

## ORIGINAL ARTICLE

# Bioinformatics advances in saliva diagnostics

Ji-Ye Ai<sup>1</sup>, Barry Smith<sup>2</sup> and David TW Wong<sup>1</sup>

**There is a need recognized by the National Institute of Dental & Craniofacial Research and the National Cancer Institute to advance basic, translational and clinical saliva research. The goal of the Salivaomics Knowledge Base (SKB) is to create a data management system and web resource constructed to support human salivaomics research. To maximize the utility of the SKB for retrieval, integration and analysis of data, we have developed the Saliva Ontology and SDxMart. This article reviews the informatics advances in saliva diagnostics made possible by the Saliva Ontology and SDxMart.**

*International Journal of Oral Science* (2012) 4, 85–87; doi:10.1038/ijos.2012.26; published online 15 June 2012

**Keywords:** BioMart; database; ontology; saliva

### SALIVA RESEARCH

The need to advance saliva research is strongly recognized by the Strategic Plan of the National Institute of Dental and Craniofacial Research.<sup>1</sup> The National Cancer Institute has also recognized saliva as a promising cancer biomarker source.<sup>2</sup> The ability to monitor health status, disease onset, progression, recurrence and treatment outcome through non-invasive means is highly important to advancing health care management. Saliva (oral fluid) is a perfect medium to be explored, offering the potential for a non-invasive, easy to obtain means for detecting and monitoring disease.

The adoption of saliva testing would allow a patient to collect their own specimens at home, yielding savings in health costs, convenience for the patient and facilitating multiple sampling. Specimen collection is less objectionable to patients than in the case of other bodily fluids and easier in children and older individuals. The analysis of saliva can thus provide a cost-effective approach for the screening of large populations. Due to these significant advantages, developing biomarkers in saliva for the detection of serious illnesses such as oral and systemic cancers has been on the national healthcare agenda for several years (Government Performance & Results Act 2008).<sup>3</sup> One mandate formulated in the Government Performance & Results Act report is that by the year 2013 proof of principle will be obtained for the ability of saliva to monitor health and diagnose for one systemic disease.

### CHALLENGES AND OPPORTUNITIES

A vast amount of saliva omics data has been generated by recent studies using high throughput technologies.<sup>4–7</sup> However, there are still barriers which researchers must overcome before such data can be exploited, such as lack of computationally accessible salivary data and information, and inability to cross-reference the salivaomics data that could potentially be made available through different proteomics,

transcriptomics, genomics and metabolomics studies. For these reasons, there is an urgent need to create the Salivaomics Knowledge Base (SKB), a data management system and web resource constructed to support saliva diagnostics research, and we present below the informatics advances brought about through the SKB and through the associated tools and resources.

### BIOMEDICAL ONTOLOGY

Ontologies are controlled structured vocabularies designed to provide consensus-based means to ensure consistent description of data by scientists working in disparate domains. As applied in the biomedical domain, ontology plays a key role in providing consensus-based controlled vocabularies serving the consistent annotation of biological and medical data and information, most conspicuously within the framework of the Gene Ontology<sup>8</sup> and now of its sister ontologies within the Open Biomedical Ontologies Foundry (<http://obofoundry.org>).

The Basic Formal Ontology (BFO) is a formal ontological framework developed by Barry Smith, Pierre Grenon and others, which serves as the starting point for some 100 ontology projects primarily in the biomedical domain (<http://www.ifomis.uni-saarland.de/bfo/>). The BFO framework can be readily extended to the treatment of families of ontologies of other types, above all to the treatment of relations between ontologies of different levels of granularity, from genes to species and from a single patient to epidemics at a geographical scale (combining applications of BFO to the medical and to the geographical domain). The framework may also be used as a tool for dealing with the relations between distinct perspectives on the biomedical domain, including culturally generated perspectives of the sort which are studied by linguists and anthropologists.<sup>9</sup>

<sup>1</sup>University of California at Los Angeles School of Dentistry and Dental Research Institute, Center for Health Sciences, University of California, Los Angeles, USA and <sup>2</sup>Department of Philosophy, University at Buffalo, Buffalo, USA

Correspondence: Dr DTW Wong, University of California at Los Angeles School of Dentistry and Dental Research Institute, 73-017 Center for Health Sciences, 10833 Le Conte Avenue, University of California, Los Angeles, CA 90095-1668, USA

E-mail: dtww@ucla.edu

Dr JY Ai, Department of Philosophy, 135 Park Hall, University at Buffalo, Buffalo, NY 14260, USA

E-mail: jiyelai@gmail.com

Received 1 March 2012; accepted 26 April 2012

Two BFO-based ontologies of special significance for our work here are the Ontology for Biomedical Investigations (OBI)<sup>10</sup> and the Ontology for General Medical Sciences. OBI addresses the need for a controlled vocabulary to support integration of experimental data. The OBI is an ontology designed to serve the coordinated representation of designs, protocols, instrumentation, materials, processes, data and types of analysis in all areas of biological and biomedical investigation. Ontology for General Medical Science is an ontology of the entities involved in the clinical encounter. Thus, it includes very general terms that are used across medical disciplines, including: 'disease', 'disorder', 'disease course', 'diagnosis', 'patient' and 'health-care provider'.<sup>11</sup>

### A DENTAL RESEARCH ONTOLOGY CONSORTIUM

To advance the consistency of data in the dental research community, Smith *et al.*<sup>12</sup> propose an approach to building a consensus-based ontology to support dental research (ODR). In analogy to efforts in other fields,<sup>11</sup> a consortium of research groups specializing in different areas of study would undertake such an effort, each building different components of ontology to support dental research. Initial efforts in this direction, by scientists in dental research and biomedical ontology at University at Buffalo and University of California, include work on the ontology of oral pathology, oral maxillofacial anatomy, dental disease and dental procedures, and as we discuss below, the Saliva Ontology.<sup>13</sup>

Integral to his work is a plan to allow a seamless connection between the use of ontology to support dental research in the dental domain and the use of existing ontology resources developed in other areas of biology and medicine, by reusing elements and strategies from them. The anatomy work is based on the Foundational Model of Anatomy. The ontology of dental procedures use the framework established by OBI.<sup>14</sup> The work on dental diseases is carried out in conjunction with the development of Ontology for General Medical Science.<sup>11</sup>

### SALIVA ONTOLOGY

The Saliva Ontology (SALO) (Figure 1) is a detailed ontology of this bodily fluid that is optimized to meet the needs of both the clinical diagnostic community and the cross-disciplinary community of omics researchers. The SALO is created through cross-disciplinary interaction with saliva experts, protein experts, diagnosticians and ontologists. To aid development and testing of SALO, we develop a corpus of saliva-relevant literature in SKB to assist in characterizing core terms and synonyms within the ontology and to provide links between SALO

content and relevant items in PubMed. SKB will also incorporate the results of experiments in data and text mining using the ontology.

SALO will incorporate links to existing ontologies and terminology resources involving treatment of saliva-relevant phenomena. We will also identify and represent within SALO relationships to saliva-relevant types represented in ontologies such as the Gene Ontology, the Protein Ontology<sup>15</sup> and the Chemical Entities of Biological Interest ontology,<sup>16</sup> and also provide links to corresponding SNOMED CT terms where available. SALO is a public domain resource and entirely web-based. Each term in the ontology has its own URL which points to a webpage providing definitions, PubMed sources, references to annotations to SKB and to external databases. SALO and Blood Ontology<sup>17</sup> are the foundation for a unified body fluids ontology resource—Body Fluids Ontology.<sup>18</sup>

### BIOMART

BioMart is a free, open-source, federated database system. It is cross-platform and supports many popular relational database management systems, including MySQL, Oracle, PostgreSQL, SQL Server and DB2. The software is data-agnostic, and can therefore be easily adapted to existing data sets. It is expandable and customizable through a plug-in system, and is open-source so the community can participate in deeper development. Furthermore, BioMart can seamlessly connect geographically disparate databases, facilitating collaboration between different groups. These features have catalyzed the creation of BioMart Central Portal, a first of its kind community-supported effort to create a single access point integrating many different, independently administered biological databases. Anybody can contribute an independently maintained resource to the Central Portal, allowing it to be exposed to and shared with the research community, and linking it with the other resources in the portal. Users can take advantage of the common interface to quickly utilize different sources without learning a new system for each. The system also simplifies cross-database searches that might otherwise require several complicated steps. Several integrated tools streamline common tasks, such as converting between ID formats and retrieving sequences. The combination of a wide variety of databases, an easy-to-use interface, robust programmatic access and the array of tools make Central Portal a one-stop shop for biological data querying.<sup>19</sup>

### SDXMART—A BIOMART PORTAL FOR SALIVAOMICS DATA

SDxMart is a BioMart data portal that hosts salivary proteomic, transcriptomic, metabolomic and microRNA data and offers access to the data by using the BioMart interface and querying environment. The

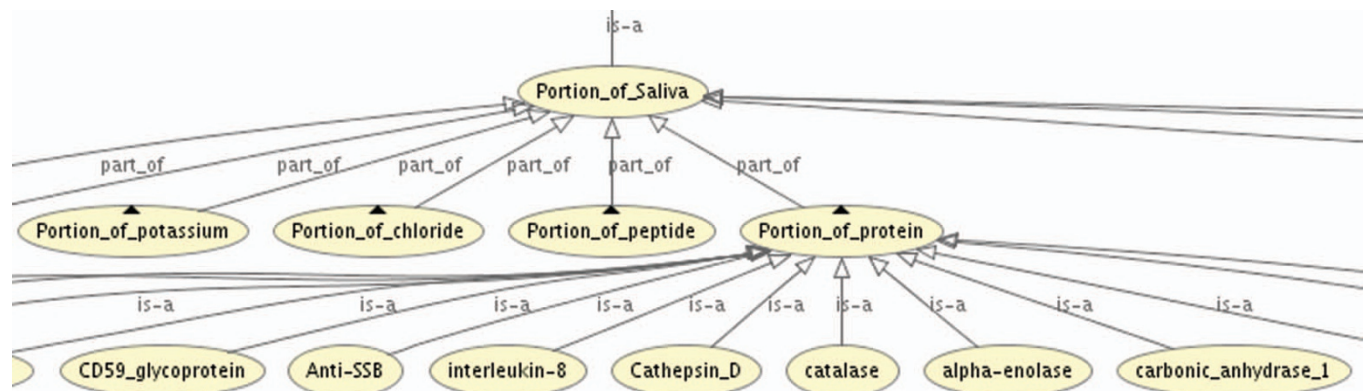


Figure 1 A fragment of the basic Saliva Ontology in its current form.

SDxMart is designed to provide a variety of queries to facilitate saliva biomarker discovery including complex queries that integrate genomic, clinical and functional information.

The SDxMart holds data from projects of oral diseases and systemic diseases including oral cancer, Sjögren's syndrome, pancreatic cancer and breast cancer. The types of datasets are: (i) proteomics; (ii) transcriptomics; (iii) microRNA; and (iv) metabolomics. In addition, the SDxMart is imported with several public databases including Ensembl genome database (Ensembl release 37),<sup>20</sup> and the number of resources is continuously growing.

## SUMMARY

The SKB is being created to facilitate researchers using salivary data from multiple perspectives. It is being built in tandem with the SALO and SDxMart which will allow the SKB to interoperate with other omics databases as part of a general strategy to facilitate integration of heterogeneous and disparate data sources that enable system biology approaches. Either SALO or SDxMart is a first and only resource of its kind in the field of dentistry.

## ACKNOWLEDGEMENTS

This work was supported by the US National Institutes of Health (grant number U01DE017790), the Felix & Mildred Yip Endowed Professorship and the Barnes Family Trust Fund.

---

- 1 Garcia I. NIDCR's 2009–2013 strategic plan. *J Dent Hyg* 2009; **83**(4): 153–154.
- 2 Phillips C. Rinse and spit: saliva as a cancer biomarker source. *NCI Cancer Bull* 2006; **2**: 39.
- 3 NIH. *FY 2006 NIH Annual Performance Report/FY 2008 NIH Performance Plan*. Bethesda: NIH, 2008: 82.
- 4 Hu S, Li Y, Wang J *et al*. Human saliva proteome and transcriptome. *J Dent Res* 2006; **85**(12): 1129–1133.
- 5 Huang CM, Zhu W. Profiling human saliva endogenous peptidome via a high throughput MALDI–TOF–TOF mass spectrometry. *Comb Chem High Throughput Screen* 2009; **12**(5): 521–531.

- 6 Takeda I, Stretch C, Barnaby P *et al*. Understanding the human salivary metabolome. *NMR Biomed* 2009; **22**(6): 577–584.
- 7 Ng DP, Koh D, Choo S *et al*. Saliva as a viable alternative source of human genomic DNA in genetic epidemiology. *Clin Chim Acta* 2006; **367**(1/2): 81–85.
- 8 Ashburner M, Ball CA, Blake JA *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**(1): 25–29.
- 9 Grenon P, Smith B, Goldberg L. Biodynamic ontology: applying BFO in the biomedical domain. *Stud Health Technol Inform* 2004; **102**: 20–38.
- 10 OBI Consortium. Available at: <http://obi-ontology.org/page/Consortium> (accessed 14 October 2010)
- 11 Scheuermann R, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. In: Proceedings of the 2009 AMIA Summit on Translational Bioinformatics; 15–17 March 2009; San Francisco, CA, USA. BioMed Central: London, UK, 2009, pp116–120.
- 12 Smith B, Goldberg LJ, Ruttenberg A *et al*. Ontology and the future of dental research informatics. *J Am Dent Assoc* 2010; **141**(10): 1173–1175.
- 13 Ai J, Smith B, David WT. Saliva Ontology: an ontology-based framework for a Salivaomics Knowledge Base. *BMC Bioinformatics* 2010; **11**: 302.
- 14 Brinkman RR, Courtot M, Derom D *et al*. Modeling biomedical experimental processes with OBI. *J Biomed Semantics* 2010; **1**(Suppl 1): S7.
- 15 Natale DA, Arighi CN, Barker WC *et al*. Framework for a protein ontology. *BMC Bioinformatics* 2007; **8**(Suppl 9): S1.
- 16 Degtyarenko K, de Matos P, Ennis M *et al*. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 2008; **36**(Database issue): D344–D350.
- 17 Almeida MB, Proetti AB, Ai J *et al*. The Blood Ontology: an ontology in the domain of hematology. In: Proceedings of the International Conference of Biomedical Ontologies; 28–30 July 2011; Buffalo, NY, USA. University at Buffalo: Buffalo, NY, USA, 2011, pp227–229.
- 18 Ai J, Almeida M, Andrade A *et al*. Towards body fluids ontology: a unified application ontology for basic and translational science. In: Proceedings of the International Conference on Biomedical Ontology; 28–30 July 2011; Buffalo, NY, USA. University at Buffalo: Buffalo, NY, USA, 2011, pp381–386.
- 19 Guberman JM, Ai J, Arnaiz O *et al*. BioMart Central Portal: an open database network for the biological community. *Database (Oxford)* 2011; **2011**: bar041.
- 20 Birney E, Andrews TD, Bevan P *et al*. An overview of Ensembl. *Genome Res* 2004; **14**(5): 925–928.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>