# WEB INFORMATION SYSTEM
# FOR THE STUDY OF ORAL HEALTH

José Melo[1], Joel P. Arrais[1], Pedro Lopes[1], Nuno Rosa[2], Maria José Correia[2],
Marlene Barros[2] and José Luís Oliveira[1]

[1]*Department of Electronics, Telecommunications and Informatics (DETI), Institute of Electronics and Telematics
Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal*
[2]*Health Sciences Department, Portuguese Catholic University, 3504-505 Viseu, Portugal*

Keywords:     Oral health, Data integration, Proteins, Diseases, Web services.

Abstract:     The human oral cavity is a complex ecosystem where multiple interactions occur and whose comprehension is critical to understand several disease mechanisms. In order to comprehend the composition of the oral cavity at a molecular level, it is needed to compile and integrate the biological information resulting from specific techniques, especially from proteomic studies of saliva. The objective for this work was to compile and curate a specific group of proteins related to the oral cavity, providing a tool to conduct further studies over the salivary proteome. Despite previous efforts to identify the protein components of saliva in healthy individuals and in several oral and systemic disorders, a resource compiling and reviewing all of these proteins is still lacking. In this paper we present a platform that integrates in a single endpoint all available information for proteins associated with the oral cavity. The proposed tool allows researchers in the biomedical sciences to explore organisms, proteins and diseases, constituting a unique tool to analyse meaningful interactions for oral health.

## 1  INTRODUCTION

Information available online is increasing in an accelerated way. In order to be processed, this increased amount of data requires the constant development of computer applications that must adapt to increasingly complex requirements, particularly those related to the integration of heterogeneous data and composition of distributed services.

Oral health is an area of research where these problems are particularly relevant. Being a very specific area of study, researchers are faced with many problems in obtaining clinically relevant information concerning the oral cavity using an easy and transparent way. This information must be stored and managed using tools that should provide the user with functionalities to retrieve, store, and search this data.

Usually, databases for molecular biology are centred either on a specific organism, such as *SGD for Saccharomyce*s (Cherry et al., 1998), or on specific research topic, such as STRING for protein-

protein interaction (Mering et al., 2003). In addition, databases like Entrez (Maglott et al., 2005) or the Universal Protein Resource (UniProt) (Bairoch et al., 2005) play a major role as hubs of biomolecular information, storing data from multiple topics and several organisms.

Despite this effort to create long lasting hubs of biomedical data has been very successful, one should not ignore the major contribute that many, more specific, databases provide to the actual state of science. They are of special interest for small communities that share common research interests. The UMD-DMD database (Humbertclaude et al., 2007), specialized on Duchenne Muscular Dystrophy, is one example.

Aware of the redundancy of features shared by many of those databases, and of the lack of technical expertise from the curators, several frameworks have been proposed to ease the task of deploying new databases. Examples include LOVD (Fokkema et al., 2005), specialized on annotating locus specific databases, GMODWeb (D O'Connor et al., 2008) for organism specific databases, or Molgenis (Swertz et al., 2010) that allows deploying more generic

biomedical databases. Despite the validity of those frameworks, there is none focused on simulating the behaviour of a single human organ or set of adjacent organs. This need to partition the data of the "whole" human system is relevant because, on the one hand, it reduces the time and resources involved in searching, processing and curating information, and on the other, it facilitates the use of algorithms to retrieve biologically meaningful results.

The main objective of this project was the development of a web information system that can collect genotypic information about oral health and that can be useful both for researchers and dentists. A comprehensive integrated resource of the saliva proteins, currently missing in the field of oral biology, would enable researchers to understand the basic constituents, diversity, and variability of the salivary proteome, allowing the definition and characterization of the human oral physiome. This goal was achieved at two levels: (1) *Oralome* for the application developer, which consists in a proprietary database, and a set of tools to retrieve biomolecular information from the major platforms like NCBI (National Center for Biotechnology Information) and UniProt; (2) for the end-user, a web portal (*OralCard*) directed to researchers and dentists, with a set of tools for searching and filtering data from the database, and the possibility to add new information.

Through OralCard web portal, users are able to perform their queries and search among a list of provided results. For each entity, users will be able to consult and analyze a list of dependencies and information retrieved from other major databases. To demonstrate the usefulness of this project we also present its application in the oral cavity research domain. A platform designed to integrate protein data related to this field will be implemented. This will include salivary proteins obtained in proteomic studies by different research groups, as well as proteins potentially produced and excreted by microorganisms assigned to the oral cavity. The ultimate goal is to present a tool for the community that contains accurate, manually curated and updated data regarding the oral cavity, to enable interactions studies, categorization and exploration.

We expect this work to be a valuable resource for investigators aiming to clarify the oral biology, identify molecular disease markers, develop diagnostic tests and improve prognostic, as well as providing information for the design of biological pathways setting the ground for the discovery of new therapeutic agents.

## 2 MOTIVATION

The oral cavity consists of a complex ecosystem where a variety of proteins from numerous origins are present. Being able to estimate the impact of the interactions among those proteins is crucial to understand the underlying disease mechanisms and hopefully to develop new treatment methods.

Saliva is the watery and usually frothy substance produced in the oral cavity of humans and most other animals. It is a unique clear fluid, composed of a complex mixture of electrolytes, proteins, and represented by enzymes, immunoglobulins and other antimicrobial factors, such as mucosal glycoproteins, traces of albumin and some polypeptides and oligopeptides, of importance for oral health (de Almeida et al., 2008).

Whole saliva is secreted mainly from three pairs of major salivary glands: the parotid, the submandibular and the sublingual glands. Approximately 90% of total salivary volume results from the activity of these three pairs of glands, with the bulk of the remainder from minor salivary glands located at various oral mucosal sites (Greabu et al., 2009). Whole saliva also contains proteins from gingival crevicular fluid, oral mucosa and oral microbiota. The various components of saliva from these sources, together with the plasma proteins that appear in saliva, define the physiological behaviour of the oral cavity, the oral physiome (Oralome).

Saliva is an ideal translational research tool and diagnostic medium and is being used in novel ways to provide molecular biomarkers for a variety of oral conditions, such as oral cancer (Nagler, 2009, Shpitzer et al., 2009), dental caries (Rudney et al., 2009) and periodontitis (Rudney et al., 2009, Gonçalves et al., 2010) , as well as systemic disorders such as breast cancer (Streckfus et al., 2008), Sjögren's syndrome (Hu et al., 2007), diabetes mellitus (Rao et al., 2009), cystic fibrosis (Livnat et al., 2010) and diffuse systemic sclerosis (Giusti et al., 2007). The ability to analyse saliva to monitor health and disease is a highly desirable goal for oral health promotion and research (Seymour et al., 2010, Wong, 2006). The most important advantage in collecting saliva is that it is obtained in a non-invasive way and is of easy access.

Over the past thirty years, there have been many efforts to determine and identify the main salivary proteins and peptides. Nevertheless, the fluctuating nature of saliva from different individuals, huge dynamic protein concentration ranges and the protein detection limits of most proteomic techniques have made the saliva proteome elusive to
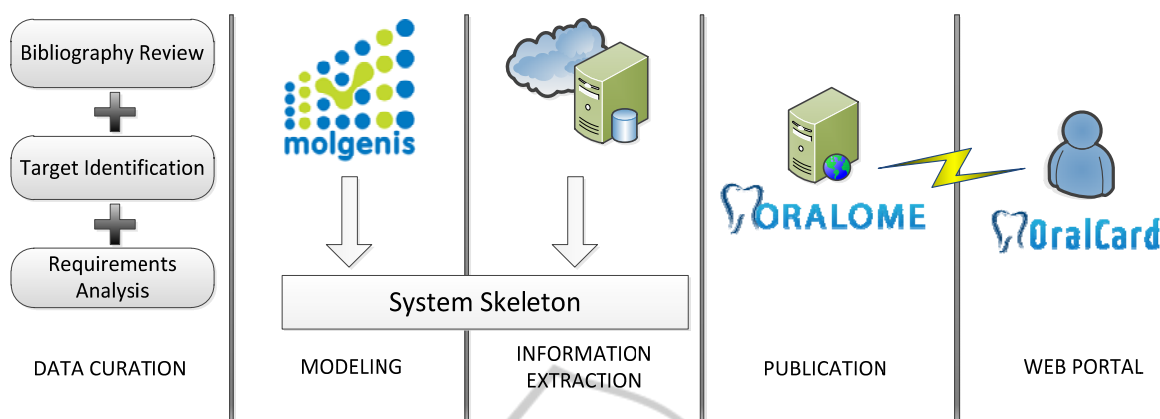
Figure 1: Pipeline for biomedical data integration.

define (Helmerhorst and Oppenheim, 2007). Even when healthy individual's saliva is considered, with multi-dimensional separations and advanced bioinformatics search software tools, proteins identified in different saliva proteomics experiments are often inconsistent with each other except for the most abundant proteins. To overcome the poor coverage, potential bias, and complementary nature of each experimental measurement of the human saliva proteome, it is necessary for biomedical researchers to collect and evaluate all reliable publicly-available saliva protein data sets generated from different analytical and computational platforms for healthy individuals as well as in disease conditions.

A comprehensive integrated resource of the saliva proteins would provide a high amount of comparative power for interpreting proteomics profile changes in patient's saliva, and may supplement or compensate for limitations and biases associated with the set of controls for a given study. It would also improve the ability for finding protein biomarkers that are known to occur in healthy human saliva, for instance where a protein is differentially expressed in a patient sample related to the quantities observed in the study control.

Oralome will have as a vital component an integrated database, by compiling and manually reviewing all of the existing experimental data performed on healthy individual samples as well as in several oral and systemic diseases. It will include a collection of microbial proteins expected to be present in saliva due to their presence in the genomes of the oral microbiota (Chen et al., 2010, Nelson et al., 2010) and a subset of microbial proteins determined experimentally (Xie et al., 2008).

We think OralCard will be a fundamental resource to clarify human oral biology and to establish protein biomarkers for salivary diagnostic processes based on the analysis of saliva samples both in health and in disease, exploring the Oralome functionalities. With OralCard, clinical samples from patient's saliva may be better analysed contributing to improved diagnostic methods and to the development of more effective therapies.

## 3 METHODS

Our solution is based upon two major components: (1) the Oralome for information picking, gathering and storage; and (2) the OralCard, a portal with a rich set of services for the end-user (Figure 1).

Oralome, our backend pipeline, comprises four functional phases: data curation, modelling, information extraction and publication.

For the OralCard portal frontend, we have selected two iterations: establishment of a relationship between the previously created database and the enterprise tier, and the design of views for the entities using web application frameworks.

### 3.1 Compilation and Curation of the Saliva Proteome

Regarding specific domains, such as the oral cavity, manual data curation is the key iteration where systems restrictions are imposed. These are essential to correctly focus the system and to establish a common platform for further iterations. In here three main activities are performed: bibliography review, target identification and requirements analysis. Reviewing bibliography consists on analysing state-of-the-art work and deciding where the final system uniqueness will be. Next, the target identification process involves exploring both relevant

bibliography and databases, and filtering which data and/or features should be available in the final system. Once data and features are selected, a careful requirements analysis process must be conducted. This envisages getting a first sense on the technological requirements related to the desired features and integrated data.

By the time this work was done, to our knowledge, there was no database to join the proteomes of major and minor salivary glands. For this reason, the first step was to compile this information from different sources. The proteome data of major salivary glands (parotid, submandibular/sublingual) were obtained from the Salivary Proteome Knowledge Base and from Yates Lab, The Scripps Research Institute. The proteome of human minor salivary gland secretion was obtained from Oppenheim Laboratory, Henry M. Goldman School of Dental Medicine, Boston University. The proteins identified in different works were compared and repeated entries eliminated.

Biological information is constantly being updated. Since first publication of saliva proteomes, many of the original identified proteins, catalogued as different entries in biological databases, have been merged with other ones and some were deleted due to misidentification. Therefore, all information concerning the identified proteins was manually curated and updated. The update of the IPI (International Protein Index) entries was carried out with the "IPI History Search" (*www.ebi.ac.uk/IPI*) tool. All other updates have been made using the UniProt database.

## 3.2 Oral Cavity Data Integration

The orthogonal nature and innate heterogeneity associated with life sciences resources have always hampered easier developments regarding the integration of distributed data. Furthermore, research in this field has entered a cycle where computational solutions lag one step behind technological requirements in biology. This brought about a growing disparity regarding bioinformatics software, where few well-known and widely used resources, such as UniProt or NCBI, co-exist with hundreds of smaller independent tools.

Although the oral cavity presents a stricter scope, it involves assorted life sciences fields, from microorganisms to proteins or diseases. Establishing new connections amongst these diverse entities creates a high degree of complexity, thus requiring the development of new ad-hoc data integration software solutions. On the one hand, large

warehouses that might contain this domain-specific information also contain plenty other resources. Consequently, researchers are overwhelmed by huge datasets, making their data of interest impossible to find. For instance, discovering oral cavity information amongst UniProt is a nightmarish task.

From a technological perspective, there are miscellaneous strategies for solving data integration problems. Though, they all rely on three elementary concepts: warehousing, middleware and link integration. Warehouse approaches intend to support an efficient decision making process, requiring the aggregation of all desired data in a huge central dataset (Santos and Bernardino, 2008). Middleware-based solutions rely on the development of specific wrappers to mediate connections between users' requests and original data servers (Barbosa et al., 2002). At last, link-based integration attempts to connect heterogeneous data types by creating graphs or networks based on pointers between distinct data units (Lopes et al., 2011). These approaches can be distinguished by the way they treat aggregated data. Warehouses replicate entire resources, creating a truly integrated environment whereas middleware or link-based solutions only provide a streamlined access to data, resulting in virtual integration.

Although many examples can be found for each integration strategy, the most common solutions involve developing a hybrid architecture, where some data is replicated whilst other are simply connected through links and identifiers. This approach was integrated into the Oralome project.

Once target resources and data were identified, the modelling iteration started. This task consisted on the design of a common information model to support oral cavity data from distinct resources.

Before the actual data integration, a system skeleton needed to be deployed. As mentioned above in this article, there are several frameworks designed for rapid prototyping of data portals for life science projects, such as LOVD or GMODWeb. For this specific task, we have chosen the Molgenis framework for its agility on creating a database and application, complete with data exploration web workspace, REST and SOAP web services, and R interface out of the box. For the data integration process, Molgenis provides easy and direct data input, whether through the web interface, through any of the available services, or through a provided database API. Therefore, custom data wrappers, collecting data from miscellaneous resources, can be easily implemented. Oralome required the deployment of general-purpose wrappers, combining external data in the newly deployed Molgenis

instance. These wrappers allow for systematic information extraction from resources such as UniProt, NCBI or STRING, amongst others. These resources provide several ways for information retrieval, such as REST interfaces or APIs for Java development.

Executing this streamlined data integration workflow curated oral cavity data is collected and re-organized in a publicly available web framework.

## 3.3 Oralome Development

Oralome will consist in a set of tools and a database which will provide access to information related to several entities, such as microorganisms, proteins, diseases and pathways, integrating crucial data regarding the oral cavity.

The upper entity is a microorganism which has several proteins associated to it. A protein for itself has others ids linked to it, such as OMIM (Online Mendelian Inheritance in Man), KEGG (Kyoto Encyclopedia of Genes and Genomes), PDB (Protein Data Bank) and GO (Gene Ontology) terms. The main subject for this tool consists in two groups of proteins: (1) a subset of microbial proteins determined experimentally, and (2) microbial proteins expected to be present in saliva. Regarding the first group, beside the information retrieved from UniProt, Oralome will integrate information related to the environment in which a protein was identified (health or disease, regulation, age group, and the particular source where it resides, for instance, mucosa or tongue).

For the Oralome tool development we chose the Molgenis framework for generating all the necessary tools and features needed to start fulfilling our database and to rapidly view this data in an easy way.

Molgenis consists in a framework written in Java, which accepts two XML files as input: a database and a user interface descriptor files. Using the first file, users can specify how the database will be structured, its entities and relations; the second file specifies the layout for the web interface. Molgenis generates a Java model and a database API which are used to deploy the related SQL tables, web services and web interface into a web server (Figure 2).

In order to start using this framework we needed to have preinstalled a MySQL server to store its database, an Apache Tomcat web server for deploying web services and a simple user interface, and the Java Development Kit for generation of SQL and HTML code.

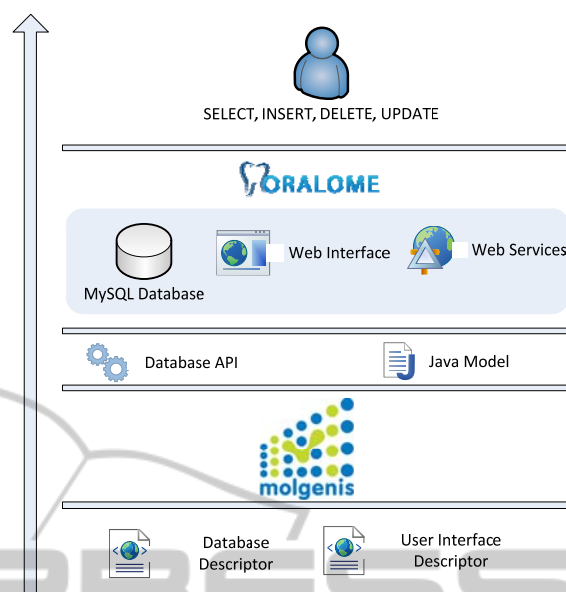Along with the Oralome application, we



Figure 2: Molgenis architecture.

developed tools and wrappers to obtain specific information on each of the elements that build up the system (proteins, diseases, pathways, and others). For this, we carried out a first survey of sources where this information would be available, and built a runnable script to update the Oralome database.

This data fetch is made easier using Molgenis. It bundles a Database API that has the advantage of hiding complex SQL commands.

To import and filter the information needed in our database, we used Java as the programming language because it is highly compatible with the most APIs provided by the major external services resources (UniProt, KEGG, and NCBI Entrez Utilities).

## 3.4 OralCard Development

In order to take advantage of the Oralome functionalities, we proposed a tool that would enable searching over the oral cavity database and showing different and customized views for each entity. This led to the OralCard web application, a fundamental resource for salivary diagnostic processes of protein biomarker studies of health and disease based on the analysis of saliva samples.

For information retrieval from Oralome, we decided to use Hibernate, an object-relational mapping tool for Java. This architecture is illustrated in the following diagram (Figure 3).
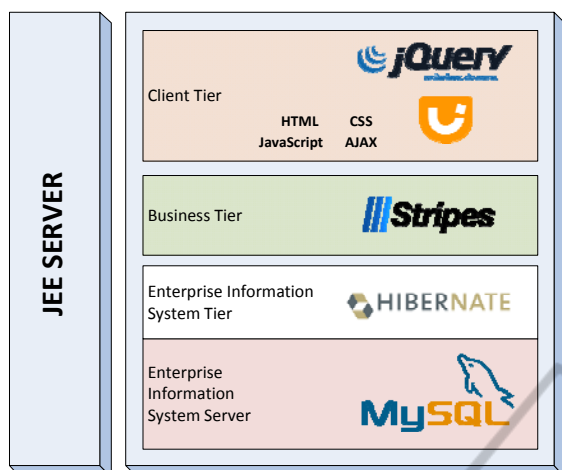
101

Figure 3: OralCard web application architecture.

OralCard frontend was developed using Stripes, a web framework that makes easier the development of Java web applications, by introducing some useful tools.

Stripes enabled us to take full control over URLs, making easy the task of accessing an entity by only knowing its id. For instance, researchers can have direct access to the protein *P22894* (*Neutrophil collagenase*), introducing the address *http://bioinformatics.ua.pt/oralcard/proteins/view/P22894*.

By using the Stripes framework, we were able to improve the user interface, introducing frameworks such as jQuery and jQueryUI. These tools contributed to present information in a more user friendly way, taking control over tables and AJAX interactions.

Finally, the OralCard web application takes advantage of the CSS benefits. It is designed to separate the document content written in JSP from the document presentation, including elements such as the page layout, used colours and fonts.

## 4 RESULTS

### 4.1 Oralome Functionalities

The web interface feature provided by Oralome reflects data contained in the database. It is available online at *http://bioinformatics.ua.pt/oralome*.

Both oral cavity researchers and developers can use this tool. Oralome provides two distinct entry points, each fitting a particular user type needs.

This workspace enables searching, filtering, browsing and viewing all collected data in a typical web interface. Moreover, researchers can combine data and download it for personal usage.

Along with the web interface for researchers, Oralome encompasses a set of web services for programmatic data access. Oralome's API is available for R, HTTP, REST and SOAP. These interfaces enable any developer to build custom solutions using any development framework, consuming Oralome's curated data related to the oral cavity.

In addition to the described oral cavity data exploration scenarios, Oralome allows the deployment of an unlimited number of customized applications. Using any the remote API interfaces, developers can use Oralome data to enrich already existing applications or to develop new ones.

This key Oralome feature envisages the usage of collected curated oral cavity information in education scenarios or by medical dentists. Oralome web framework's openness will enable the creation of an entire application ecosystem built around expertly curated information, aiming the delivery of improved dental health care.

### 4.2 OralCard Interface

OralCard web application is available online at *http://bioinformatics.ua.pt/oralcard*.

The researcher is first presented with a home page where he can insert search items. The protein customized view offers several tools. The *Data* tab presents the PANTHER classification system web application; the *Structure* tab provides a list of PDB structures related to the protein (Figure 4), where the user can download and visualize the related models; the *Interactions* tab features a STRING tool to visualize specific network diagrams containing related genes; the *References* tab contains all the publication related to the protein, and links to PubMed; it also features views for the related diseases, gene ontologies and for the sources of the oral cavity in which the protein was studied.

## 5 EVALUATION

As an application example, we used the Oralome tool to evaluate how saliva proteins contribute to or reflect impaired healing of oral cavity tissues in diabetic patients (Lamster et al., 2008). Then we identified all the salivary proteins whose amount is changed in patients with diabetes mellitus. Subsequently, in order to understand which salivary proteins molecular functions were altered in diabetes, we found that binding and catalytic activity functions have changed more evidently. That led to

| Id ▲ | Title | Date ⇕ | Image ⇕ | File ⇕ | Widget ⇕ |
|---|---|---|---|---|---|
| 1A85 | MMP8 WITH MALONIC AND ASPARAGINE BASED INHIBITOR | 1998-04-03 | View | Download | |
| 1A86 | MMP8 WITH MALONIC AND ASPARTIC ACID BASED INHIBITOR | 1998-04-03 | View | Download | |
| 1BZS | CRYSTAL STRUCTURE OF MMP8 COMPLEXED WITH HMR2909 | 1998-11-04 | View | Download | |
| 1I73 | COMPLEX OF PRO-LEU-L-TRP PHOSPHONATE WITH THE CATALITIC DOMAIN OF MATRIX METALLO PROTEINASE-8 (MET80 FORM) | 2001-03-07 | View | Download | |
| 1I76 | COMPLEX OF 2-(BIPHENYL-4-SULFONYL)-1,2,3,4-TETRAHYDRO-ISOQUINOLINE-3-CARBOXYLIC ACID (D-TIC DERIVATIVE) WITH T CATALITIC DOMAIN OF MATRIX PROTEINASE-8 (MET80 FORM) | 2001-03-08 | View | Download | |

Figure 4: Structure view for the Neutrophil collagenase protein.

the search for pathways in which the binding molecular function could be acting, in other words, which signalling pathways could be involved. Our results indicate that, the blood coagulation pathway (directly related to healing) was one in which the proteins altered in *diabetes mellitus* patient's saliva were involved. Then traced molecular networks reflecting the interactions between proteins involved in this pathway. We found that two key molecules in this network are two blood coagulation cascade inhibitors.

This methodology has allowed us to identify molecular reasons for the impaired healing of oral tissues in diabetic patients, as well as some key molecules in this process, which may be good molecular markers useful for diagnostic and even good targets for therapeutic agents.

## 6 CONCLUSIONS

In this paper we described a pipeline for creating web-based databases specialized in a set of adjacent human organs. We believe that the proposed work is of major importance to research projects that need an easy and agile solution to share the result of studies.

The information of the oral cavity is dispersed through different databases focused on more general systems. In addition to being dispersed, data is not always standardized, which makes their integration

and comprehensive study a colossal task. This work resulted in the development of an integrated database which comprises a comprehensive catalogue and characterization of the human oral proteome. It aims at becoming a fundamental resource to clarify human oral biology and in establishment of protein biomarker for salivary diagnostic processes.

We presented Oralome as a set of tools combining a database, web services and user interface, useful for joining specific results from several major databases, such as UniProt or NCBI. This framework generated a database, web services, and a web application where users can access the downloaded data.

Finally we proposed OralCard as an example of a web application that takes advantages of the Oralome functionalities. It works as a search engine where the researcher can input any looked-for search item, concerning to the field of oral biology. It uses the Oralome database and some of the new web standards for presenting specific information regarding a protein (JavaScript, AJAX and CSS).

Some of the advantages of the Oralome approach are the time and resources reduction involved in searching, processing and curating information, as facilitating the use of algorithms to retrieve biologically meaningful results. As a particular example of Oralome use, OralCard offer the advantage in gathering several major external tools in one single web application, becoming easier and more confortable to researchers accessing important

information regarding one specific protein. Using the oral cavity as a particular case study we have shown how it can be used to obtain a fully functional tool that enables both interaction's categorization and exploration.

We believe this project will be a valuable resource for investigators to clarify the oral cavity biology, identify molecular disease markers, to develop diagnostics tests and improve prognostic, as well as providing invaluable help in discovering new therapeutic agents.

# REFERENCES

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R. & Magrane, M. 2005. The universal protein resource (UniProt). *Nucleic acids research*, 33, D154.

Barbosa, A. C. P., Porto, F. a. M. & Melo, R. N. 2002. Configurable data integration middleware system. *Journal of the Brazilian Computer Society*, 8, 12-19.

Chen, T., Yu, W. H., Izard, J., Baranova, O. V., Lakshmanan, A. & Dewhirst, F. E. 2010. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database: the journal of biological databases and curation*, 2010.

Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A., Dwight, S. S., Hester, E. T., Jia, Y., Juvik, G., Roe, T. Y. & Schroeder, M. 1998. SGD: Saccharomyces genome database. *Nucleic acids research*, 26, 73.

D O'connor, B., Day, A., Cain, S., Arnaiz, O., Sperling, L. & Stein, L. D. 2008. GMODWeb: a web framework for the Generic Model Organism Database. *Genome Biology*, 9, R102.

De Almeida, P. V., Gregio, A., Machado, M., De Lima, A. & Azevedo, L. R. 2008. Saliva composition and functions: a comprehensive review. *J Contemp Dent Pract*, 9, 72-80.

Fokkema, I. F. a. C., Den Dunnen, J. T. & Taschner, P. E. M. 2005. LOVD: Easy creation of a locus specific sequence variation database using an "LSDB in a box" approach. *Human mutation*, 26, 63-68.

Giusti, L., Bazzichi, L., Baldini, C., Ciregia, F., Mascia, G., Giannaccini, G., Del Rosso, M., Bombardieri, S. & Lucacchini, A. 2007. Specific proteins identified in whole saliva from patients with diffuse systemic sclerosis. *The Journal of Rheumatology*, 34, 2063.

Gonçalves, L. D. R., Soares, M. R., Nogueira, F., Garcia, C., Camisasca, D. R., Domont, G., Feitosa, A. C. R., Pereira, D. A., Zingali, R. B. & Alves, G. 2010. Comparative proteomic analysis of whole saliva from chronic periodontitis patients. *Journal of proteomics*, 73, 1334-1341.

Greabu, M., Battino, M., Mohora, M., Totan, A., Didilescu, A., Spinu, T., Totan, C., Miricescu, D. & Radulescu, R. 2009. Saliva—a diagnostic window to the body, both in health and in disease. *J Med Life*, 2, 124-132.

Helmerhorst, E. & Oppenheim, F. 2007. Saliva: a dynamic proteome. *Journal of dental research*, 86, 680.

Hu, S., Wang, J., Meijer, J., Ieong, S., Xie, Y., Yu, T., Zhou, H., Henry, S., Vissink, A. & Pijpe, J. 2007. Salivary proteomic and genomic biomarkers for primary Sjögren's syndrome. *Arthritis & Rheumatism*, 56, 3588-3600.

Humbertclaude, V., Tuffery-Giraud, S., Ben Yaou, R., Hamroun, D., Khau Van Kien, P., Leturcq, F., Chelly, J., Claustres, M. & Béroud, C. 2007. G.P.9.10 Clinical development of the French UMD-DMD database. *Neuromuscular Disorders*, 17, 817-818.

Lamster, I. B., Lalla, E., Borgnakke, W. S. & Taylor, G. W. 2008. The relationship between oral health and diabetes mellitus. *The Journal of the American Dental Association*, 139, 19S.

Livnat, G., Bentur, L., Kuzmisnsky, E. & Nagler, R. 2010. Salivary profile and oxidative stress in children and adolescents with cystic fibrosis. *Journal of Oral Pathology & Medicine*, 39, 16-21.

Lopes, P., Dalgleish, R. & Oliveira, J. L. 2011. WAVe: web analysis of the variome. *Human mutation*.

Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. 2005. Entrez Gene: gene-centered information at NCBI. *Nucleic acids research*, 33, D54.

Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. & Snel, B. 2003. STRING: a database of predicted functional associations between proteins. *Nucleic acids research*, 31, 258.

Nagler, R. M. 2009. Saliva as a tool for oral cancer diagnosis and prognosis. *Oral oncology*, 45, 1006-1010.

Nelson, K. E., Weinstock, G. M., Highlander, S. K., Worley, K. C., Creasy, H. H. H., Wortman, J. R. R., Rusch, D. B., Mitreva, M., Sodergren, E. & Chinwalla, A. T. 2010. A catalog of reference genomes from the human microbiome. *Science* (New York, NY), 328, 994-999.

Rao, P. V., Reddy, A. P., Lu, X., Dasari, S., Krishnaprasad, A., Biggs, E., Roberts Jr, C. T. & Nagalla, S. R. 2009. Proteomic identification of salivary biomarkers of type-2 diabetes. *Journal of Proteome Research*, 8, 239-245.

Rudney, J., Staikov, R. & Johnson, J. 2009. Potential biomarkers of human salivary function: a modified proteomic approach. *Archives of oral biology*, 54, 91-100.

Santos, R. J. & Bernardino, J. 2008. Real-time data warehouse loading methodology. *Proceedings of the 2008 international symposium on Database engineering \&\#38; applications*. Coimbra, Portugal: ACM.

Seymour, G. J., Cullinan, M. P. & Heng, N. C. 2010. Oral Biology: Molecular Techniques and Applications (*Methods in Molecular Biology*), Humana Press.

Shpitzer, T., Hamzany, Y., Bahar, G., Feinmesser, R., Savulescu, D., Borovoi, I., Gavish, M. & Nagler, R.

2009. Salivary analysis of oral cancer biomarkers. *British journal of cancer*, 101, 1194-1198.

Streckfus, C. F., Mayorga-Wark, O., Arreola, D., Edwards, C., Bigler, L. & Dubinsky, W. P. 2008. Breast cancer related proteins are present in saliva and are modulated secondary to ductal carcinoma in situ of the breast. *Cancer investigation*, 26, 159-167.

Swertz, M., Dijkstra, M., Adamusiak, T., Van Der Velde, J., Kanterakis, A., Roos, E., Lops, J., Thorisson, G., Arends, D. & Byelas, G. 2010. The MOLGENIS toolkit: rapid prototyping of biosoftware at the push of a button. *BMC Bioinformatics*, 11, S12.

Wong, D. T. 2006. Salivary diagnostics powered by nanotechnologies, proteomics and genomics. The *Journal of the American Dental Association*, 137, 313.

Xie, H., Onsongo, G., Popko, J., De Jong, E. P., Cao, J., Carlis, J. V., Griffin, R. J., Rhodus, N. L. & Griffin, T. J. 2008. Proteomics analysis of cells in whole saliva from oral cancer patients via value-added three-dimensional peptide fractionation and tandem mass spectrometry. *Molecular & Cellular Proteomics*, 7, 486.