



# Metabolonote: a wiki-based database for managing hierarchical metadata of metabolome analyses

**Takeshi Ara<sup>1,2†</sup>, Mitsuo Enomoto<sup>1,2</sup>, Masanori Arita<sup>2,3</sup>, Chiaki Ikeda<sup>1,2</sup>, Kota Kera<sup>4</sup>, Manabu Yamada<sup>1,2</sup>, Takaaki Nishioka<sup>2,5</sup>, Tasuku Ikeda<sup>2,5</sup>, Yoshito Nihei<sup>2,5</sup>, Daisuke Shibata<sup>1</sup>, Shigehiko Kanaya<sup>2,5</sup> and Nozomu Sakurai<sup>1,2\*</sup>**

<sup>1</sup> Department of Technology Development, Kazusa DNA Research Institute, Kisarazu, Japan

<sup>2</sup> National Bioscience Database Center (NBDC), Japan Science and Technology Agency (JST), Tokyo, Japan

<sup>3</sup> RIKEN Center for Sustainable Resource Science, Yokohama, Japan

<sup>4</sup> Department of Research & Development, Kazusa DNA Research Institute, Kisarazu, Japan

<sup>5</sup> Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma, Japan

## Edited by:

Reza M. Salek, European  
Bioinformatics Institute, UK

## Reviewed by:

Subha Madhavan, Georgetown  
University, USA

Macha Nikolski, Bordeaux

Bioinformatics Center, France

Janna Hastings, European

Bioinformatics Institute, UK

## \*Correspondence:

Nozomu Sakurai, Metabolomics  
Team, Department of Technology  
Development, Kazusa DNA Research  
Institute, 2-6-7 Kazusa-kamatari,  
Kisarazu, Chiba 292-0818, Japan  
e-mail: sakurai@kazusa.or.jp

## \*Present address:

Takeshi Ara, KAGOME Tomato  
Discoveries Laboratory, Graduate  
School of Agriculture, Kyoto  
University, Kyoto, Japan

Metabolomics – technology for comprehensive detection of small molecules in an organism – lags behind the other “omics” in terms of publication and dissemination of experimental data. Among the reasons for this are difficulty precisely recording information about complicated analytical experiments (metadata), existence of various databases with their own metadata descriptions, and low reusability of the published data, resulting in submitters (the researchers who generate the data) being insufficiently motivated. To tackle these issues, we developed Metabolonote, a Semantic MediaWiki-based database designed specifically for managing metabolomic metadata. We also defined a metadata and data description format, called “Togo Metabolome Data” (TogoMD), with an ID system that is required for unique access to each level of the tree-structured metadata such as study purpose, sample, analytical method, and data analysis. Separation of the management of metadata from that of data and permission to attach related information to the metadata provide advantages for submitters, readers, and database developers. The metadata are enriched with information such as links to comparable data, thereby functioning as a hub of related data resources. They also enhance not only readers’ understanding and use of data but also submitters’ motivation to publish the data. The metadata are computationally shared among other systems via APIs, which facilitate the construction of novel databases by database developers. A permission system that allows publication of immature metadata and feedback from readers also helps submitters to improve their metadata. Hence, this aspect of Metabolonote, as a metadata preparation tool, is complementary to high-quality and persistent data repositories such as MetaboLights. A total of 808 metadata for analyzed data obtained from 35 biological species are published currently. Metabolonote and related tools are available free of cost at <http://metabolonote.kazusa.or.jp/>.

**Keywords:** metabolomics, metadata management system, Semantic MediaWiki, metadata sharing, data compatibility

## INTRODUCTION

Vast amounts of metabolome data are produced daily by high-throughput mass spectrometers; however, publication of these data and their reuse are limited. Metabolomics is a technological area that focuses on comprehensive detection of small molecules derived from the biosynthesis of living organisms (Oliver et al., 1998; Dunn, 2008). It is applied to a wide range of fields such as biology (Kopka et al., 2004; Khoo and Al-Rubeai, 2007; Mashego et al., 2007), exploration of biomarkers in medical sciences (Jansson et al., 2009; Koulman et al., 2009), quality evaluation of foods (Pongsuwan et al., 2008; Fitzgerald et al., 2009), and assessment of environmental pollution (Lin et al., 2006; Krauss et al., 2010). Mass spectrometry (MS) is one of the popular choices for compound detection because of its advantage in factors such as sensitivity and throughput (Zhang et al., 2012; Dunn and Hankemeier, 2013). However, in terms of publication of the experimental data in public

databases, metabolomics lags far behind the other “omics” such as transcriptome and proteome (Griffin and Steinbeck, 2010). Several metabolomics-specific reasons appear to be associated with this issue. First, a vast amount of metadata – detailed information about data – usually accompanies the metabolome data, which are obtained from complicated procedures and conditions of metabolite extraction, instrumental analysis, and computational processing. Therefore, standards for minimal information about metadata description are still under discussion (Fernie et al., 2011; Griffin et al., 2011). One major data generation bottleneck is identification of metabolites (Wishart, 2009; Bowen and Northen, 2010), which causes finalization of the data analysis to take an extended amount of time. During this period, a part of the complicated metadata may be lost or become difficult to trace. Second, many databases specially constructed for each purpose already exist, and the required metadata are different for each. Consequently,

submitters have to describe complicated metadata according to each metadata format, even when the data are derived from the same raw data. Finally, the most serious issue is that reuse of published metabolome data is quite limited because determination of the comparability of data for further analysis is difficult for database readers, owing to the complexity of the experimental procedures. These issues make general researchers less motivated to publish their own data in the databases.

The issues outlined above have to be resolved in order to improve publication and dissemination of metabolomic data. The standardization of minimum information for metadata description has been under discussion since 2005 by the Metabolomics Standards Initiative (Fiehn et al., 2007), and continued by COSMOS since 2012 (Steinbeck et al., 2012; Salek et al., 2013a); thus, standardization is likely to be resolved in the near future. Recent progress in metabolite annotation tools (Horai et al., 2010; Fukushima and Kusano, 2013; Kind et al., 2013; Hufsky et al., 2014) will accelerate the production of data. Use of metadata management tools such as ISACreator (Rocca-Serra et al., 2010) and XperimentR (Tomlinson et al., 2013) will reduce the risk of metadata loss. On the other hand, sharing of metadata is still problematic. ISA-Tab, which is designed for metadata sharing (Sansone et al., 2008; Rocca-Serra et al., 2010), is utilized in the metabolome data repository MetaboLights (Steinbeck et al., 2012; Haug et al., 2013). However, MetaboLights does not provide APIs for semantic search of the metadata. To reuse metadata from other systems, developers have to reconstruct the metadata from ISA-Tab files, which are retrieved in a synchronized manner from the MetaboLights website. MetabolomeXchange<sup>1</sup>, a portal website that gathers metabolome data resources published in databases such as MetaboLights, GMD (Kopka et al., 2005), and Metabolomics Workbench<sup>2</sup>, manages only limited items of metadata, with details remaining in the original databases. Therefore, cross-database searching for metadata details is not provided in the set of APIs they provide. The issue of metadata sharing is one of the reasons why determination of comparable data is still difficult for database readers. Consequently, no increase in the motivation of researchers to publicize data can be expected.

To overcome these fundamental issues in metabolomics, we developed a Semantic MediaWiki-based database called Metabolonote. We briefly mentioned Metabolonote in a previous article about our web portal KOMICS, in which we introduced the metabolomics tools and databases that we have developed (Sakurai et al., 2014). In this paper, we report on it in more detail. The ideas underlying the development of Metabolonote include complete separation of the management of metadata from that of data, and permission to attach related information to the metadata. Metabolonote is a metadata-specific database; thus, submitters (the majority of whom are assumed to be general researchers who generate the data and want to deposit and publicize them) can start describing complicated metadata without the substance of the experimental data. Further, an easy-to-use wiki system helps submitters to reduce loss of metadata. The metadata-specific database

is implemented with APIs for semantic searching and retrieving of metadata, and enables computational sharing of metadata among multiple databases without invading their own property values. Delegation of metadata management to Metabolonote facilitates the construction of novel databases by database developers. The permission to attach related information to the metadata enables submitters to publish their knowledge of other data that are comparable to their data. We believe that this will promote the reuse of metabolome data and motivate researchers to publish their data. Thus, these features are advantageous for all submitters, readers, and database developers.

## MATERIALS AND METHODS

Metabolonote is constructed as an extension called “Omic-snoteCore” for Semantic MediaWiki<sup>3</sup>, a content management system written in PHP, and a set of wiki page data (Properties, Templates, and Forms) prepared based on the “Togo Metabolome Data” (TogoMD) metadata format (described in the Section “The TogoMD Format”) (Figure 1). Functions related to form-editing, property displaying, and data storage in the extensions of MediaWiki<sup>4</sup>, Semantic MediaWiki, and Semantic Forms, are partially modified. The Metabolonote website is currently running on Red Hat Enterprise Linux Server release 5.6 with Apache 2.2, PHP 5.3, and MySQL 5.0. A sample program called “MNSearchDemo,” which includes practical examples of API usage for semantic searching and retrieving of metadata, is written in PHP. These programs and settings files are available free of cost at Metabolonote’s help page<sup>5</sup>.

## RESULTS

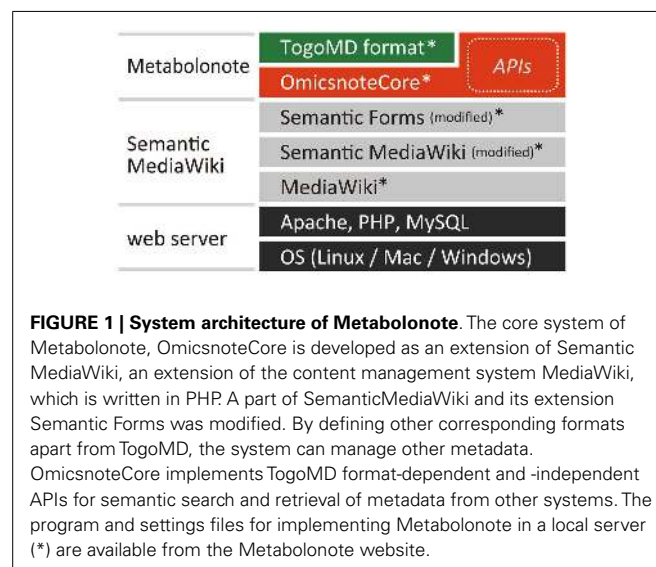
### SYSTEM ARCHITECTURE

We chose Semantic MediaWiki as the base system for Metabolonote because the following features satisfy our metadata

<sup>3</sup><http://semantic-mediawiki.org/>

<sup>4</sup><http://www.mediawiki.org/>

<sup>5</sup><http://metabolonote.kazusa.or.jp/Help:Contents>



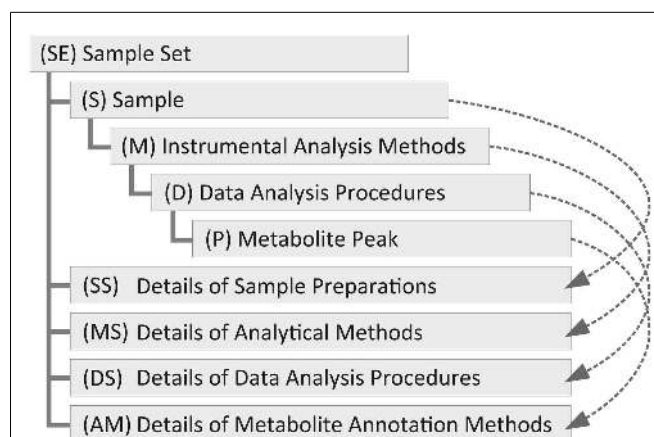
<sup>1</sup><http://metabolomexchange.org/>

<sup>2</sup><http://www.metabolomicsworkbench.org/>

management requirements: (1) The contents can be edited via a web browser. (2) The contents can be shared as webpages. (3) APIs for computationally searching and retrieving the contents can be implemented. (4) Users' inputs are managed as values of properties in a wiki page; this data structure is suitable for managing a set of metadata consisting of several defined items that need to be described. (5) The form-editing interface provided by the extension Semantic Forms is suitable for inputting the formatted metadata. (6) The concept of subpages in MediaWiki, in which a hierarchy of wiki pages is represented as a directory path by a URL separated by a slash "/", is suitable for representing the hierarchy of metadata (see The TogoMD Format) and to provide a way to uniquely access each metadata. Finally (7), additional information can be attached to the metadata in HTML format or MediaWiki's proprietary markup language. Therefore, we constructed the core system of Metabolonote, OmicsnoteCore, as an extension of Semantic MediaWiki (**Figure 1**). OmicsnoteCore provides functions for managing the hierarchical structure of metadata, user access controls, and APIs to search and retrieve the metadata. We defined a format "TogoMD" (see below) as a set of metadata for use on the OmicsnoteCore. However, OmicsnoteCore can be used to manage other metadata apart from TogoMD simply by submitting the definition of that set of metadata.

### THE TogoMD FORMAT

To manage the metadata in Metabolonote, we defined a novel metadata and data format called the TogoMD format. This novel format was conceptualized with the following considerations. The metabolomic metadata can be separated into hierarchical classes, specifically, a class for information about the purpose of the study, a class for sample preparation information, a class for analytical method information, and a class for data analysis information (**Figure 2**). In general, multiple samples are used per study; for example, biological replications of treatment and control groups. A sample can be analyzed in several ways and multiple times; for example, by liquid chromatography–MS, gas chromatography–MS, NMR, and their analytical replications. The raw data generated by each analysis can be analyzed in several ways. For instance, computational tools and procedures should be different when the data are used as fingerprints, for metabolite annotations to make metabolic profiling data, or used for getting tandem mass spectrum data. Each set of data generated through the process should be related to each class at a different level of the hierarchy. For instance, the raw data generated by the analytical apparatus should be related to the analytical method class, and the processed data should be related to the data analysis class. To share the metadata among outside systems, a method for uniquely accessing the metadata at each level of the hierarchy, possibly by unique identifiers (IDs), is required. We first evaluated the utility of ISA-Tab, which is used in MetaboLights (Sansone et al., 2008; Rocca-Serra et al., 2010). In ISA-Tab, the hierarchy of the classes mentioned above is connected with sample names and names assigned to the protocols (a set of common procedures). Therefore, control of the nomenclature of these names is required to establish a method to uniquely and computationally access the metadata at each level of the hierarchy, and modifications of the predefined format such as addition of an extra column for IDs is possibly required to



**FIGURE 2 | Hierarchy of the metadata classes of the TogoMD format.**

The metadata of metabolome analysis is divided into various classes; specifically, a class for the purpose of study information with a set of samples (SE), samples (S), analytical methods (M), and data analysis (D). These classes constitute a tree structure. We define classes for commonly used (shared) procedures for sample preparation details (SS), analytical method details (MS), data analysis details (DS), and annotation method details (AM) under the top-level class. Their instances are referred from the instances of the other classes (dashed arrows). The data class for peak information (P) is not used in Metabolonote. The parentheses are the prefixes of their instance ID.

control it well. Therefore, we defined TogoMD, which includes a rule for ID assignment (details are given in Systematic ID Design for the Metadata). Similar to the ISA-Tab protocol, frequently utilized procedures can be written as referent information. As each metadata has description and comment fields for free description by the submitter, almost the same metadata as those in ISA-Tab can be described in accordance with the recommendation of MSI (Fiehn et al., 2007) and also should be conducted in line with that of COSMOS (Steinbeck et al., 2012) Working Package 2 in the future. The relationships between the description fields of ISA-Tab and TogoMD used in Metabolonote are shown in Table S1 in Supplementary Material. Most of the fields in TogoMD, including author description, are simpler than those of ISA-Tab. TogoMD also defines formats for the data files. The details of the formats are described in the online help at the Metabolonote website<sup>6</sup>.

### SYSTEMATIC ID DESIGN FOR THE METADATA

To provide a simple way to uniquely access the metadata at each level of the hierarchy, an ID assignment rule for the metadata is a prerequisite in OmicsnoteCore. A metadata ID consists of alphabetic prefixes representing the class of the metadata, followed by a digit part representing the instance of the class, then an underscore connecting the instances. In the case of TogoMD, the following four classes and the prefixes shown in the parentheses are defined. The class for study purpose consists of a set of samples (SE) that corresponds to Study in ISA-Tab, with sample information (S), analytical methods (M), and data analysis (D) (**Figure 2**). A data class for the detected peaks (P) is

<sup>6</sup><http://metabolonote.kazusa.or.jp/TogoMetabolomeDataFormat>

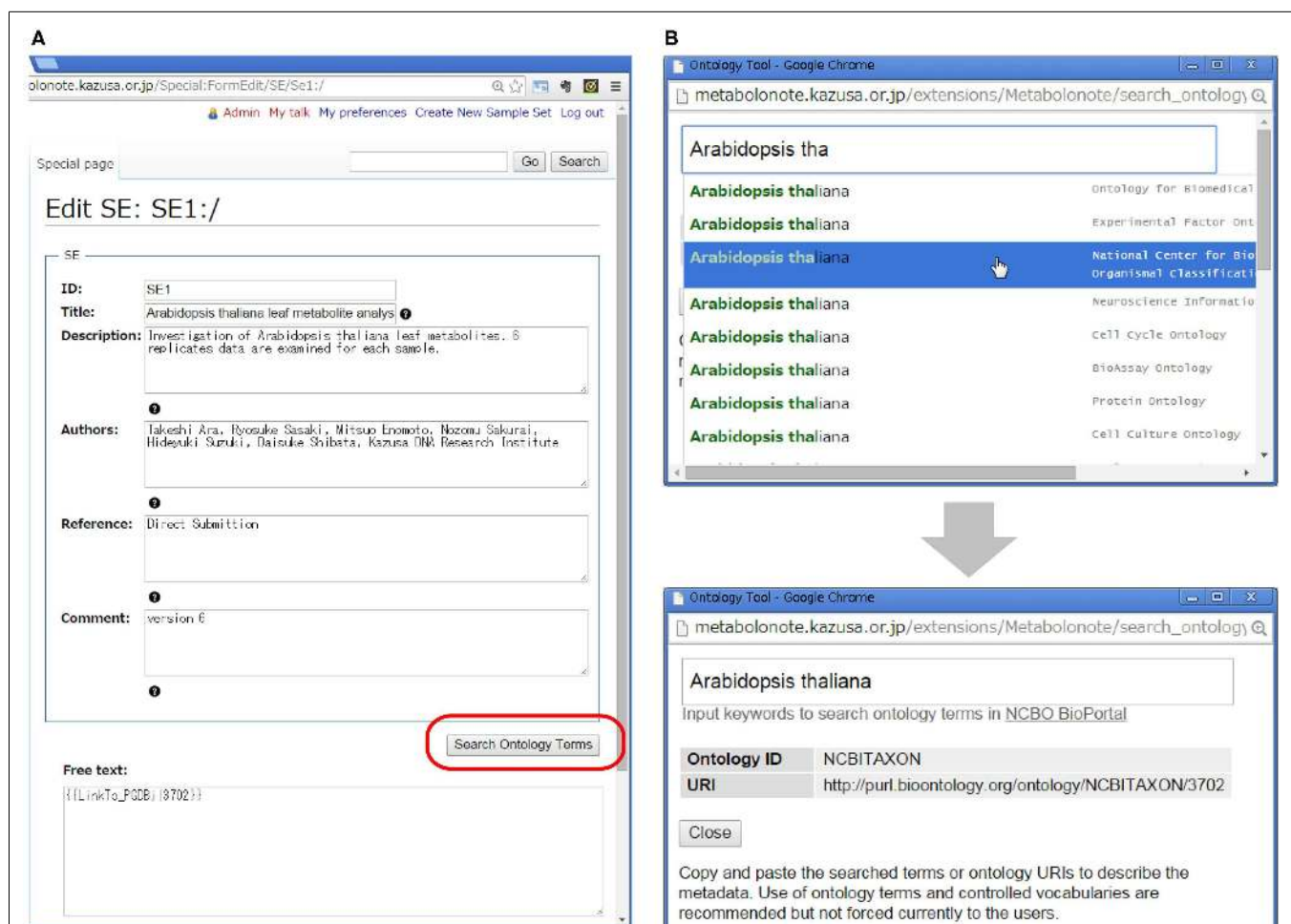
also defined subsequent to class D; however, it is not used in Metabolonote. As an instance, a metadata ID related to an analytical method is represented as SE1\_S01\_M01, and that for the processed data derived from the analysis data is represented as SE1\_S01\_M01\_D01. The instances of the classes for the metadata of frequently referred and shared procedures, specifically, detailed information about sample preparation (SS), analytical methods (MS), data analysis (DS), and metabolite annotation methods (AM), are placed beneath the instance of SE; for example, SE1\_MS1 (Figure 2, dashed arrows). The metadata are displayed in web browsers by accessing the URL with the IDs. In this case, the separators of the hierarchy level are replaced by a colon plus slash “:/” for the top-level class and a slash “/” for the rest. For example, <http://metabolonote.kazusa.or.jp/SE1:/S01/M01> for SE1\_S01\_M01. The digit part of the top-level instance (SE) is managed and assigned by the administrator of Metabolonote, whereas those of the other instances are assigned by the submitter themselves. Any array of digits can be assigned unless the same prefix plus array of digits are placed under the same metadata ID. The set of ID rules allows the submitters to assign meaningful digits to their own data. For example, they can assign S01–S03 for triplicates

of drug-treated samples, S11–S13 for the controls, and S91–S95 for authentic compounds. The only thing the administrator of the system has to do with regards to ID management is to control the uniqueness of the ID of the top-level instance. The concatenation of the metadata with the separators facilitates clear understanding of the hierarchy of the metadata. More detailed information about ID nomenclature is given on the Metabolonote help page<sup>7</sup>.

### EDITING AND PUBLICATION OF METADATA

Creation and editing of the metadata can be done via web browsers without installing any special software. Figure 3A shows a form-editing interface for an SE-instance with a minimized number of description items in TogoMD format. The submitters can search ontology terms in the BioPortal at the National Center for Biomedical Ontology (NCBO) (Whetzel et al., 2011) by clicking the “Search Ontology Terms” button (Figure 3B). Submitters currently do not have to use the proper ontology terms for fields such as species names, but the function helps to reduce incidents

<sup>7</sup>[http://metabolonote.kazusa.or.jp/Help/Data\\_Structure\\_ID](http://metabolonote.kazusa.or.jp/Help/Data_Structure_ID)



**FIGURE 3 | Form-editing window of Metabolonote. (A)** Metadata and attachment of additional information can be edited using forms. **(B)** The function “Search Ontology Terms,” available at the button represented by the red round circle in **(A)**, is used to search for ontology terms in BioPortal in NCBO.



of misspelling of specific terms, and helps to query the ID in the Taxonomy database of the National Center for Biotechnology Information (NCBI) (Federhen, 2012). In the “Free text” area, submitters can attach additional information in HTML or MediaWiki’s markup language. This area can be used, for example, to make links to the record of other databases in which data related to the metadata are deposited. Metabolonote provides several wiki templates to facilitate the creation of links by users. To prevent inappropriate writing and the creation of links to malicious websites, only system administrators are currently allowed to change the public/private state of the metadata.

### EXEMPLARY USAGE AS A HUB OF DATA RESOURCES

By adding extra information in the free text area, a metadata can become a hub with rich information about related data resources.

**Figure 4A** is an example of metadata of computationally analyzed data obtained from the liquid chromatography–MS analysis of leaves of a model plant, *Arabidopsis* (SE1\_S01\_M01\_D01). Several links to the other databases are attached. In Metabolonote, all of the metadata at the upper levels of the hierarchy are displayed as a single webpage. A link to PGDBj (Asamizu et al., 2014), which provides integrated information around plant genome data, is attached to the metadata of the sample set (SE). At the metadata of the samples, a link to the record of KomicMarket (Sakurai et al., 2014), the peak annotation database is present. For the analytical methods metadata, a link to the corresponding raw data in MassBase (Sakurai et al., 2014) is present. Links to two databases are attached to the data analysis procedures metadata; namely, Bio-MassBank<sup>8</sup>, the mass spectrum library from organisms, and KomicMarket2’s temporary website (KM2)<sup>9</sup>, which provides peak annotation data in TogoMD format. The metadata in Metabolonote is referred from the record pages in Bio-MassBank by a URL link, which complements the detailed description of metadata in Bio-MassBank. **Figure 4B** shows an attachment of a link to the other data resource that is comparable. **Figure 4C** is an example of the attachment of an image representing the process of data generation. Detailed information about the process in Metabolonote is referred as a Supplementary Material from Kera et al. (2014). A link to the article website is also attached. **Figure 4D** is another example of attachment of pictures for aiding understanding about the analyzed samples. This information aids in the reuse of published data by the readers. As shown in **Figure 4**, the metadata prepared by the submitters becomes a hub of related data resources, and give rich information to the readers.

### APIs FOR SEARCHING AND RETRIEVING THE METADATA

Metabolonote provides APIs for searching semantically and retrieving the metadata, by which computational sharing of metadata from the other systems is realized. The description of items in a metadata class is designated as “property” in Semantic MediaWiki. The values of the property can be searched by the APIs. Two APIs for semantic search (search for properties that have specific values) and 12 for data retrieval are currently provided. The “Metadata Search” function in Metabolonote was developed by

internally using these APIs. The KM2 website, which distributes the peak annotation data in TogoMD format, was also developed using the APIs. No metadata management system is implemented in KM2, and KM2 only manages the data with their corresponding metadata IDs. When the metadata are searched for by the semantic search APIs, a list of metadata IDs is returned. Then, KM2 displays the results for those data whose IDs are on the list. Sharing of metadata via APIs significantly aids the database developers by releasing them from designing and developing a metadata management system for their own web applications. A sample PHP program called “MNSearchDemo” is available at the Metabolonote website. MNSearchDemo is a practical example of API usage, and by which developers can create a simple database like KM2 in local servers.

### CURRENT STATISTICS OF METADATA

The metadata for 808 sets of analysis data (D) obtained from 420 experimental analyses (M) of 149 samples (S) examined in 51 sample sets (SE) consist of 35 biological species, including plants, algae, bacteria, and animals, are currently published in Metabolonote. Six analytical methods (MS) have links to the comparable data. Links to the eight other databases, namely, MassBase, KomicMarket, Bio-MassBank, PGDBj, KM2, PRIME (Sakurai et al., 2013), MetaboLights, and MetabolomeExpress (Carroll et al., 2010), are included.

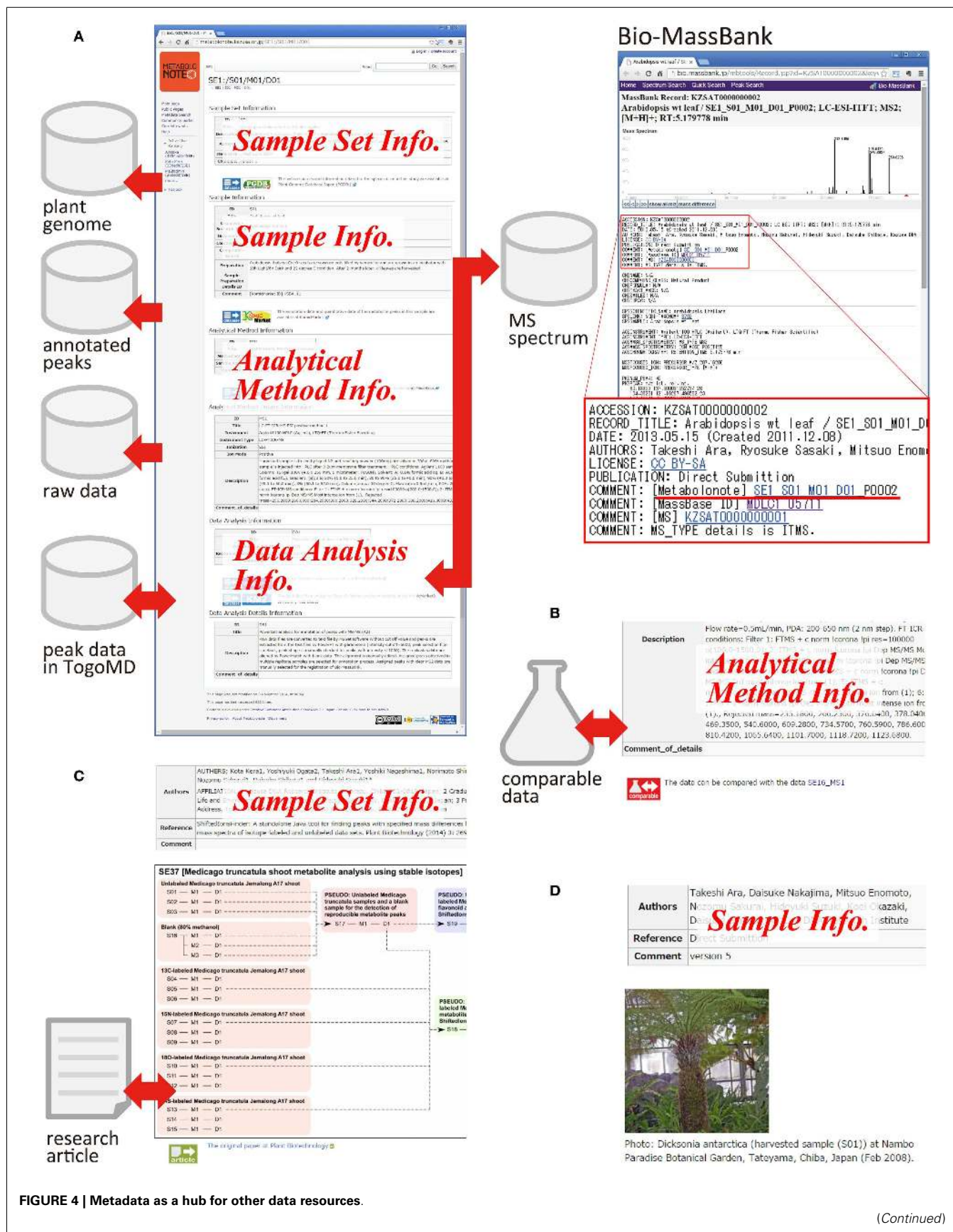
## DISCUSSION

### BENEFITS OF METADATA-CENTRIC RESOURCE MANAGEMENT

By separating the management of metadata from that of data, and permitting the attachment of related information to the metadata, Metabolonote provides benefits to all submitters (who are assumed to be general researchers who generate and deposit their data), readers, and database developers. The submitters become motivated to publish their data by attaching additional information that makes their work attractive. Piwowar et al. (2007) surmised that increases in the citation of research articles for which the data are published can be attributed to increase exposure of the work. However, their statement is based on microarray data, which is more easily comparable than metabolome data. Even when standardized descriptions are obligated by the journals (Salek et al., 2013b), issues of data comparability remain because of the complexity of experimental conditions. It is not practicable to annotate all comparability between data stored in data repositories. Thus, provision of comparability information by the submitters themselves is one practical solution to this problem that can lead to increased use of their data and citation of their work. When rich information is provided by submitters, it is advantageous for the readers to understand about the data and to find related data resources, especially comparable data, by which reuse of metabolome data will be promoted. For the database developers, delegation of the metadata management system to Metabolonote is advantageous as it aids in the development of novel databases like KM2. If the metadata sharing became common, submitters would be released from repeated description and management of their metadata in multiple databases. Metadata that are described in a uniform format helps readers to understand the data. Separate management of metadata and experimental raw data, and then

<sup>8</sup><http://bio.massbank.jp/>

<sup>9</sup>[http://nedo3.kazusa.or.jp/km2\\_tmp/](http://nedo3.kazusa.or.jp/km2_tmp/)



**FIGURE 4 | Continued**

Examples of the metadata are shown: **(A)** Metadata for processed data obtained from *Arabidopsis* (SE1\_S01\_M01\_D01, the original metadata are available at <http://metabolonote.kazusa.or.jp/SE1/S01/M01/D01>). A reciprocal link is made with the mass spectrum database Bio-MassBank. **(B)** Metadata with a link to the comparable data (SE12\_MS1, the original metadata are

available at <http://metabolonote.kazusa.or.jp/SE12/MS1>). **(C)** Metadata with images representing analytical procedures (SE37, the original metadata are available at <http://metabolonote.kazusa.or.jp/SE37/>). The metadata are used as Supplementary Material in the research article. **(D)** Metadata with a picture of the sample (SE4, the original metadata are available at <http://metabolonote.kazusa.or.jp/SE4/>).

accessibility to them, may also help an increasing requirement for data sharing in the biomedical research field, where unwillingness to share data is of major issues (Piwowar et al., 2008; Wang et al., 2014).

**COMPLEMENTARY USAGE WITH THE OTHER REPOSITORIES**

The aspect of Metabolonote as a support system for completion of metadata is complementary to other curated, high-quality, and persistent repositories such as MetaboLights. Metabolonote allows submitters to publish immature descriptions of metadata. This promotes publication of metabolome data that were not previously considered publishable owing to inadequateness of the metadata. Depending on the purpose of a study, the required metadata for the published data should be different. For example, in the case of analysis of species-specific MS/MS spectrum, information about species and the type of MS would be sufficient if the data are abundantly provided. Conversely, in the case of comparison of metabolic profiles, further detailed metadata such as extraction and chromatography separation are essential. Permission to publish immature metadata is advantageous for submitters as it allows them to get feedback or requests from readers for more detailed metadata. Providing all of their data with completed metadata is a burden for submitters. However, when they receive requests from readers who are interested in their work, they can focus on curation and enrichment of the metadata for the requested data. Such requests from readers should help to significantly motivate the submitters. Further, if metadata improved through these processes become a sufficiently high-quality, then the submitters can deposit them to MetaboLights. They can make a link to MetaboLights from the metadata in Metabolonote, and readers can then judge for themselves the quality of the metadata.

**METADATA-SPECIFIC SYSTEM**

To the best of our knowledge, Metabolonote is the first database specified for managing metabolomics metadata. Most life science databases manage the metadata along with the data. There are several metadata-specific systems, such as NCBI BioProject (Barrett et al., 2012) and GEOmetadb (Zhu et al., 2008). These are aimed at managing vast amounts of data that have been deposited. BioSamples Database (Gostev et al., 2012) is a metadata-specified system for biological samples and was developed to connect multiple-omics data in the European Bioinformatics Centre. The concept of Metabolonote in metadata-centric management of various data is similar to that of BioSamples. However, BioSamples does not support description and sharing of the metadata of metabolome experiments, and is not targeted at authentic compounds that can be analyzed in metabolite identification. Other metadata-centric systems can be found in the field of digital library management.

Free software such as DSPACE<sup>10</sup> and Islandora<sup>11</sup> are well known, but our investigations reveal that there is no system that supports hierarchical structured metadata. A portal for metabolomics databases, MetabolomeXchange is a metadata-specific system because it does not manage experimental data. However, as mentioned in the Introduction, cross-database searching by detailed metadata is not yet supported. Therefore, from all appearances, Metabolonote is a novel database that manages rich metadata separately from the management of experimental data, and is aimed at the sharing of hierarchical structured metadata. As any kind of data can be related to the metadata, OmicsnoteCore may be highly suited to manage metadata from multiple-omics analyses and study fields where novel analytical methods are frequently applied. OmicsnoteCore is also used as a laboratory data management system in some laboratories by defining other metadata formats apart from TogoMD.

**FUTURE PERSPECTIVES**

To assist knowledge discoveries through intensive analysis of vast amounts of metabolome data, further improvements are required. For advanced use of metabolome data, detailed facet search with various parameters such as parts of the body, sampling environment, type of chromatography column, and name of data analysis tool, will be essential in the future. However, to realize this we have to tackle a contradictory matter, namely, subdivision of metadata description items and their efficient collection. We are currently in discussions about a format that will extend TogoMD to describe sub-items in the fields for “description” and “comment.” The description of sub-items should not be mandatory but should be easy to fill in by the submitters or by the article curators. An easy-to-use and intuitive user interface should also be developed. Another requirement in order to realize facet searching is preparation of semantic web technologies. Semantic search APIs are already implemented in Metabolonote. However, for advanced search across resources on other databases, the metadata should be queried by a semantic query language such as SPARQL<sup>12</sup>. For this purpose, preparation of the metadata in Resource Description Framework (RDF)<sup>13</sup> and preparation of a set of necessary ontologies and controlled vocabularies by using pre-existing ones such as Mass Spectrum ontology<sup>14</sup> are currently underway. As mentioned in the Section “Complementary Usage with the Other Repositories,” requests and feedback from readers help submitters to improve the quality of the metadata. A system for efficient

<sup>10</sup><http://www.dspace.org/>

<sup>11</sup><http://islandora.ca/>

<sup>12</sup><http://www.w3.org/TR/sparql11-overview/>

<sup>13</sup><http://www.w3.org/TR/rdf11-concepts/>

<sup>14</sup><http://bioportal.bioontology.org/ontologies/MS>

communication between submitters and readers is therefore being considered.

## ACKNOWLEDGMENTS

This work was supported in part by the National Bioscience Database Center (NBDC) of the Japan Science and Technology Agency (JST), and in part by the Kazusa DNA Research Institute Foundation.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00038>

## REFERENCES

- Asamizu, E., Ichihara, H., Nakaya, A., Nakamura, Y., Hirakawa, H., Ishii, T., et al. (2014). Plant genome database Japan (PGDBj): a portal website for the integration of plant genome-related databases. *Plant Cell Physiol.* 55, e8. doi:10.1093/pcp/pct189
- Barrett, T., Clark, K., Gevorgyan, R., Gorenkov, V., Gribov, E., Karsch-Mizrachi, I., et al. (2012). BioProject and bioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids Res.* 40, D57–D63. doi:10.1093/nar/gkr1163
- Bowen, B. P., and Northen, T. R. (2010). Dealing with the unknown: metabolomics and metabolite atlases. *J. Am. Soc. Mass Spectrom.* 21, 1471–1476. doi:10.1016/j.jasms.2010.04.003
- Carroll, A. J., Badger, M. R., and Harvey Millar, A. (2010). The metabolome-express project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. *BMC Bioinformatics* 11:376. doi:10.1186/1471-2105-11-376
- Dunn, W. B. (2008). Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys. Biol.* 5, 011001. doi:10.1088/1478-3975/5/1/011001
- Dunn, W. B., and Hankemeier, T. (2013). Mass spectrometry and metabolomics: past, present and future. *Metabolomics* 9, 1–3. doi:10.1007/s11306-013-0507-z
- Federhen, S. (2012). The NCBI taxonomy database. *Nucleic Acids Res.* 40, D136–D143. doi:10.1093/nar/gkr1178
- Fernie, A. R., Aharoni, A., Willmitzer, L., Stitt, M., Tohge, T., Kopka, J., et al. (2011). Recommendations for reporting metabolite data. *Plant Cell* 23, 2477–2482. doi:10.1105/tpc.111.086272
- Fiehn, O., Robertson, D., Griffin, J., Van Der Werf, M., Nikolau, B., Morrison, N., et al. (2007). The metabolomics standards initiative (MSI). *Metabolomics* 3, 175–178. doi:10.1007/s11306-007-0070-6
- Fitzgerald, M. A., McCouch, S. R., and Hall, R. D. (2009). Not just a grain of rice: the quest for quality. *Trends Plant Sci.* 14, 133–139. doi:10.1016/j.tplants.2008.12.004
- Fukushima, A., and Kusano, M. (2013). Recent progress in the development of metabolome databases for plant systems biology. *Front. Plant Sci.* 4:73. doi:10.3389/fpls.2013.00073
- Gostev, M., Faulconbridge, A., Brandizi, M., Fernandez-Banet, J., Sarkans, U., Brazma, A., et al. (2012). The biosample database (BioSD) at the European bioinformatics institute. *Nucleic Acids Res.* 40, D64–D70. doi:10.1093/nar/gkr937
- Griffin, J. L., Atherton, H. J., Steinbeck, C., and Salek, R. M. (2011). A Metadata description of the data in “A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human”. *BMC Res Notes* 4:272. doi:10.1186/1756-0500-4-272
- Griffin, J. L., and Steinbeck, C. (2010). So what have data standards ever done for us? The view from metabolomics. *Genome Med.* 2, 38. doi:10.1186/gm159
- Haug, K., Salek, R. M., Conesa, P., Hastings, J., De Matos, P., Rijnbeek, M., et al. (2013). MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* 41, D781–D786. doi:10.1093/nar/gks1004
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 45, 703–714. doi:10.1002/jms.1777
- Hufsky, F., Scheubert, K., and Böcker, S. (2014). New kids on the block: novel informatics methods for natural product discovery. *Nat. Prod. Rep.* 31, 807–817. doi:10.1039/c3np70101h
- Jansson, J., Willing, B., Lucio, M., Fekete, A., Dicksved, J., Halfvarson, J., et al. (2009). Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE* 4:e6386. doi:10.1371/journal.pone.0006386
- Kera, K., Ogata, Y., Ara, T., Nagashima, Y., Shimada, N., Sakurai, N., et al. (2014). ShiftedIonsFinder: a standalone Java tool for finding peaks with specified mass differences by comparing mass spectra of isotope-labeled and unlabeled data sets. *Plant Biotechnol.* 31, 269–274. doi:10.5511/plantbiotechnology.14.0609c
- Khoo, S. H., and Al-Rubeai, M. (2007). Metabolomics as a complementary tool in cell culture. *Biotechnol. Appl. Biochem.* 47, 71–84. doi:10.1042/BA20060221
- Kind, T., Liu, K. H., Lee Do, Y., Defelice, B., Meissen, J. K., and Fiehn, O. (2013). LipidBlast *in silico* tandem mass spectrometry database for lipid identification. *Nat. Methods* 10, 755–758. doi:10.1038/nmeth.2551
- Kopka, J., Fernie, A., Weckwerth, W., Gibon, Y., and Stitt, M. (2004). Metabolite profiling in plant biology: platforms and destinations. *Genome Biol.* 5, 109. doi:10.1186/gb-2004-5-6-109
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., et al. (2005). GMD@CSB.DB: the Golm metabolome database. *Bioinformatics* 21, 1635–1638. doi:10.1093/bioinformatics/bti236
- Koulman, A., Lane, G. A., Harrison, S. J., and Volmer, D. A. (2009). From differentiating metabolites to biomarkers. *Anal. Bioanal. Chem.* 394, 663–670. doi:10.1007/s00216-009-2690-3
- Krauss, M., Singer, H., and Hollender, J. (2010). LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal. Bioanal. Chem.* 397, 943–951. doi:10.1007/s00216-010-3608-9
- Lin, C. Y., Viant, M. R., and Tjeerdema, R. S. (2006). Metabolomics: methodologies and applications in the environmental sciences. *J. Pestic. Sci.* 31, 245–251. doi:10.1584/jpestics.31.245
- Mashego, M. R., Rumbold, K., De Mey, M., Vandamme, E., Soetaert, W., and Heijnen, J. J. (2007). Microbial metabolomics: past, present and future methodologies. *Biotechnol. Lett.* 29, 1–16. doi:10.1007/s10529-006-9218-0
- Oliver, S. G., Winson, M. K., Kell, D. B., and Bagan, F. (1998). Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 16, 373–378. doi:10.1016/S0167-7799(98)01214-1
- Piwowar, H. A., Becich, M. J., Bilofsky, H., and Crowley, R. S. (2008). Towards a data sharing culture: recommendations for leadership from academic health centers. *PLoS Med.* 5:e183. doi:10.1371/journal.pmed.0050183
- Piwowar, H. A., Day, R. S., and Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2:e308. doi:10.1371/journal.pone.0000308
- Pongsuwan, W., Bamba, T., Harada, K., Yonetani, T., Kobayashi, A., and Fukusaki, E. (2008). High-throughput technique for comprehensive analysis of Japanese green tea quality assessment using ultra-performance liquid chromatography with time-of-flight mass spectrometry (UPLC/TOF MS). *J. Agric. Food Chem.* 56, 10705–10708. doi:10.1021/jf8018003
- Rocca-Serra, P., Brandizi, M., Maguire, E., Sklyar, N., Taylor, C., Begley, K., et al. (2010). ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics* 26, 2354–2356. doi:10.1093/bioinformatics/btq415
- Sakurai, N., Ara, T., Enomoto, M., Motegi, T., Morishita, Y., Kurabayashi, A., et al. (2014). Tools and databases of the KOMICS web portal for preprocessing, mining, and dissemination of metabolomics data. *Biomed Res. Int.* 2014, 194812. doi:10.1155/2014/194812
- Sakurai, T., Yamada, Y., Sawada, Y., Matsuda, F., Akiyama, K., Shinozaki, K., et al. (2013). PRIME Update: innovative content for plant metabolomics and integration of gene expression and metabolite accumulation. *Plant Cell Physiol.* 54, e5. doi:10.1093/pcp/pcs184
- Salek, R. M., Haug, K., and Steinbeck, C. (2013a). Dissemination of metabolomics results: role of MetaboLights and COSMOS. *Gigascience* 2, 8. doi:10.1186/2047-217X-2-8
- Salek, R. M., Steinbeck, C., Viant, M. R., Goodacre, R., and Dunn, W. B. (2013b). The role of reporting standards for metabolite annotation and identification in metabolomic studies. *Gigascience* 2, 13. doi:10.1186/2047-217X-2-13
- Sansone, S. A., Rocca-Serra, P., Brandizi, M., Brazma, A., Field, D., Fostel, J., et al. (2008). The first RSBI (ISA-TAB) workshop: “Can a simple format work for complex studies?” *OMICS* 12, 143–149. doi:10.1089/omi.2008.0019
- Steinbeck, C., Conesa, P., Haug, K., Mahendrakar, T., Williams, M., Maguire, E., et al. (2012). MetaboLights: towards a new COSMOS of metabolomics data management. *Metabolomics* 8, 757–760. doi:10.1007/s11306-012-0462-0



- Tomlinson, C. D., Barton, G. R., Woodbridge, M., and Butcher, S. A. (2013). XperimentR: painless annotation of a biological experiment for the laboratory scientist. *BMC Bioinformatics* 14:8. doi:10.1186/1471-2105-14-8
- Wang, F., Vergara-Niedermayr, C., and Liu, P. (2014). Metadata based management and sharing of distributed biomedical data. *Int. J. Metadata Semant. Ontol.* 9, 42–57. doi:10.1504/IJMSO.2014.059126
- Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., et al. (2011). BioPortal: enhanced functionality via new Web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic Acids Res.* 39, W541–W545. doi:10.1093/nar/gkr469
- Wishart, D. S. (2009). Computational strategies for metabolite identification in metabolomics. *Bioanalysis* 1, 1579–1596. doi:10.4155/bio.09.138
- Zhang, A., Sun, H., Wang, P., Han, Y., and Wang, X. (2012). Modern analytical techniques in metabolomics analysis. *Analyst* 137, 293–300. doi:10.1039/c1an15605e
- Zhu, Y., Davis, S., Stephens, R., Meltzer, P. S., and Chen, Y. (2008). GEOmetadb: powerful alternative search engine for the gene expression omnibus. *Bioinformatics* 24, 2798–2800. doi:10.1093/bioinformatics/btn520

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 November 2014; accepted: 13 March 2015; published online: 07 April 2015.

Citation: Ara T, Enomoto M, Arita M, Ikeda C, Kera K, Yamada M, Nishioka T, Ikeda T, Nihei Y, Shibata D, Kanaya S and Sakurai N (2015) Metabolonote: a wiki-based database for managing hierarchical metadata of metabolome analyses. *Front. Bioeng. Biotechnol.* 3:38. doi: 10.3389/fbioe.2015.00038

This article was submitted to *Bioinformatics and Computational Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2015 Ara, Enomoto, Arita, Ikeda, Kera, Yamada, Nishioka, Ikeda, Nihei, Shibata, Kanaya and Sakurai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.