

Using Semantic Web Technologies to Power LungMAP, a Molecular Data Repository

Michelle C. Krzyzanowski, Josh L. Levy, Grier P. Page, Nathan C. Gaddis, Robert F. Clark

RTI International is a registered trademark and a trade name of Research Triangle Institute.



What drives advancements in biomedical research?



Importantly, open data that is easily accessible is key to progressing ongoing biomedical research and our understanding of what is known.

There is a need to develop standards for biomedical data

Large scale data is continually being generated by:



Hospitals



Academic Institutions



Industry and Biotech

As the amount of data collected grows exponentially, this necessitates a need to standardize stored data format and develop tools to store analyze the data to make it **easy to find**, **accessible** and **reusable**.

Working cooperatively to standardize data format and integration



a Feasible Standard

Establish a Primary Data Storage Location Recruit Researchers To Contribute and Submit Data

Our data follows a standard format, creating a freely accessible public source.

LungMAP: http://lungmap.net

🖄 Lung**MAP**

About LungMAP Resources Help ÷



LUNGMAP

LungMAP seeks to improve lung health. We offer a web-based data resource (BREATH) for investigation into processes that regulate lung development.

▼ Search

Donor tissue supplied by the University of Rochester Medical Center through the United Network for Organ Sharing and the Neonatal and Pediatric Organ Donor Program.

We are extremely grateful to the families who have generously given such

Data Inventory (mouse over to see # of experiments) Annotated Images

Histological stain

LungMAP stemmed from a NHLBI Initiative





We Treat Kids Better













National Heart, Lung, and Blood Institute

RTI's Contributions

As part of the Data Coordinating Center (DCC):

LungMAP Portal (Website)	BREATH (Database)
Website development and maintenance	Operating procedures for data management
Creation of web tools to browse and interpret data	Data processing, integration and maintenance into BREATH DB
Maintenance of Cloud services used	SPARQL queries
	Management of ontology of lung development, including cells, structures, and cross-species comparison

What data does LungMAP provide?



The proper development of an organism is carefully orchestrated by:

- Gene expression
- Protein-protein interactions
- Cell-cell interactions

All are critical for correct development of all organ systems and the organism as a whole.

What data does LungMAP provide?



Any changes to the expression, availability or interaction of genes, proteins and metabolites may result in improper development.

Therefore, mapping what genes are expressed, what proteins are present and the anatomical placement of cells may indicate what markers or combination of markers can lead to improper development.

Ontologies: Useful for mapping anatomical information



Using anatomical terms, lists of genes, proteins, lipids, etc., we apply their relationships through ontologies and triple store databases.

Anatomical terms become entities and their known biological hierarchy to each other establishes the relations.

Ontologies: Useful for mapping anatomical information

A subset of LungMAP's ontology



Using anatomical terms, lists of genes, proteins, lipids, etc., we apply their relationships through ontologies and triple store databases.

Anatomical terms become entities and their known biological hierarchy to each other establishes the relations.

LungMAP has been designed with the researcher in mind

At the web portal, experimental and biologic sample data is visualized for user-case scenarios, such as:

A researcher interested in browsing available data of a particular experiment type.

A researcher interested in finding data from all experiment types related to a specific term of interest.

A researcher seeking specific reagents or detecting certain genes or proteins during lung development.

Searching for information on a gene

A user arrives to LungMAP and conducts a search "**Acta1**"

After arriving at the gene information page, they click to view all **Single-cell RNA-seq experiments**

🖄 Lung MAP		About LungMAP Resources Help 🛃
Gene	•	Search
Details for ACTA1		
Gene Details 2D Image Experiments IF (0) ISH (0) Omics Experiments NanoString RNA (0) Proteomics (0) RNA-seq (1) Single-cell RNA-seq (2)	Symbol	Acta1
	Name	actin, alpha 1, skeletal muscle
	Synonyms	AA959943, Acta-2, Acta2, Acts, Actsk-1, Actsk1
	Species	Mus musculus (Taxon ID: 10090)
	Location	Chr 8:128161741-128161824, minus strand
	External Links	UniGene MGI International_Mouse_Knockout_Project MGI_GO_Links ENSEMBL Entrez

JOIN OUR MAILING LIST | CONTACT US | HOW TO CITE LUNGMAP | JOIN OUR FORUM

Supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number U01HL122638 © 2017 RTI International

Searching for information on a gene

A user arrives to LungMAP and conducts a search "**Acta1**"

After arriving at the gene information page, they click to view all **Single-cell RNA-seq experiments**

The following ontology patterns are used to retrieve all **single-cell RNA-seq experiments associated with Acta1**



Single-cell RNA-seq experiments for gene Acta1



JOIN OUR MAILING LIST | CONTACT US | HOW TO CITE LUNGMAP | JOIN OUR FORUM

Supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number U01HL122638 © 2017 RTI International

Using Ontologies and Triple Stores for Image Annotation

LungMAP features a large inventory of microscopic images of lung tissue.

To understand what is pictured, labels and annotations are required.

The following ontology patterns are used to create and retrieve annotations of images, including position on the image, label and annotator.

Image annotation uses OpenLayers.



Using Ontologies and Triple Stores for Image Annotation



LungMAP promotes collaboration



Biomedical research is everevolving. Therefore, data and associated ontologies must also be flexible to change.

Future Works

- Integromics tool for analyzing biological molecules across omics data types
- RESTful API for accessing BREATH database
- Improved data visualization tools
- Expansion of a story builder tool that enables users to create pages describing findings derived from BREATH data
- New data types: nanoDESI, methylation, metabolomics

Acknowledgements

Data Coordinating Center (RTI/Duke)

Nathan Gaddis Josh Levy Martin Duparc Michelle Krzyzanowski Stephen Hwang Grier Page Mary-Anne Ardini Robert Clark Carol Hill

Ontology Working Group

Gail Deutsch Helen Pan Susan Wert

Consortium Members

Namasivayam Ambalavanan Charles Ansong **Jacqueline Bagwell** Cliburn Chan Charles Frevert Davera Gabriel Sina Gharib James Hagood Carol Hill Jeanne Holden-Wiltse Anil Jegga Tom Mariani Anna Maria Masci Wei Shi David Warburton Kathryn Wikenheiser-Brokamp



NIH Award Number U01HL122638

Thank you.