

Database

Open Access

## Integrated olfactory receptor and microarray gene expression databases

Nian Liu<sup>\*1,2</sup>, Chiquito J Crasto<sup>1,3,5</sup> and Minghong Ma<sup>4</sup>

Address: <sup>1</sup>Center for Medical Informatics, Yale University School of Medicine, New Haven, CT 06520, USA, <sup>2</sup>Department of Anesthesiology, Yale University School of Medicine, New Haven, CT 06520, USA, <sup>3</sup>Department of Neurobiology, Yale University School of Medicine, New Haven, CT 06520, USA, <sup>4</sup>Department of Neuroscience, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA and <sup>5</sup>Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Email: Nian Liu<sup>\*</sup> - [nian.liu@yale.edu](mailto:nian.liu@yale.edu); Chiquito J Crasto - [chiquito.crasto@yale.edu](mailto:chiquito.crasto@yale.edu); Minghong Ma - [minghong@mail.med.upenn.edu](mailto:minghong@mail.med.upenn.edu)

<sup>\*</sup> Corresponding author

Published: 30 June 2007

Received: 12 December 2006

BMC Bioinformatics 2007, 8:231 doi:10.1186/1471-2105-8-231

Accepted: 30 June 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/231>

© 2007 Liu et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Gene expression patterns of olfactory receptors (ORs) are an important component of the signal encoding mechanism in the olfactory system since they determine the interactions between odorant ligands and sensory neurons. We have developed the Olfactory Receptor Microarray Database (ORMD) to house OR gene expression data. ORMD is integrated with the Olfactory Receptor Database (ORDB), which is a key repository of OR gene information. Both databases aim to aid experimental research related to olfaction.

**Description:** ORMD is a Web-accessible database that provides a secure data repository for OR microarray experiments. It contains both publicly available and private data; accessing the latter requires authenticated login. The ORMD is designed to allow users to not only deposit gene expression data but also manage their projects/experiments. For example, contributors can choose whether to make their datasets public. For each experiment, users can download the raw data files and view and export the gene expression data. For each OR gene being probed in a microarray experiment, a hyperlink to that gene in ORDB provides access to genomic and proteomic information related to the corresponding olfactory receptor. Individual ORs archived in ORDB are also linked to ORMD, allowing users access to the related microarray gene expression data.

**Conclusion:** ORMD serves as a data repository and project management system. It facilitates the study of microarray experiments of gene expression in the olfactory system. In conjunction with ORDB, ORMD integrates gene expression data with the genomic and functional data of ORs, and is thus a useful resource for both olfactory researchers and the public.

### Background

The mammalian olfactory system, with its ability to detect numerous odor molecules in the environment, helps animals locate food, mates, and predators. The detection sensitivity and specificity largely depend on the olfactory receptors (ORs) expressed in the primary sensory neurons

within the neuroepithelium in the nose [1]. In rodents, there are approximately 10 million sensory neurons expressing > 1000 different types of OR genes in each nostril [2]. Each neuron expresses only one specific OR, which exhibits an optimal response to specific odorants [3]. Axons from the sensory neurons that express the same

OR converge onto one or a few glomeruli per olfactory bulb [4]. The expression patterns of OR genes and the unique projection of their sensory neurons provide a molecular basis for olfactory signal coding in the brain.

Diverse sets of data have resulted from olfactory research carried out at the cellular, functional and behavioral levels [5-9]. Since the first identification of rat OR genes in 1991, thousands of ORs have been found, some in tissues not associated with the olfactory system [10,11]. Functional imaging in the olfactory bulb has produced two-dimensional "odor maps" (unique activity patterns) in the glomerular layer [12-14]. Three key databases in the SenseLab [ORDB (OR database), OdorDB (odor database) and OdorMapDB (odor map database)], housed at the Yale University School of Medicine, provide an integrated, multidisciplinary model for the olfactory pathway [15-18]. The information contained in these databases illustrates the chain of events that starts from the exposure of the animals to odorous environment, to the binding of odorant ligands with specific ORs, and to the resulting spatial activity patterns in the olfactory bulb.

Recent microarray studies using Affymetrix (Santa Clara, CA) gene-chips have resulted in the accumulation of a large amount of OR gene expression data. These data show the differential expression of OR genes in the olfactory and non-olfactory tissues [11]. Building an efficient microarray database specific for the ORs and integrating it with the SenseLab system has been challenging. There are many large-scale and well-established databases, e.g., ArrayExpress at the European Bioinformatics Institute [19], the Stanford Microarray Database [20], and the Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI) [21], for public data repositories of the microarray experiments. However, they do not fit the needs specific to the chemosensory research community, i.e., to have a domain-specific database integrated with the SenseLab system.

This paper describes the Olfactory Receptor Microarray Database (ORMD), an informatics tool dedicated to disseminating information related to rodent OR microarray experiments. Currently archived in the ORMD are gene expression data of ORs in the olfactory epithelium as well as other tissues. ORMD is integrated with ORDB (which relies on an architecture that is common to all databases in SenseLab), through dynamic link in the webpage for each OR. Both databases are designed to facilitate the experimental research on elucidating the mechanisms underlying perception of smell.

## Construction and content

### Methods

The database schema of ORMD has been implemented in Oracle. It is primarily a traditional relational database. Some tables, however, implement the EAV (Entity-Attribute-Value) data model [22]. Since the number and names of the parameters defining the experimental conditions may vary and be unpredictable, the EAV model has been chosen to store the values of some parameters for the experiments. This model ensures flexibility of the system. Also, an EAV-like approach has been used to store gene expression datasets, with the experiment being the entity, the probe set name being the attribute, and the signal, detection call, and P-value being the values. The Web application has been developed in Java Server Pages (JSP) and run on Apache Tomcat Server.

### Olfactory Receptor Database – ORDB

ORDB is one of the major databases in the SenseLab system [23]. As previously reported [18], the archived OR properties are categorized as descriptive attributes and sequence data. The descriptive attributes include animal species name (i.e., organism), strain, source tissue, chromosomal location, data source, sequence laboratories [the principal investigator's (PI's) lab that cloned the OR gene or identified the gene from the genome], and references (including links to GenBank and Medline sources for the OR gene). The sequence data include the nucleotide residues of genomic DNA or cDNA (complementary DNA) and the amino acid residues of receptor proteins. Recently, functional data, including gene expression, molecular modelling and activity regulation, have been added as a new category of information about the receptors. In addition, ORDB is continuously being expanded to include recently identified receptors, as well as new species whose ORs have been cloned and identified [16]. Currently, ORDB contains receptors from 50 different species with complete genomic repertoires for several species, including mouse, rat and human. It represents the work of more than 100 laboratories around the world.

Figure 1 is an example ORDB webpage showing the information for an individual OR, ORL2135, a name based on the ORDB's nomenclature. The receptor also has a common name, e.g., MOR182-4, given by the laboratory that cloned or otherwise identified the gene. The gene for ORL2135 is located on mouse chromosome 16. Celera Genomics (Rockville, MD) is the source of the DNA sequence. The sequence type of receptors can be either genomic DNA or cDNA. In this figure, the sequence type of ORL2135 is genomic DNA. For receptors with the sequence type of cDNA, e.g., ORL466, the source tissues are usually provided. ORL466 has two putative odorant ligands: hexanol and heptanol. A URL link to ORMD (described below) is provided in the webpage. This link

allows users to access the gene expression data related to the OR.

#### Olfactory Receptor Microarray Database – ORMD

ORMD is a secure, Web-based repository for the OR gene expression data from the microarray experiments. The tools to access the data are the same for both regular and

logged-in users. The availability of individual microarray data, however, depends on the display attribute of the experiments set by the data owners, the ownership of the data, and the logged-in status of the user. While users without login can access the data that have been made "public" by owners, users with login may access both the public data and their own private data. Only logged-in

Chemosensory Receptors	
Name:	ORL2135
Description	
CR Type	. <a href="#">ORL (Olfactory Receptor Like)</a> <a href="#">Show other</a>
Organism	. <a href="#">Mus musculus (mouse)</a> <a href="#">Show other</a>
Source Tissue	
Chromosome	16
Genbank	<a href="#">AY073097</a>
Medline	<a href="#">21864068</a>
Common Name	MOR182-4/WE
Strain	129X1/SvJ, 129
Source	. <a href="#">CELERA Gen</a>
Sequence Lab	. <a href="#">FIRESTEIN :</a> . <a href="#">TRASK B</a>
Length	. <a href="#">Full-Length</a>
Ligand	
Nucleotides	ATGGAGAAGAAC ATCTATGTCACC TTAGCCTTTGCI TTTGCCCAATAJ CTGCTCTATCTA TTATTATTTAGA CTTAACACATTG CTGAACATGAAC ATGTACGTGAGC TACAGCTTAAGA
Amino Acids	MEKNSTLLTEFV LRGITDRSEI FAQYYFFCISAT TEIFLLVAMA LNTLVVFICASS IQISTSVTII YSLRNKEVLGAL KKFIFK
Type of Sequence	. <a href="#">Genomic DNA</a> <a href="#">Show other</a>
Accession Number	AY073097
Receptor 2D (GPCRDB)	<a href="#">Q8VGP9 MOUSE-rbdg.html</a>
Molecular Models	
Functional Analysis Laboratory	
Functional Analysis Reference	
Microarray Experiment	<a href="#">18060</a>

**Figure 1**

**SenseLab ORDB archiving the olfactory receptors.** This sample webpage lists the attributes (Left column) and the available values (Right column) for a receptor. The data include many descriptive attributes (such as the organism and the sequence lab), the nucleotide and amino acid sequences of the gene and its protein product, and the functional data (such as gene expression from microarray experiments). The hyperlinks lead to related pages in ORDB or other websites for detailed attribute information about the receptor. Note that the bottom part of the webpage has been cut and moved up showing the full list of the receptor attributes.

users may create projects and enter the details of data resulting from their microarray experiments. For each experiment, the gene expression datasets and the raw data files can be uploaded and stored in the database. Currently archived in the database are 31 microarray experiments from mice using the standard (Murine Genome Array U74Av2) and custom-designed Affymetrix gene-chips covering the mouse ORs. Fourteen experiments, with sample sources from the olfactory epithelium and varied body organs including brain, testes, heart, spleen, etc, have been made available to the public.

Figure 2 is an example webpage showing the details of an individual project. The project shown in this page investigates the OR expression patterns in the septal organ, a small island of olfactory neuroepithelium in the mouse nasal septum. The menu bar located on the left of the page allows users to use the database more effectively. The content of the page shows the project name, name of the PI, project description, and a list of associated experiments. The name of each experiment (the first column in the table) is hyperlinked to another webpage that contains details of the experiments. Each experiment is described by gene-chip (i.e., probe array type) name, sample type, protocol, and name of the operator (the technician carrying out the experiment). The owner of the data retains privileges which include modifying the content of project information and its experimental details.

Gene-chip information, including names of the probe sets on the chip and the related ORs, is stored in the database and used to annotate the expression data. Figure 3 is an example webpage showing gene expression levels in the experiment "OE1", an experiment in the project that investigates the OR expression profiles in different zones of the neuroepithelium. The table in Figure 3 shows names of the probe sets, signal intensities, and statistical calls of detections of the genes in the hybridization. It also shows the OR name used in ORDB for the probe set, with hyperlinks (implemented using receptor IDs generated by ORDB) that direct users to the webpage for that OR in ORDB. Hyperlinks showing publications related to each gene in PubMed are also provided. Users can also view the expression levels of a particular gene in all publicly accessible experiments in the database.

#### **Microarray experiment data upload/download**

Each microarray experiment is associated with three types of files: (1) An experiment-file that describes the details and conditions associated with hybridization; (2) A dataset file that describes the gene expression levels; and (3) Raw data files that include the hybridization image files. All these files may be uploaded into ORMD using a single Web interface (Figure 4A). The experiment and dataset files are text-based. The Web server parses these files and

stores the information in the database. The probe set names in the dataset file are used to link the expression levels with individual ORs. The raw data files are stored in the Oracle database as binary data type. When an experiment file is uploaded, a corresponding "experiment" (name) will be automatically generated in the database. When the user uploads the dataset or raw data files, the corresponding experiment name is selected. The Web interface for uploading data is easy to use and intuitive.

Gene expression datasets can also be exported from ORMD. Figure 4B shows the Web interface for data export. In order to provide comparable datasets from different experiments, the data from the same gene-chip type are exported. Once the chip type is selected, the list of related experiments is automatically refreshed. Users may choose one or more datasets and export and save them (into their local computers) as text-formatted or Microsoft Excel files. Users can then use third-party tools to analyse the exported data. Figure 4C shows part of a sample Excel file of gene expression data from the experiments named "Embryo," "Heart," and "Kidney." These files describe the OR expression profiles for these three organs. In addition, the system allows users to choose multiple experiments and carry out pair-wise scatterplots to visualize the gene expression patterns between samples. Raw data files are also available for download from the experiment detail page, allowing researchers to share the data with colleagues and others in the community.

#### **A project and experiment management system**

ORMD provides an informatics tool for researchers to manage their projects and experiments. It allows users to create/edit projects which may include multiple experiments. Individual experiments may be associated with one or more projects. This hierarchy helps researchers to systematically organize their data. It also allows efficient retrieval of the microarray datasets based on the projects.

ORMD primarily serves as a private data repository for those who will deposit data, but also provides a resource of OR gene expression information to the community. The decision as to which datasets should be made available to the public lies with the data owners (i.e., account users). While private experimental data can be made public, the publicly available data can be made private by the owners. This allows the account users the freedom to protect their unpublished data. In general, the criterion for making data public is the publication of the study in journals. The owners can also make available the unpublished data that they are willing to share with the community. The system provides the capacity to allow the public data to be adequately annotated, in particular for the MIAME-compliance of the published data [24].

Olfactory Receptor Microarray Database								Logged-in: Minghong Ma	
<div>Home</div> <div>Search</div> <div>Projects</div> <div>Experiments</div> <div>GeneChips</div> <div>DataExport</div> <div>DataAnalysis</div>	<div>Project Name:</div>	ORs in the septal organ							
	<div>PI:</div>	Minghong Ma							
	<div>Created Date:</div>	2005-02-03 15:14:58.0							
	<div>Start Date:</div>								
	<div>Description:</div>	Olfactory receptors expressed in the septal organ, as compared to the main olfactory epithelium and the respiratory epithelium							
	<div>Update</div>								
Experiment List									
<div>Downloads</div> <div>Library</div> <div>Dataset</div>	Experiment	Probe Array Type	Sample Type	Protocol	Operator	Date	Description	Comments	
	<a href="#">OE_NOV2702</a>	cvyMORF	TOTAL RNA, 3 ug	Midi_Euk2v3	Shrikant Mane	2002-11-27 12:27:00.0		CUSTOM CHIP	
	<a href="#">RE-Nov2702</a>	cvyMORF	TOTAL RNA, 3 ug	Midi_Euk2v3	Shrikant Mane	2002-11-27 12:27:00.0		CUSTOM CHIP	
	<a href="#">SO-Nov2702</a>	cvyMORF	TOTAL RNA, 3 ug	Midi_Euk2v3	Shrikant Mane	2002-11-27 12:54:00.0		CUSTOM CHIP	
	<a href="#">MOE-septum-June2402</a>	cvyMORF	REHYBRIDIZATION	Midi_Euk2v3	Sheila Westman	2002-06-24 13:29:00.0	ming hong ma, book 1, p.40-41, sample 3	small sample protocol, double amplification	
<div>Uploads</div> <div>Library</div> <div>Data</div>	<a href="#">RE-June2402</a>	cvyMORF	TOTAL RNA, 1 ug	Midi_Euk2v3	Sheila Westman	2002-06-20 11:41:00.0	MING HONG MA, BOOK 1, P 40-41, SAMPLE #	small sample protocol, double amplification	
<div>Admin</div> <div>Profile</div> <div>AddProject</div> <div>SetExpDisplay</div>									
<div>Logout</div>									

**Figure 2**

**ORMD archiving microarray experiments of the olfactory receptors.** This webpage shows a single project and the associated experiments. Listed in the left column of the page are menu items allowing users to search and browse the projects and experiments, to upload/download data files, and to export the gene expression data. The project is annotated with a PI-defined project name, PI name, archiving date, and description. The table lists all the associated experiments, with information about the array type, sample type, operator, etc. The experiment names are hyperlinks that lead to a page showing the experiment detail and gene expression data.

Olfactory Receptor Microarray Database						Logged-in: Minghong Ma	
Home Search	Experiment: OE1						
	Project: ORs in different zones						
Projects	Probe Set Name	Signal	Detection	P-Value	ORDB_Name	PubMed_id	
Experiments	<a href="#">136_MOR184-8CDS_at</a>	7.5	A	.962402	<a href="#">ORL2133</a> ●	<a href="#">21864068</a>	
GeneChips	<a href="#">136_MOR184-8UTRL_at</a>	1110.9	A	.067627	<a href="#">ORL2133</a> ●	<a href="#">21864068</a>	
DataExport	<a href="#">137_MOR182-3CDS_at</a>	457.2	P	.023926	<a href="#">ORL2134</a> ●	<a href="#">21864068</a>	
DataAnalysis	<a href="#">138_MOR182-4CDS_at</a>	424.2	P	.000732	<a href="#">ORL2135</a> ●	<a href="#">21864068</a>	
Downloads	<a href="#">139_MOR266-1CDS_at</a>	437.6	M	.056152	<a href="#">ORL2224</a> ●	<a href="#">21864068</a>	
	<a href="#">140_MOR127-1CDS_at</a>	3088.4	P	.001221	<a href="#">ORL2244</a> ●	<a href="#">21864068</a>	
	<a href="#">140_MOR127-1UTRL_at</a>	261.5	A	.171387	<a href="#">ORL2244</a> ●	<a href="#">21864068</a>	
Library	<a href="#">141_MOR261-1CDS_at</a>	329.9	A	.398926	<a href="#">ORL1439</a> ●	<a href="#">21864068</a>	
Dataset	<a href="#">AFFX-BioB-5_at</a>	1325.9	P	.004017	<a href="#">ORL1074</a> ●	<a href="#">99384009</a>	
Uploads	<a href="#">AFFX-BioB-M_at</a>	1268.2	P	.00039	<a href="#">ORL1076</a> ●	<a href="#">99384009</a>	
	<a href="#">AFFX-BioB-3_at</a>	681.2	P	.001248	<a href="#">ORL1073</a> ●	<a href="#">99384009</a>	
	<a href="#">AFFX-BioC-5_at</a>	2467.4	P	.00039	<a href="#">ORL1078</a> ●	<a href="#">20256252</a>	
	<a href="#">AFFX-BioC-3_at</a>	1530.7	P	.00007	<a href="#">ORL1077</a> ●	<a href="#">99384009</a>	
	<a href="#">AFFX-BioDn-5_at</a>	1849.2	P	.000044	<a href="#">ORL108</a> ●	<a href="#">96158919</a>	
Admin	<a href="#">AFFX-BioDn-3_at</a>	14000.8	P	.000127	<a href="#">ORL1079</a> ●	<a href="#">20183981</a>	
	<a href="#">AFFX-CreX-5_at</a>	18405.8	P	.000044	<a href="#">ORL1099</a> ●	<a href="#">12034333</a>	
	<a href="#">AFFX-CreX-3_at</a>	34862.2	P	.000044	<a href="#">ORL1080</a> ●	<a href="#">99238494</a>	
Profile	<a href="#">AFFX-DapX-5_at</a>	72.2	A	.327079	<a href="#">ORL110</a> ●	<a href="#">96149403</a>	
AddProject	<a href="#">AFFX-DapX-M_at</a>	175.9	A	.095667	<a href="#">ORL1100</a> ●	<a href="#">12034333</a>	
SetExpDisplay	<a href="#">AFFX-DapX-3_at</a>	13.3	A	.979978	<a href="#">ORL11</a> ●	<a href="#">91191556</a>	
Logout	<a href="#">AFFX-LysX-5_at</a>	35.4	A	.699409	<a href="#">ORL1104</a> ●	<a href="#">11802173</a>	
	<a href="#">AFFX-LysX-M_at</a>	73.2	A	.737173	<a href="#">ORL1105</a> ●	<a href="#">11802173</a>	

**Figure 3**

**A single experiment in ORMD archiving the gene expression data.** For each probe set, the signal intensity, detection call (A-absent; P-present), and statistical P-value are provided. For each probe set, the target gene name is provided in the column "ORDB\_name" (corresponding to the name used in SenseLab). The hyperlinks on the gene name lead to the detail page of chemosensory receptors in ORDB (see Figure 1). The PubMed\_IDs are hyperlinks directing users to original publications related to the genes.



### A

Upload Experiment File (with .EXP extension):

Browse...

Upload Experiment File

- OR -

Upload Dataset File:

Experiment: MOEEXP ▼

Chip Type: cvyMORF ▼

Data File:  Browse...

Upload Dataset File

- OR -

Upload Raw Data File:

Experiment: —Select an Experiment— ▼

Chip Type: —Select the Correct Chip Type— ▼

Data File:  Browse...

Upload Raw Data Files

### B

#### Export ORMD Expression Data

---

Step 1

Chip Type: cvyMORF ▼

Step 2

Choose Experiments: 

BRAIN  
 EMBRYO  
 HEART  
 Kidney  
 LUNG

Step 3

Choose Export Format: ☐ Text ☒ Excel

Export

### C

1	Chip Type: cvyMORF									
2		EMBRYO				HEART			Kidney	
3	Probe Set Name	Signal	Detection	P_Value		Signal	Detection	P_Value	Signal	Dete
4	AFFX-BioB-5_at	1332	A	0.095667		1801	P	0.039661	1665.8	P
5	AFFX-BioB-M_at	1814.2	P	0.002275		2671.4	P	0.006187	1605	P
6	AFFX-BioB-3_at	1302.3	P	0.026111		1371.8	P	0.042962	1144	P
7	AFFX-BioC-5_at	4100.2	P	0.000581		5400.7	P	0.001796	3210.3	P
8	AFFX-BioC-3_at	1787.4	P	0.002023		2814.6	P	0.003212	1698.3	P
9	AFFX-BioDn-5_at	2211.6	P	0.003595		2737.4	P	0.00034	2178.2	P
10	AFFX-BioDn-3_at	28258	P	0.000857		31483.5	P	0.000258	20267.8	P

**Figure 4**

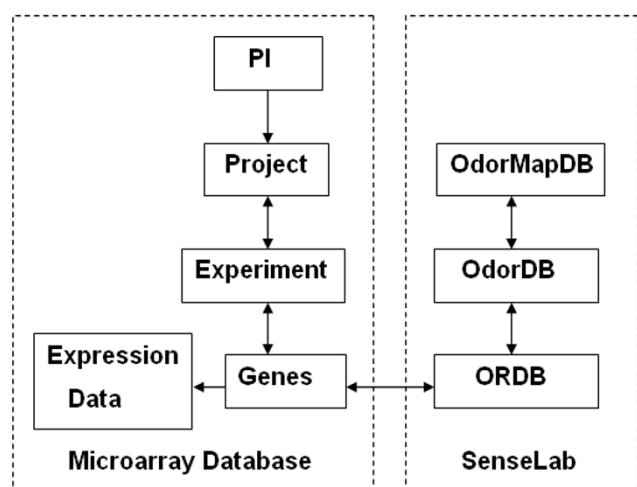
**The Web interface in ORMD for data upload and export.** A. Data upload. For each experiment, three separate uploads are carried out depending on the file type: 1) the description text file; 2) the dataset file showing the gene expressions; and 3) the raw data files. The description and dataset files are parsed by the system and the values are stored in the database. The raw data files are stored in the database as binary data type. B. Data export. The user first needs to choose a gene-chip type (Step 1), after which the dropdown list of the related experiments will be automatically refreshed. The user may select one or more experiments (Step 2) and choose an appropriate export file format, i.e., text or MS Excel (Step 3). C. A sample Excel file of gene expression data from three experiments: "Embryo," "Heart," and "Kidney".

## Discussion

This paper describes a database system for the storage and presentation of microarray gene expression data of olfactory receptors. The database is integrated with the olfactory databases in the SenseLab system. It is designed to facilitate experimental research in the olfactory field.

## Database integration

Investigating how odor information is transduced and processed by the olfactory system is essential to our understanding of the sense of smell. In essence, SenseLab uses the olfactory system as a model to develop informatics tools to facilitate experimental neuroscience research. The



**Figure 5**  
**Relationships among the major components within ORMD and among the SenseLab olfactory databases.**  
 In ORMD, the users (PI) may create projects. The projects and experiments have a many-to-many relationship, i.e., each project may have multiple experiments and each experiment may be associated with one or more projects. Microarray gene expression data are uploaded for individual experiments and annotated with the "genes" which are probed in the particular gene-chips. The olfactory receptor "genes" in ORMD have links with the corresponding olfactory receptors archived in ORDB of the SenseLab. The receptors in ORDB are also linked to the putative odorant ligands, if there are any, in OdorDB. The odor-induced activity patterns in the brain ("odor maps") are archived in OdorMapDB.

OdorDB, ORDB, and OdorMapDB in SenseLab archive the odor types, ORs, and odor map images, respectively. These olfactory databases are integrated, allowing a clear description of the chain of events from the odor stimuli to the unique activity pattern in the brain.

SenseLab has been developed to be flexible. Its common architecture allows adding new databases and integrating with other systems [22]. ORMD, the creation and development of which is described in this study, is a database associated with olfaction that archives the OR gene expression data from microarray experiments. ORMD is linked to the ORDB in SenseLab through webpages providing genomic, proteomic and associated information for individual receptor genes (Figure 5). Since they store different, yet complimentary, types of data, the links that integrate ORMD and ORDB are beneficial to both databases. Users of ORMD may have easy access to the SenseLab to request gene information (for which microarray experiments have been carried out) that includes the nucleotide sequence, protein sequence, and odorant lig-

ands known to interact (excite or inhibit) these ORs. On the other hand, users of ORDB may access microarray experiments and the related gene expression data for their OR gene of interest.

ORMD and ORDB differ not only in the type of data archived but also in the scope of user accessibility. ORDB is a knowledge-based database, with its content originating from published data. The data include the "normal" genetic information of the receptors in a given species. The volume of the ORDB grows as investigators identify new genes or extend their research into new species. On the other hand, ORMD archives gene expression data from diverse microarray experiments, dependent on tissue source and affected by biological or disease conditions. Whereas ORDB primarily provides a resource to the research community and general public (though facilities for private storage of cloned OR gene data prior to publication or deposition into GenBank exists), ORMD provides a data repository, management and sharing tool for researchers with user accounts and also allows public access to dedicated datasets.

#### **Informatics tools for experimental olfactory research**

The eventual utility of both ORDB and ORMD in olfactory research will be evident as the microarray approach is increasingly used to investigate the gene expression patterns of ORs in the olfactory system [11,25]. Receptor genes archived in ORDB are characterized by their sequences, species, chromosome locations, etc. The expression of the receptor genes, however, depends on species, developmental stage, and tissue source. Strong expression of some genes in certain regions may help researchers uncover the relationships between animal behaviors and the stimulating odor types. Since the OR gene family contains hundreds to thousands of members in each species, the gene-chip approach provides a high-throughput, combinatorial, and powerful tool to examine the expression of the identified genes in the species simultaneously. The integrated olfactory databases described in this paper help archive and present large amounts of gene expression data, thus facilitating experimental research in understanding the molecular mechanism underlying olfactory detection and discrimination.

ORs have also been found in many other organs, such as testes, liver, and spleen. For example, the olfactory receptor hOR17-4 is found in human spermatogenic cells and may play a role in chemical communication between sperm and egg [26]. The expression of olfactory receptors in non-olfactory tissues may be due to the fact that the ORs are members of a superfamily of membrane receptors known as G-protein-coupled receptors [27]. Although the functions of ORs in other body organs remain elusive, a comprehensive investigation of ORs using microarray



techniques will enhance our understanding of signal transduction in biological systems beyond olfaction.

### A generic data management system

Although ORMD is currently a data depository and management system for Affymetrix gene-chip data, it can serve as an open-source database easily adapted to house other types of microarray data. Many journals require that published microarray data conform to the MIAME Consortium. ORMD primarily serves as a private data repository, not as a portal for publication of experimental results. Although the database does not enforce the requirements set by the Consortium for private data, it allows owners to annotate the data made public according to the MIAME checklist. In general, storing good quality data always remains a high priority from system administrative as well as scientific points of view. It will be helpful that a work-group of account users recommend and enforce the MIAME compliance as a requirement for all data made accessible to the public.

As an informatics tool, ORMD is a secure management system for microarray projects and experiments. It can be used to facilitate microarray studies in olfactory as well as other systems. The authenticated login to access the private data and the regular backup of the database ensure security of the system and protection of the data.

### Conclusion

We have described the development of ORMD and its integration with the established OR gene database ORDB. ORDB provides information on the receptor genes and proteins, while ORMD provides microarray gene expression data of the ORs. These databases include hyperlinks that connect the genes and their expression data. Together, they provide a resource for researchers using different investigative approaches to understand how mammalian organisms perceive odors.

### Availability and requirements

Project name: ORMD

Project home page: <http://neurolab.med.yale.edu/ormd/>

ORDB home page: <http://senselab.med.yale.edu/ORDB/>

Operating system(s): Platform independent

Programming language: SQL, Java

Other requirements: Access to Oracle database

License: The SQL schema is freely available from the web-site

Any restrictions to use by non-academics: None

### Authors' contributions

NL carried out the design and implementation of ORMD Web interface and Oracle database schema, participated in integration of ORMD and ORDB, and drafted the manuscript. CJC carried out the update and maintenance of ORDB and participated in integration of ORMD and ORDB. MM carried out the system evaluation of ORMD and was a major data contributor of the current system. All authors read and approved the final manuscript.

### Acknowledgements

The work was supported by NIH grants K22LM008422, T15LM07056, P20LM07253, P01DC04732 and R01DC06213. The authors thank Dr. Rixin Wang and Ms. Jin Yang for technical support and Mr. George Michel for critical reading of the manuscript.

### References

1. Mombaerts P: **Genes and ligands for odorant, vomeronasal and taste receptors.** *Nat Rev Neurosci* 2004, **5**(4):263-278.
2. Zhang X, Rodriguez I, Mombaerts P, Firestein S: **Odorant and vomeronasal receptor genes in two mouse genome assemblies.** *Genomics* 2004, **83**(5):802-811.
3. Zhao H, Ivic L, Otaki JM, Hashimoto M, Mikoshiba K, Firestein S: **Functional expression of a mammalian odorant receptor.** *Science* 1998, **279**(5348):237-242.
4. Mombaerts P, Wang F, Dulac C, Chao SK, Nemes A, Mendelsohn M, Edmondson J, Axel R: **Visualizing an olfactory sensory map.** *Cell* 1996, **87**(4):675-686.
5. Johnson BA, Farahbod H, Leon M: **Interactions between odorant functional group and hydrocarbon structure influence activity in glomerular response modules in the rat olfactory bulb.** *J Comp Neurol* 2005, **483**(2):205-216.
6. Laska M, Teubner P: **Olfactory discrimination ability for homologous series of aliphatic alcohols and aldehydes.** *Chemical Senses* 1999, **24**(3):263-270.
7. Linster C, Hasselmo ME: **Behavioral responses to aliphatic aldehydes can be predicted from known electrophysiological responses of mitral cells in the olfactory bulb.** *Physiol Behav* 1999, **66**(3):497-502.
8. Shepherd GM: **Synaptic organization of the mammalian olfactory bulb.** *Physiol Rev* 1972, **52**(4):864-917.
9. Xu F, Greer CA, Shepherd GM: **Odor maps in the olfactory bulb.** *J Comp Neurol* 2000, **422**(4):489-495.
10. Buck L, Axel R: **A novel multigene family may encode odorant receptors: a molecular basis for odor recognition.** *Cell* 1991, **65**(1):175-187.
11. Zhang X, Rogers M, Tian H, Zhang X, Zou DJ, Liu J, Ma M, Shepherd GM, Firestein SJ: **High-throughput microarray detection of olfactory receptor gene expression in the mouse.** *Proc Natl Acad Sci USA* 2004, **101**(39):14168-14173.
12. Johnson BA, Ho SL, Xu Z, Yihon JS, Yip S, Hingco EE, Leon M: **Functional mapping of the rat olfactory bulb using diverse odors reveals modular responses to functional groups and hydrocarbon structural features.** *J Comp Neurol* 2002, **449**(2):180-194.
13. Schaefer ML, Young DA, Restrepo D: **Olfactory fingerprints for major histocompatibility complex-determined body odors.** *J Neuroscience* 2001, **21**(7):2481-2487.
14. Xu F, Liu N, Kida I, Rothman DL, Hyder F, Shepherd GM: **Odor maps of aldehydes and esters revealed by fMRI in the glomerular layer of the mouse olfactory bulb.** *Proc Natl Acad Sci USA* 2003, **100**(19):11029-11034.
15. Shepherd GM, Mirsky JS, Healy MD, Singer MS, Skoufos E, Hines MS, Nadkarni PM, Miller PL: **The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data.** *Trends Neurosci* 1998, **21**(11):460-468.

16. Crasto C, Marenco L, Miller P, Shepherd G: **Olfactory Receptor Database: a metadata-driven automated population from sources of gene and protein sequences.** *Nucleic Acids Res* 2002, **30(1)**:354-360.
17. Liu N, Xu F, Marenco L, Hyder F, Miller P, Shepherd GM: **Informatics approaches to functional MRI odor mapping of the rodent olfactory bulb: OdorMapBuilder and OdorMapDB.** *Neuroinformatics* 2004, **2(1)**:3-18.
18. Skoufos E, Marenco L, Nadkarni PM, Miller PL, Shepherd GM: **Olfactory receptor database: a sensory chemoreceptor resource.** *Nucleic Acids Res* 2000, **28(1)**:341-343.
19. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca-Serra P, Sansone SA: **ArrayExpress--a public repository for microarray gene expression data at the EBI.** *Nucleic Acids Res* 2003, **31(1)**:68-71.
20. Gollub J, Ball CA, Binkley G, Demeter J, Finkelstein DB, Hebert JM, Hernandez-Boussard T, Jin H, Kaloper M, Matese JC, Schroeder M, Brown PO, Botstein D, Sherlock G: **The Stanford Microarray Database: data access and quality assessment tools.** *Nucleic Acids Res* 2003, **31(1)**:94-96.
21. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R: **NCBI GEO: mining millions of expression profiles--database and tools.** *Nucleic Acids Res* 2005, **33(Database issue)**:562-566.
22. Nadkarni P, Marenco L, Chen R, Skoufos E, Shepherd G, Miller P: **Organization of heterogeneous scientific data using the EAV/CR representation.** *J Am Med Inform Assoc* 1999, **6(6)**:478-493.
23. Miller PL, Nadkarni P, Singer M, Marenco L, Hines M, Shepherd G: **Integration of multidisciplinary sensory data: a pilot model of the human brain project approach.** *J Am Med Inform Assoc* 2001, **8(1)**:34-48.
24. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M: **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.** *Nat Genet* 2001, **29(4)**:365-371.
25. Getchell TV, Liu H, Vaishnav RA, Kwong K, Stromberg AJ, Getchell ML: **Temporal profiling of gene expression during neurogenesis and remodeling in the olfactory epithelium at short intervals after target ablation.** *J Neurosci Res* 2005, **80(3)**:309-329.
26. Spehr M, Gisselmann G, Poplawski A, Riffell JA, Wetzel CH, Zimmer RK, Hatt H: **Identification of a testicular odorant receptor mediating human sperm chemotaxis.** *Science* 2003, **299(5615)**:2054-2058.
27. Liu AH, Zhang X, Stolovitzky GA, Califano A, Firestein SJ: **Motif-based construction of a functional map for mammalian olfactory receptors.** *Genomics* 2003, **81(5)**:443-456.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

