

BRIEF COMMUNICATION

The Allergic Airway Inflammation Repository – a user-friendly, curated resource of mRNA expression levels in studies of allergic airways

D. R. Gawel^{1,*}, A. Rani James^{1,*}, M. Benson¹, R. Liljenström¹, A. Muraro², C. E. Nestor¹, H. Zhang^{1,*} & M. Gustafsson^{1,*}

¹Department of Clinical and Experimental Medicine, Centre for Individualised Medicine, Linköping University, Linköping, Sweden;

²Department of Women and Child Health, Referral Centre for Food Allergy Diagnosis and Treatment, Veneto Region, Padua University Hospital, Padua, Italy

To cite this article: Gawel DR, Rani James A, Benson M, Liljenström R, Muraro A, Nestor CE, Zhang H, Gustafsson M. The Allergic Airway Inflammation Repository – a user-friendly, curated resource of mRNA expression levels in studies of allergic airways. *Allergy* 2014; DOI: 10.1111/all.12432.

Keywords

allergy; mRNA microarray; repository.

Correspondence

Mika Gustafsson, PhD, The Centre for Individualised Medicine, Linköping University Hospital, 58185 Linköping, Sweden.

Tel.: +46-72-227-65-05

Fax: +46(0)-10-103-42-73

E-mail: mika.gustafsson@liu.se

*These authors contributed equally to the paper and serve as joint first and last authors.

Abstract

Public microarray databases allow analysis of expression levels of candidate genes in different contexts. However, finding relevant microarray data is complicated by the large number of available studies. We have compiled a user-friendly, open-access database of mRNA microarray experiments relevant to allergic airway inflammation, the Allergic Airway Inflammation Repository (AAIR, <http://aair.cimed.ike.liu.se/>). The aim is to allow allergy researchers to determine the expression profile of their genes of interest in multiple clinical data sets and several experimental systems quickly and intuitively. AAIR also provides quick links to other relevant information such as experimental protocols, related literature and raw data files.

Accepted for publication 22 April 2014

DOI:10.1111/all.12432

Edited by: Stephan Weidinger

High-throughput data stored in public databases like the gene expression omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and ArrayExpress (<http://www.ebi.ac.uk/arrayexpress>) constitute formidable resources for researchers. Given the explosion of high-throughput data generation over the last decade, these repositories have grown exponentially in size. Currently, GEO contains more than 150 000 mRNA microarray studies alone, as well as several other forms of high-throughput data. Because scientific journals require that high-throughput data from publications are stored in public databases, these grow at very high rates (1). The relevance to allergy researchers lies in that such databases can be used for a range of different *in silico* experiments. For example, to find the expression levels of *IL-4* in experiments relevant to allergy, this gene and a given cell type or experimental context can be searched for. This will generate a list of multiple

different experiments, which can be further explored to find the expression profiles of the gene as well as information about experimental conditions and a reference to the publication. It is also possible to identify co-expressed genes, which may help to identify novel, functionally related candidate genes which can be subsequently linked to data in other external databases such as Online Mendelian Inheritance in Man (OMIM) and Medline. Advanced searches can also be made and combined with a large number of other constantly developing databases, such as OMIM or Medline. It is, in fact, possible to perform research solely based on data sets from different databases, to generate and test hypotheses completely *in silico* (2, 3). These possibilities and developments suggest that increased use of public high-throughput databases is a valuable complement to allergy research. Ideally, such database searches should be part of the routine

Figure 1 The Allergic Airway Inflammation Repository web interface. The search enquiry for expression analysis involves four steps. (1) Either a gene name, an entrez id, or a probe id is entered in the search box A. (2) The choice of ID in the search is specified in the drop-down menu indicated by B. (3) A specific study of interest is then chosen in the drop-down menu C. A detailed description of that dataset is shown upon selection, with information on experimental conditions, available comparisons, and a link to the corresponding dataset on GEO. (4) Finally one or more diagrams from the different comparisons are chosen to be shown from the tick box list D. The user can also be provided with a list of all comparisons, together with data on fold change and *P*-values, ranked by *P*-values, by using the search box E.

experimental design and validation, similar to making a literature search. However, the scale and complexity of many repositories often acts as a barrier to their use by basic researchers. For example, finding relevant experiments among the 150 000 available data sets may involve making multiple searches across different cell types, diseases or experimental conditions, as well as individual assessments of the relevance of each study which can be both time-consuming and technically challenging.

To make this valuable data more accessible to allergy researchers, we have constructed an independent, open-access database of allergy-relevant data sets from studies related to allergic airway inflammation, with an intuitive, user-friendly interface: the Allergic Airway Inflammation Repository (AAIR, <http://aair.cimed.ike.liu.se/>).

The database provides the user with information on the expression values of user-defined genes in a given study and under given experimental conditions, as well as statistical components such as mean and median values, fold change and *P*-values from comparisons between the different experimental conditions in the selected studies. It can also provide a list of the associated probes with a given gene, ranked by *P*-values of comparisons over different experimental conditions, in order to easily determine under which experimental conditions a given gene is differentially expressed.

The AAIR includes a short introduction of its content and how to use it, as well as links to (i) a list of selected data sets relevant to allergic airway inflammation, (ii) a large-scale repository for comprehensive searches, (iii) a database that contains information about mRNA expression levels in different cell types and tissues, as well as summary information about gene function (<http://biogps.org>).

The selected data sets include data from studies on allergic airway inflammation induced by allergens as well as microbes. These studies have been performed on CD4⁺ and CD8⁺ T cells, PBMC and epithelial cells from different levels of the respiratory tract (3–13). The database also includes a study of monozygotic twins, discordant for seasonal allergic rhinitis (14). Microarray data of RNA expression levels in different T-cell subsets have also been included (15).

The repository can be used either to (i) search for the expression levels of a gene of interest in an individual study or (ii) obtain a summary table of fold change ratio over mean expression levels in different study groups, together with *P*-values for the difference (Fig. 1). Importantly, the database also provides a detailed description of the experimental protocol used in each study and simple links to the articles related to each study and the raw data for download.

For detailed information of our methods and statistical analyses, we refer to the supplementary document 1 (Appendix S1).

In summary, we have constructed a user-friendly microarray repository relevant to allergic airway inflammation. We hope this will serve as a valuable complement to large-scale databases, and also increase the use of them.

Author contributions

D.R.G. designed research, acquired and analysed data and wrote the paper; M.B. designed research and wrote the paper; R.L. designed research and wrote the paper; A.M. acquired data; C.E.N. designed research and wrote the paper; A.R.J. designed research, developed the database and website and wrote the paper; H.Z. designed research; and M.G. designed research and acquired data.

Funding

This work was supported by the Swedish Research Council.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Methods.

References

1. Wilhite SE, Barrett T. Strategies to explore functional genomics data sets in NCBI's GEO database. *Methods Mol Biol* 2012;**802**:41–53.
2. Pedicini M, Barrenas F, Clancy T, Castiglione F, Hovig E, Kanduri K et al. Combining network modeling and gene expression microarray analysis to explore the dynamics of Th1 and Th2 cell regulation. *PLoS Comput Biol* 2010;**6**:e1001032.
3. Barrenas F, Chavali S, Alves AC, Coin L, Jarvelin MR, Jornsten R et al. Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms. *Genome Biol* 2012;**13**:R46.
4. Nestor CE, Barrenas F, Wang H, Lentini A, Zhang H, Bruhn S et al. DNA methylation changes separate allergic patients from healthy controls and may reflect altered CD4⁺ T-cell population structure. *PLoS Genet* 2014;**10**:e1004059.
5. Wagener AH, Zwinderman AH, Luiten S, Fokkens WJ, Bel EH, Sterk PJ et al. dsRNA-induced changes in gene expression profiles of primary nasal and bronchial epithelial cells from patients with asthma, rhinitis and controls. *Respir Res* 2014;**15**:9.
6. Aguerri M, Calzada D, Montaner D, Mata M, Florido F, Quiralte J et al. Differential gene-expression analysis defines a molecular pattern related to olive pollen allergy. *J Biol Regul Homeost Agents* 2013;**27**:337–350.
7. Wagener AH, Zwinderman AH, Luiten S, Fokkens WJ, Bel EH, Sterk PJ et al. The impact of allergic rhinitis and asthma on human nasal and bronchial epithelial gene expression. *PLoS One* 2013;**8**:e80257.
8. Giovannini-Chami L, Marcet B, Moreilhon C, Chevalier B, Illie MI, Lebrigand K et al. Distinct epithelial gene expression phenotypes in childhood respiratory allergy. *Eur Respir J* 2012;**39**:1197–1205.
9. Bosco A, McKenna KL, Firth MJ, Sly PD, Holt PG. A network modeling approach to analysis of the Th2 memory responses underlying human atopic disease. *J Immunol* 2009;**182**:6011–6021.
10. Kicic A, Hallstrand TS, Sutanto EN, Stevens PT, Kobor MS, Taplin C et al. Decreased fibronectin production significantly contributes to dysregulated repair of asthmatic epithelium. *Am J Respir Crit Care Med* 2010;**181**:889–898.
11. Chamberland A, Madore AM, Tremblay K, Laviolette M, Laprise C. A comparison of two sets of microarray experiments to define allergic asthma expression pattern. *Exp Lung Res* 2009;**35**:399–410.
12. Vroiling AB, Jonker MJ, Luiten S, Breit TM, Fokkens WJ, van Drunen CM. Primary nasal epithelium exposed to house dust mite extract shows activated expression in allergic individuals. *Am J Respir Cell Mol Biol* 2008;**38**:293–299.
13. Tsitsiou E, Williams AE, Moschos SA, Patel K, Rossios C, Jiang X et al. Transcriptome analysis shows activation of circulating CD8⁺ T cells in patients with severe asthma. *J Allergy Clin Immunol* 2012;**129**:95–103.
14. Sjogren AK, Barrenas F, Muraro A, Gustafsson M, Saetrom P, Wang H et al. Monozygotic twins discordant for intermittent allergic rhinitis differ in mRNA and protein levels. *Allergy* 2012;**67**:831–833.
15. Zhang H, Nestor CE, Zhao S, Lentini A, Bohle B, Benson M et al. Profiling of human CD4⁺ T-cell subsets identifies the TH2-specific noncoding RNA GATA3-AS1. *J Allergy Clin Immunol* 2013;**132**:1005–1008.

Appendix S1. Methods

Data processing and statistical calculations

The data in the AAIR database are collected from public repositories, such as the Gene Expression Omnibus (GEO). The raw expression data from each study have been quantile normalised. Mean and median values have been calculated for each experimental condition separately (e.g. patients or healthy controls). Fold change has been calculated as the ratio between the mean expression values from two experimental conditions. P-values have been calculated using the LIMMA student's t-test, using the relevant R-package in R [1]. Raw expression values were log2 transformed prior to statistical testing.

Database development and website design

The database was first created in backend, in SQL server 2012. The front end web portal /website is developed in Visual Studio 2012 (.net Framework 4.5). We have used mainly chart controls and user controls, apart from the standard ones such as dropdown list, text box, and check list. The look and behaviour of all other pages is inherited from the master page.

The charts were created with Microsoft Chart Controls, which is a tool integrated in Microsoft Visual Studio.

Data storage

The data is stored as tables in the SQL server 2012 and the cloud server. We have used Windows Azure as the cloud server, which is a service from Microsoft. Unlike similar cloud services, Azure allows the users to make relational queries against the structured or unstructured stored data. In addition to that it has specific features such as querying data, data analysis and data synchronization. In addition to the table structure we have created a stored procedure with "UNION" statement for all the tables in the database, in order to query against the search gene name. The stored procedure returns the gene name from the respective study with statistical values, such as P-value, and other values, such as probe name and entrez id.

[1] Smyth, GK (2005). Limma: linear models for microarray data. In: 'Bioinformatics and Computational Biology Solutions using R and Bioconductor'. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds), Springer, New York, pages 397-420.