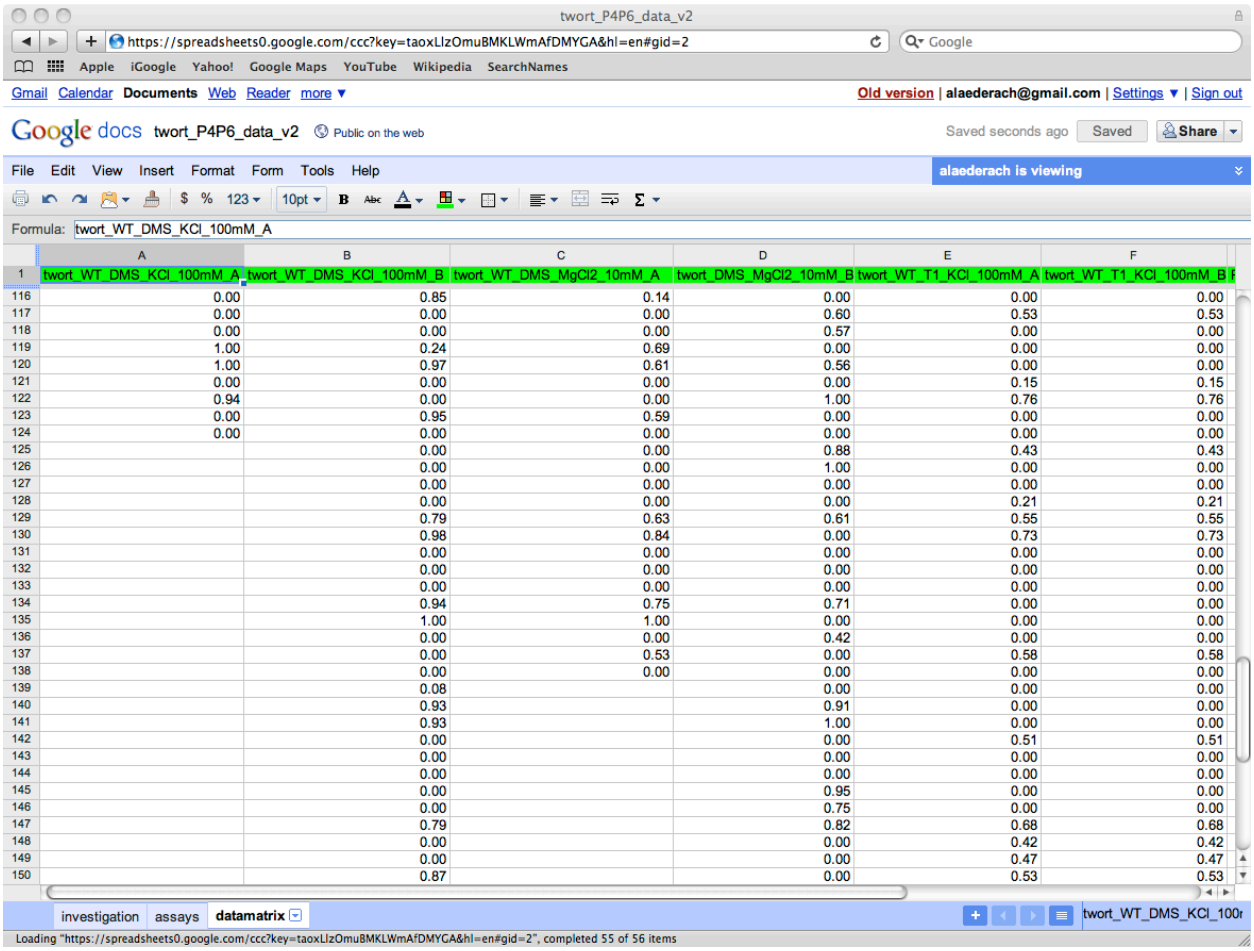
experiments, exogenous molecules (for example proteins that bind to nucleic acids), other nucleic acids, and any other variables in the experiment.

10. Study Protocol Name: This is where you would input the types of structure mapping experiments you carried out. For example, "DMS Structure Mapping," or "NMIA Structure Mapping," or "1M7 Structure Mapping." You should use the terms published in this spreadsheet (http://bit.ly/d5iyNY.) which summarizes all the ontological terms we have established to describe SNRNASM experiments.

11. STUDY CONTACTS: This section should be rather clear. You should try and include one long-term contact (e.g. a PI) so that if anyone finds this data set they have a chance of tracking you down.

5. The datamatrix tab: Although the datamatrix tab is third, we have found it easier to complete it now, which will help you define the assays tab. Presumably, you will already have your chemical or enzymatic probing data in a similar format. Figure 2 is a screen capture of the datamatrix tab. Each column in the datamatrix represents one assay, that is one probing reaction on an RNA. Usually these will be average values over multiple repeats and represent the chemical or enzymatic reactivity per nucleotide for a probe. The column headers should be unique and somewhat descriptive of the experiment. For example "twort_WT_DMS_KCL_100mM_A" is the first replica of a DMS experiment carried out on the wild-type Twort intron at 100mM KCl (Figure 2). In this case, each replica does not cover the same nucleotides and thus the columns are different lengths. This is ok, it is however important to keep track of what nucleotide the data starts at, as this will be input in the assays tab. If you do not have data for certain nucleotides, you can either leave the cell blank or set it to -999. You will be entering the sequence of your RNA or DNA in the assays tab. If you do not have quantitative data (analyzed by peak fitting software like SAFA, Shapefinder, CAFA or HitRACE) you can still report your values as "strong, medium, and weak" or any numeric scale you choose. Most importantly, when setting up your data matrix remember:

   1. Each column header must be unique
   2. Keep track of the starting nucleotide for each column, this will be reported in the assays tab.
   3. Leave blank or set cells to -999 if there is no data for that nucleotide.
   4. If you averaged the reactivity of several nucleotides, report the average value for all the nucleotides you averaged.



	A	B	C	D	E	F
1	twort_WT_DMS_KCl_100mM_A	twort_WT_DMS_KCl_100mM_B	twort_WT_DMS_MgCl2_10mM_A	twort_DMS_MgCl2_10mM_B	twort_WT_T1_KCl_100mM_A	twort_WT_T1_KCl_100mM_B
116	0.00	0.85	0.14	0.00	0.00	0.00
117	0.00	0.00	0.00	0.60	0.53	0.53
118	0.00	0.00	0.00	0.57	0.00	0.00
119	1.00	0.24	0.69	0.00	0.00	0.00
120	1.00	0.97	0.61	0.56	0.00	0.00
121	0.00	0.00	0.00	0.00	0.15	0.15
122	0.94	0.00	0.00	1.00	0.76	0.76
123	0.00	0.95	0.59	0.00	0.00	0.00
124	0.00	0.00	0.00	0.00	0.00	0.00
125		0.00	0.00	0.88	0.43	0.43
126		0.00	0.00	1.00	0.00	0.00
127		0.00	0.00	0.00	0.00	0.00
128		0.00	0.00	0.00	0.21	0.21
129		0.79	0.63	0.61	0.55	0.55
130		0.98	0.84	0.00	0.73	0.73
131		0.00	0.00	0.00	0.00	0.00
132		0.00	0.00	0.00	0.00	0.00
133		0.00	0.00	0.00	0.00	0.00
134		0.94	0.75	0.71	0.00	0.00
135		1.00	1.00	0.00	0.00	0.00
136		0.00	0.00	0.42	0.00	0.00
137		0.00	0.53	0.00	0.58	0.58
138		0.00	0.00	0.00	0.00	0.00
139		0.08		0.00	0.00	0.00
140		0.93		0.91	0.00	0.00
141		0.93		1.00	0.00	0.00
142		0.00		0.00	0.51	0.51
143		0.00		0.00	0.00	0.00
144		0.00		0.00	0.00	0.00
145		0.00		0.95	0.00	0.00
146		0.00		0.75	0.00	0.00
147		0.79		0.82	0.68	0.68
148		0.00		0.00	0.42	0.42
149		0.00		0.00	0.47	0.47
150		0.87		0.00	0.53	0.53

Figure 2: Screen capture of the datamatrix tab for T1 and DMS chemical mapping of the Twort and Tetrahymena group I introns which can be downloaded from https://spread...REIZROE&hl=en.

6. The assays tab: The assays tab is used to describe the columns in the datamatrix tab (see Figure 3). There is a one to one correspondence between the columns in the datamatrix tab and the rows in the assays tab. You will notice several important features of the assays tab:
    1. The names in the Sample Name column correspond to the names in the top row of the datamatrix tab.
    2. The Data Starts at indicates the starting nucleotide in the sequence for which there is data.
    3. You specify in the Character Type columns whether the sample is RNA or DNA, which means you can paste a DNA sequence in the Nucleotide Sequence (T instead of U).
    4. The sequence column (Charateristics[Nucleotide Sequence]) should include any mutations and be the exact sequence of the RNA that was probed, even if the data only covers a portion of that sequence.
    5. RNA Production can be either "in vitro synthesis," "cell extract," or "in vivo," depending on the source of the RNA.

6. The Protocol Ref uses a standardized vocabulary which is summarized online at http://bit.ly/d5iyNY and copied below (Figure 4). For example, since the data reported in Figure 3 is DMS structure mapping, we use the OBI:0001015 Term Accession number we looked up in http://bit.ly/d5iyNY for DMS chemical mapping.

7. Performer and Date can be approximate, and can also be the date and person completing the ISATab file.

8. Term Accession Number is looked up in Figure 4 (online at http://bit.ly/d5iyNY) in column B. Only the last digits i.e. OBI:001015 needs to be reported.

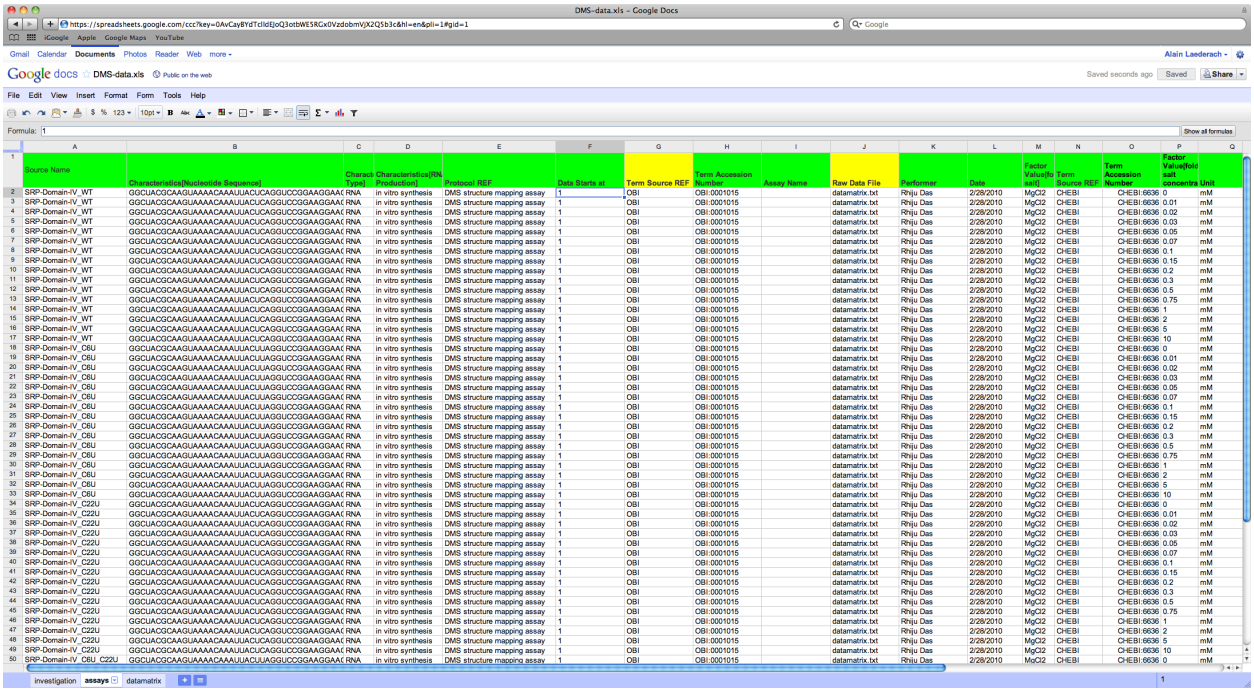
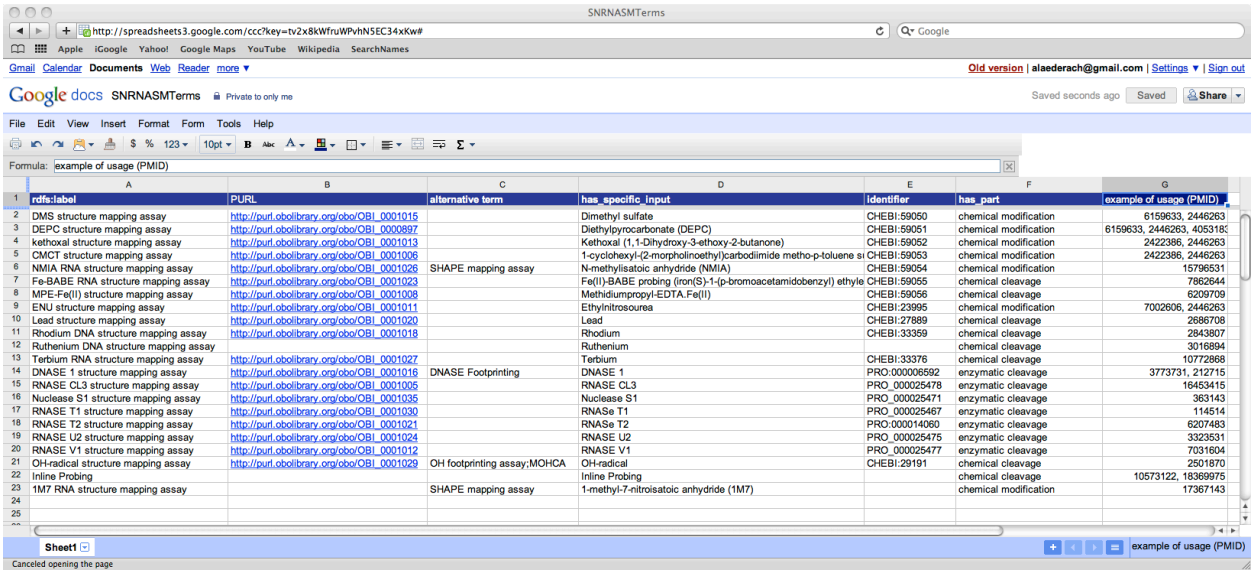


Figure 3: assays tab for DMS experiments run on a synthetic RNA molecule available from http://bit.ly/9QKNKA). Each column in this tab corresponds to a row in the datamatrix tab.

9. The final five columns in this (Figure 3) ISATab file (L-P) are optional, but in this case are important to describing the experiment in more detail. For the assays described in Figure 3, DMS probing was performed on four RNAs (one WT, two single mutants and a double mutant) at 16 different MgCl2 concentrations. We therefore report MgCl2 as a Factor Value[folding salt] (column L) along with the associated CHEBI ID (6636, Column P) which we looked up at bioportal (Figure 5). We then report the concentration in Column O and the units in Column P. If in your experiment you are simply probing different RNAs under the same solution conditions, you do not need to include Columns L-P. Your experimental conditions are specified already in the investigations tab.



	rdfs:label	PURL	alternative term	has_specific_input	identifier	has_part	example of usage (PMID)
2	DMS structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001015		Dimethyl sulfate	CHEBI:59050	chemical modification	6159633, 2446263
3	DEPC structure mapping assay	http://purl.obolibrary.org/obo/OBI_0000897		Diethylpyrocarbonate (DEPC)	CHEBI:59051	chemical modification	6159633, 2446263, 405318
4	kethoxal structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001013		Kethoxal (1,1-Dihydroxy-3-ethoxy-2-butanone)	CHEBI:59052	chemical modification	2422386, 2446263
5	CMCT structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001006		1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene si	CHEBI:59053	chemical modification	2422386, 2446263
6	NMIA RNA structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001026	SHAPE mapping assay	N-methylisatoic anhydride (NMIA)	CHEBI:59054	chemical modification	15796531
7	Fe-BABE RNA structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001023		Fe(II)-BABE probing (iron(S)-1-(p-bromoacetamidobenzyl) ethyle	CHEBI:59055	chemical cleavage	7862644
8	MPE-Fe(II) structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001008		Methidiumpropyl-EDTA.Fe(II)	CHEBI:59056	chemical cleavage	6209709
9	ENU structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001011		Ethylnitrosourea	CHEBI:23995	chemical modification	7002606, 2446263
10	Lead structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001020		Lead	CHEBI:27889	chemical cleavage	2686708
11	Rhodium DNA structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001018		Rhodium	CHEBI:33359	chemical cleavage	2843807
12	Ruthenium DNA structure mapping assay			Ruthenium		chemical cleavage	3016894
13	Terbium RNA structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001027		Terbium	CHEBI:33376	chemical cleavage	10772868
14	DNASE 1 structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001016	DNASE Footprinting	DNASE 1	PRO:000006592	enzymatic cleavage	3773731, 212715
15	RNASE CL3 structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001005		RNASE CL3	PRO_000025478	enzymatic cleavage	16453415
16	Nuclease S1 structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001035		Nuclease S1	PRO_000025471	enzymatic cleavage	363143
17	RNASE T1 structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001030		RNASe T1	PRO_000025467	enzymatic cleavage	114514
18	RNASE T2 structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001021		RNASe T2	PRO:000014060	enzymatic cleavage	6207483
19	RNASE U2 structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001024		RNASE U2	PRO_000025475	enzymatic cleavage	3323531
20	RNASE V1 structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001012		RNASE V1	PRO_000025477	enzymatic cleavage	7031604
21	OH-radical structure mapping assay	http://purl.obolibrary.org/obo/OBI_0001029	OH footprinting assay;MOHCA	OH-radical	CHEBI:29191	chemical cleavage	2501870
22	Inline Probing			Inline Probing		chemical cleavage	10573122, 18369975
23	1M7 RNA structure mapping assay		SHAPE mapping assay	1-methyl-7-nitroisatoic anhydride (1M7)		chemical modification	17367143
24							
25							

Figure 4: Standard terms and identifiers for SNRNASM experiments to use in the protocol Ref column of the assays tab. This file is regularly updated and the data obtained from http://bit.ly/d5iyNY.

10. The terms we report in Figure 4 represent a subset of the terms you can use to describe the solution conditions of your experiment. We obviously can't imagine all possible salt and buffer conditions you might use You can therefore look terms at bioportal (http://bioportal.bioontology.org/), simply search for DMS and follow the links to Chemical Entities of Biological Interest (Figure 5) and you'll see it is indeed listed as CHEBI: 59050.

11. We will also provide examples of files for time-resolved data at our central site http://snrnasm.bio.unc.edu/browse.html. In all cases, you need to think about what variables change in your experiments other than the sequence, which you report in the Sequence column, and describe those variables in as much detail as possible. The details you provide should make it possible for someone, or a computer, to decipher the main variables in your experiment.

7. Sharing your ISATab file: Once you've created an ISATab file, sharing it is actually the fun part! You simply need to upload the spreadsheet to Google Docs (click on upload under the Google docs image). When you open the spreadsheet, select the sharing option (top right hand corner) and select the settings you desire. In this case (Figure 6) we chose to make the document public on the web.

8. For search engines to index your data, there needs to exist a link to it from some other web page. We therefore encourage you to blog, e-mail, tweet and facebook the link to your data. You may also want to use a URL shortening service like http://bit.ly to provide a shorter link. We also encourage including this URL in your publications to link the two together.

9. You can automatically validate your ISA-Tab file at RMDB (http://rmdb.stanford.edu/repository/tools/validate/ as well as convert it to other formats. This simple service can catch common format errors in your file.

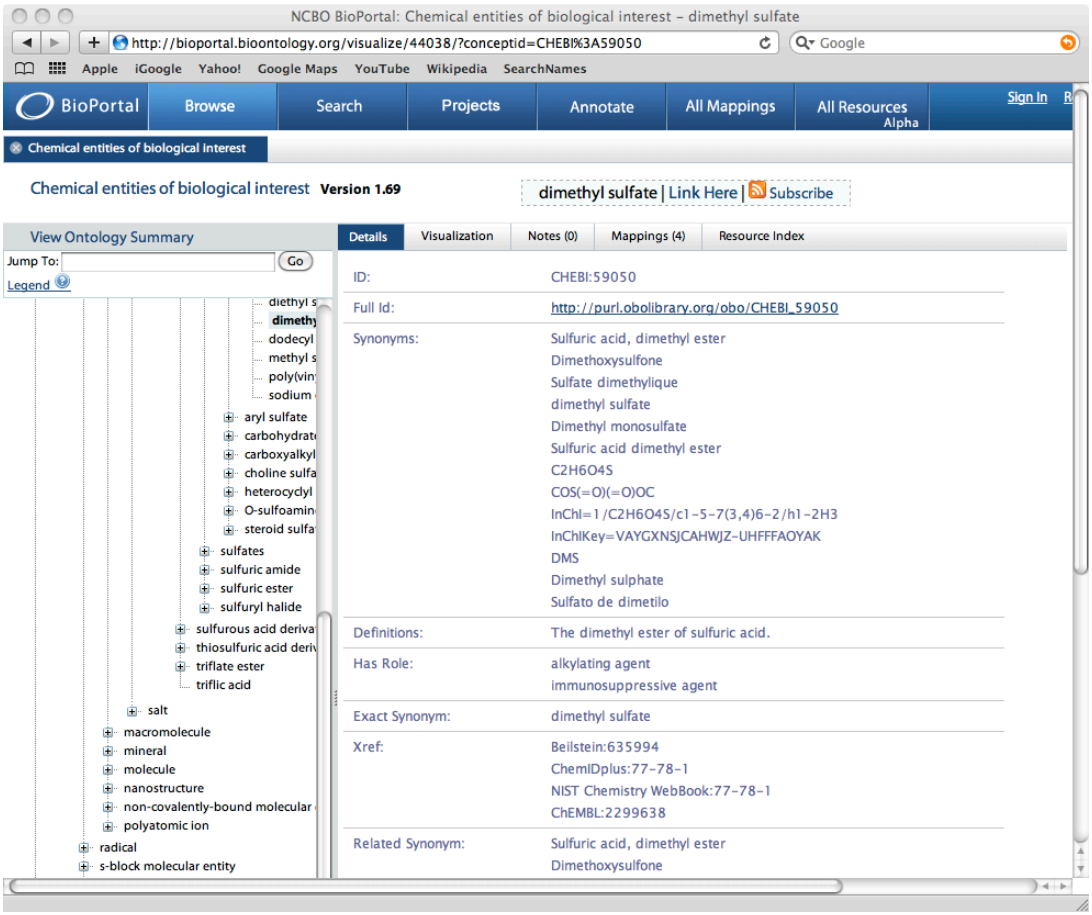


Figure 5: Bioportal lookup of DMS, indicating that the CHEBI ID for dimethyl sulfate is 59050. http://bioportal.bioontology.org/ is a central repository for finding terms to describe your assays. If you cannot find a term in the example files that describes a chemical or enzyme used in your experiments, searching bioportal is a good alternative.

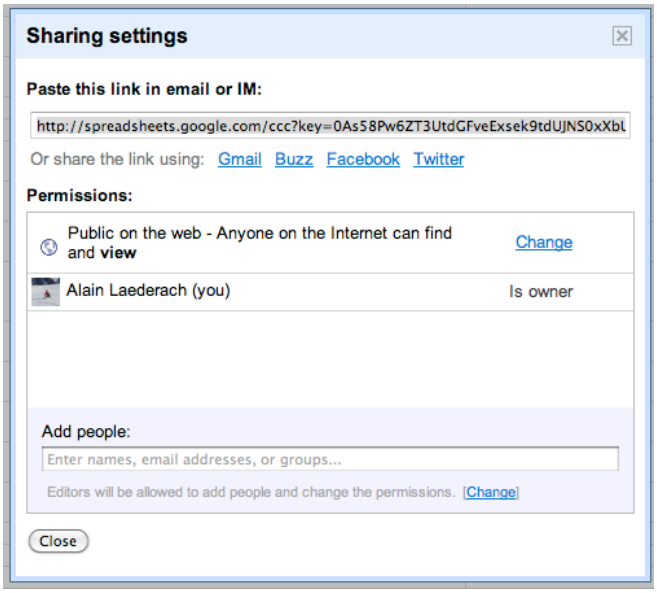


Figure 6: Sharing settings on Google Docs for an ISATab spreadsheet. In this case, we have made the data public on the Web, meaning anyone can see and download the file. These are the settings you want to use when you are ready to make your data available to the world. Note that in this case, people will not be able to edit your data, only you (or any other collaborators you choose) will be able to edit the document. The link to your data can also be copied from here to share with others or use in a URL shortening service like http://bit.ly.

Frequently Asked Questions

1.) How do I report errors in the data (i.e. multiple repeats or standard deviations)?

There are basically two types of error in SNRNASM data. To report error on the actual values, you can simply add an additional row to the assays tab (and corresponding column to the data matrix) for each assay and report it as "standard deviation," or error. You can also choose to report multiple repeats of the experiment, which will allow others to estimate error using their preferred statistical analysis.

To report error in the assignment of data to specific positions (i.e. you were only able to assign a reactivity to a range of positions), you can use a similar approach to above, where you repeat the data with alternative assignments in separate assay rows and corresponding data matrix columns. Finally, if no data was collected for a position, or there is too much uncertainty for that position, you can indicate this with a -999 reactivity in that cell.

2.) How should I align my data to the sequence?

The precise way in which each chemical probe or enzyme cleaves or modifies the nucleic acid backbone is different. For example, certain RNAses will cleave the nucleotide backbone 5' to a double-stranded nucleotide. Data should always be aligned to the sequence such that the data relevant to the structure of that particular nucleotide corresponds to that position in the sequence.

3.) How do I access the data in bulk without having to download excel files and convert them?

The Google Docs API (application program interface) http://code.google.com/apis/documents/ is a powerful tool for automatically accessing, searching and downloading data stored in the cloud. There is even a spreadsheet specific API (http://code.google.com/apis/spreadsheets/).

4.) How do I know if the data is a repeat measurement or a separate experiment?

Certain users may choose to repeat individual experiments rather than mean values. This data will be implicitly stored in the assays tab if all the study factors are identical, then the data is a repeat and not a separate experiment.