A curated collection of transcriptome datasets to investigate the molecular mechanisms of immunoglobulin E-mediated atopic diseases

Susie S. Y. Huang[1], Fatima Al Ali[1], Sabri Boughorbel[1], Mohammed Toufiq[1], Damien Chaussabel[1], Mathieu Garand*[1]

[1]Systems Biology and Immunology Department, Translational Medicine, Sidra Medicine, Doha, Qatar.

*Corresponding authors: mathieu.garand@gmail.com

Manuscript length: 2749 words
Number of tables: 2
Number of figures: 3
Supplement: 1 table, 2 figures, 1 additional information

ABSTRACT (Word count: 180)

Prevalence of allergies has reached ~50% of industrialized populations and with children under ten being the most susceptible. However, the combination of the complexity of atopic allergy susceptibility/development and environmental factors has made identification of gene biomarkers challenging. The amount of publicly accessible transcriptomic data presents an unprecedented opportunity for mechanistic discoveries and validation of complex disease signatures across studies. However, this necessitates structured methodologies and visual tools for the interpretation of results. Here, we present a curated collection of transcriptomic datasets relevant to immunoglobin E (IgE)-mediated atopic diseases (ranging from allergies to primary immunodeficiencies). 30 datasets from the Gene Expression Omnibus (GEO), encompassing 1761 transcriptome profiles, were made available on the Gene Expression Browser (GXB), an online and open-source web application that allows for the query, visualization, and annotation of metadata. The thematic compositions, disease categories, sample number, and platforms of the collection are described. Ranked gene lists and sample grouping are used to facilitate data visualization/interpretation and are available online via GXB (http://ige.gxbsidra.org/dm3/geneBrowser/list). Dataset validation using associated publications showed good concordance in GXB gene expression trend and fold-change.

Database URL: http://ige.gxbsidra.org/dm3/geneBrowser/list

Keywords: immunology, gene expression browser, atopy, immunoglobulin E, curated dataset collection.

INTRODUCTION

Allergic disease is highly prevalent and currently reaches ~50% of the populations in industrialized nations (1). Although the generation of allergic responses is well-understood, the early sensitization steps and factors contributing to the development of immunoglobin E (IgE)-mediated diseases remain unclear. IgE is the major mediator of atopic response in humans, whereas atopy represents the predisposition to become over IgE-sensitized to allergens. However, not all encounters with a potential allergen will lead to sensitization. Similarly, not all sensitizations will result in a symptomatic allergic response even in atopic individuals.

The effect of IgE spans across multiple systems. In the circulation system, IgE increases flow and permeability of the blood vessels, fluid and protein in tissues, as well as flow to the lymph nodes. In the airway, IgE decreases air conduct diameter, increases mucus congestion, and can induce blockage. In the gastro-intestinal (GI) tract, IgE increases fluid secretion, peristalsis, and expulsion-diarrhea. The immunoglobin has also been suggested to play a role in the defense against parasite infection and as a general gate keeper for any foreign materials entering the body (2,3). Ultimately, IgE is involved in the normal spectrum of reaction to expulse foreign material from the body; hence, protecting by elimination. Nevertheless, over sensitization, which is developed by unknown mechanism(s), can lead to imbalance and pathology in affected individuals.

Two main groups of immune signals initiate the production of IgE in response to an antigen: 1) the signals that drive the differentiation of CD4 naive T cell to T helper type 2 (Th2) cells and 2) the cytokines and co-stimulatory molecules secreted by Th2 cells, which subsequently promote T follicular helper cells-induced immunoglobulin B cells class switch towards IgE production. Antigen characteristics, such as concentration and localization of the encounter (i.e. tissue, mucosa, circulation), can also affect Th2 cell induction. IL-4, IL-13, and STAT6 are key mediators of Th2 responses and IgM class switch to IgE. IL-4 secretion and mast cell CD40 surface expression also contribute to the IgE class switch to IgE at the site of allergic reaction.

Systemic levels of IgE alone is not a sufficient indicator for allergy risk (4). Peripheral blood IgE level can increase upon sensitization, but not reliable enough for deducing a diagnosis of allergy or allergen type (5). Concentration, binding strength or affinity, specificity, and portion of specific IgE to total IgE are all factors in translating a humoral IgE response into a clinical symptom (6). However, genetic component exists in allergic disease. Studies have demonstrated strong heritable components of allergic diseases and atopy, estimated at 33%-76% (7,8). Genetic components of food allergy and asthma have also been reviewed (9–12). Although single susceptibility gene have been identified for certain allergies (13–16), most genome-wide association studies reported multiple loci of susceptibility or gene regions to a wide range of allergy development (17–19). Environmental factors further complicate allergic susceptibility - the hygiene hypothesis of allergic susceptibility

has received much attention in the last few decades (20). Other exposure-related factors (e.g. diet, pollution, tobacco smoke) are also likely to have contributed to the increasing susceptibility to allergic diseases in developed nations (1).

IgE may also be implicated in some primary immunodeficiency diseases (PIDs). For CD40 ligand and CD40 deficiency, circulating IgE+ plasma cells are absent (21). For others, IgE plasma levels are elevated: DOCK8 deficiency, AD-HIES Job's syndrome, Comel-Netherton syndrome, PGM3 deficiency, IPEX (immune dysregulation, polyendocrinopathy, enteropathy X-linked), and Tyk2 deficiency. Large phenotypic heterogeneity is observed among these PIDs; however, the underlying mechanisms is not completely understood. This knowledge gap mirrors our limited understanding of the involved gene products mediating phenotypes atopic diseases.

There is a crucial need to pursue genetic markers of an atopic constitution amidst growing concern over the increasing abundance of IgE atopy and that current treatments remain mostly palliative. As both environmental interaction and genetic predisposition forms the determinant of allergy development, extrapolating potential gene markers of allergy atopy from a broad range of studies with associated clinical data is highly relevant, particularly given the high variations of the disease clinical phenotypes. With the advances in high through-put platforms for transcriptomic analyses, a large number of gene expression datasets is regularly deposited on publicly accessible repositories such as the NCBI Gene Expression Omnibus (GEO). GEO currently hosts over 90 000 datasets comprising over 2 million samples (https://www.ncbi.nlm.nih.gov/geo/) and presents an enormous opportunity for data mining across multiple studies.

The web-based Gene Expression Browser (GXB) is an open-source interface that allows for custom compilation of selected datasets (i.e. of interest to users) and facilitates visualization of gene expression data. GXB has been previously described (22) and used to generate a number of Data Notes (23–28). As well, GXB is a useful tool for novel gene/function discovery (29) and in system re-analysis approaches (30).

A search strategy was implemented to identify GEO datasets relevant to IgE-related atopic disease, including PIDs, and uploaded those datasets to the GXB platform online. The associated study metadata, such as the detailed gene information, relevant literature, study design, and sample information, were also uploaded to facilitate in-depth interpretation. In creating this collection of datasets, we aimed provide a resource which will facilitate risk prediction and interception of allergic diseases. The datasets were retrieved from publicly available GEO series and selected by relevance to IgE and atopic diseases and filtered by analysis platform and species. As previously demonstrated, data mining and re-analysis/re-interpretation of large and publicly available dataset is a promising avenue (31) to elucidate complicated diseases such as IgE-related atopy.

METHODS

Justification of data selection and filter

The focus of the GEO dataset selection was primarily on whether or not the dataset involved 1) IgE production or function or 2) if the pathology studied implicated directly or indirectly IgE production or function (Table 1A). We ensured a broad search approach by combining the 7 independent search results (Table 1A) using the "Merge collection" function available on "My NCBI" (https://www.ncbi.nlm.nih.gov/myncbi/) (Table 1B) and by performing another search using an assembly of all the terms used in Table 1A (Table 1C). 203 potentially relevant datasets were identified from the initial query (Table 1D). The query results were then manually filtered to restrict datasets to human sample, expression profiling by microarrays, and relevance to IgE-related atopic diseases. The process involved inspecting the study description, design, and sample type for each dataset and resulted in 30 curated datasets. The following criteria were deemed very important: clear description of tissue type, comparisons between patient vs healthy or stimulated vs unstimulated samples, and disease category implicating IgE. Furthermore, datasets/studies that are indirectly relevant to IgE-mediated atopic disease, such as gene expression in B cells after tonsillectomy, were also included as they were deemed valuable to 1) the discovery of putative novel gene-disease association, 2) to improving our knowledge of adaptive immunity, and/or 3) to increasing our knowledge about factors affecting IgE production.

The web-based GXB platform: a visualization tool for gene expression data

GXB is a valuable tool for training researchers about the reductionist investigative approaches (see description of such program in (31)). The creation of a web-based GXB platform was previously described in detail (22). In brief, the GXB is a simple interactive interface designed for visualization of large quantities of heterogenous data (Suppl. Figure 2). The platform allows for customizable data plots with overlapping metadata information, changeable sample order, as well as generation of sharable mini-URLs that encapsulate information about the display settings in use. The dataset navigation page allows for quick identification of datasets of interest either through filtering using pre-define lists or via query terms. The user has access to multiple functionalities within the GXB graphic interface. In brief, the data-viewing interface enables interactive browsing and graphic representation of large-scale data. This interpretable format displays ranked gene lists and expression results. The interface also allows for user flexibility in terms of changing how the gene list is ranked, the method used for ranking, sample grouping (i.e. disease type), sample sorting (i.e. gender or age) and view type (i.e. bar or chart). The end user can browse through the datasets, format graph for a selected gene within a dataset, and export data (i.e. annotation, FC, signal, groups). The original GEO data and annotation are accessible from the GXB interface (Downloads tabs) as well as from their GEO page (links provided on Study tab). The associated source code and

instructions are publicly available (https://github.com/BenaroyaResearch/gxbrowser)], as well as the necessary R scripts (https://github.com/BenaroyaResearch/gxrscripts).

Construction of the dataset collection on GXB

The selected datasets were downloaded from GEO in SOFT file format and uploaded (http://ige.gxbsidra.org/dm3/geneBrowser/list) via the GXB interface (accessed by navigating the Tools menu located in the top-right corner of the GXB webpage; Tools / Chips Loaded / Upload Expression Data). The SOFT files contain metadata and normalized signal intensity data, generated by methods indicated by the author(s). These SOFT files can be analyzed directly for differentially expressed genes; thus, no additional processing was required. Using the "Sample Set Annotation Tool" of the GXB interface, the datasets were annotated according to the information provided on GEO. The raw signal data type of the dataset (ex. raw signal, $\log_2$ transformed and etc.,) are also specified for each GEO entry. The default data display is in linear scale (see further details below in "Presentation of datasets"). When necessary, the sample annotation file (which is part of the SOFT file) were edited in order to add group information; as it was important to identify groups in order to compute fold change (FC). For each dataset, individual samples were grouped based on relevant study variables. Three datasets (GSE19190, GSE75603, and GSE8507) were split via GXB's "Sample Set Annotation Tool/Group Sets" interface to provide a more meaningful group comparison (these datasets share the same GSE number in Table 2). Genes were then ranked based on FCs of the specified two-group comparison. All the information annotated and presented on GXB is assembled in a SQL database (Data Citation 1).

Data availability

The curated datasets collected that have been described in this data note were assembled from the public repository NCBI GEO website: http://www.ncbi.nlm.nih.gov/gds/. In this study, we cited each dataset GEO accession number, and raw signal and annotation files are made available for download from the GXB web-application (http://ige.gxbsidra.org/dm3/geneBrowser/list).

RESULTS

Description of datasets

After applying the filtering strategy as previously described, we curated 30 datasets encompassing 1761 transcriptome profiles relevant to IgE-related atopic diseases. Detailed information on each dataset is presented in Table 2 and the summary of the data collection is presented in an aggregation of pie charts (Figure 1). The data collection includes a wide range of studies, sample types, as well as diseases. In total, 12 different microarray platforms are represented, with the majority being Affymetrix Human Array chips (various version). 17 datasets were generated from *in vitro* studies and 13 from *ex vivo* studies. Ten sample types are represented, with the most

abundant being PBMC (n = 8) and airway epithelial cells (n = 5). Sample size of each dataset ranges from 5 to 628, with most studies having 10-50 samples. The dataset collection covered 7 main disease categories, including allergy, asthma, healthy responses, hyper IgE syndrome (HIES), dermatitis, atopic irritable bowel syndrome (IBS), and Urticaria. In the majority of the studies, comparisons are made between patients and health controls (n = 17), followed by stimulation (n = 6) and time course (n = 5). The frequency of the terms retrieved from the associated GEO descriptions and those derived from the MeSH/keywords of the published articles are visualized through word clouds presented in Figure 2.

Presentation of datasets

On the graphic interface of GXB, genes expression values are displayed in linear scale. The original signal data type can be found under the Info tab of individual datasets. The associated metadata are available under the Samples tab and this information can be used for graphical overlaying (via "Overlays" dropdown menu). The FCs are also displayed in linear scale (see Supplemental Information for a detailed example of FC calculation). For FC analysis, each subject/sample are grouped according to the experimental design (experimental variable vs controls) and/or, if available, as per the corresponding publications. It is also possible to rank the genes based on expression difference by selecting "Advanced" in the "Rank Lists" dropdown menu of the graphic interface. This display option can be more robust than FC when low expression is observed in one group. It is important to note that integration and re-analysis of the datasets is not the intent of this collection. Therefore, a more meaningful use of this collection lies in the comparison of FC expression for a gene across multiple relevant datasets (i.e. reductionist approach (31).

Dataset validation

Quality assessment of the datasets was performed by looking for key gene expression, i.e. marker genes, and highly differentiated genes as indicated in the associated publications. Table 2 includes the type of validation, the genes used, as well as the strength of the validation and the associated comments. Certain datasets were split in two to facilitate comparisons (GSE19190, GSE75603, and GSE8507); in these cases, validation is indicated for one of the two datasets as indicated in Table 2. Data validation was achieved for most datasets except for 6, due to having no publications (GSE14842, GSE37157, and GSE41861) or incomparable units (i.e. z-score) and/or absence of FC information (GSE64639, GSE54522 and GSE44956). Trend validations were done on 4 datasets that had no linked publications (GSE13619 and GSE99948) or had incomparable units to GXB (GSE70760 and GSE72542). But when possible, literature values were used for additional validation of the trend (i.e., GSE92688and GSE99948). Examples of the validation results are presented in Figure 3.

DISCUSSION

Potential application of the dataset collection

To demonstrate the potential use of the dataset collection, gene expression profile of different tissues for house dust mite (HDM) allergy were compared. Nasal epithelial (GSE19190) and PBMC-derived CD4 T cells (GSE70760) gene expression profiles (detailed in Table 2) were compared. In both dataset, genes associated with the Th2 pathway/axis were increased in HDM-sensitized patients compared against healthy controls (32,33); illustrating that Th2 response may results in the symptomatic phase. This is further supported by the IgE levels and sIgG/sIgE ratio in the same study (33).

SERPINB2, a gene coding for the inhibitor of plasminogen activator PAI-2, is markedly upregulated in patients in both datasets, hence tissues (Suppl. Figure 1A from GSE70760 and Figure 1B from GSE19190). This is consistent with previous association of asthma severity and biomarker panel including SERPINB2 from PBMC (34) and correlation of SERPINB2 expression in respiratory epithelial cells with atopic asthma severity (35). SERPINB2 has been reported to have a role in the interleukin-12-mediated signaling pathway; evidence from mice showed that SERPINB2 regulates IFNg production, causing down regulation of Th1 cytokines in macrophages (36).

The expression profiles of SERPINB2 in both tissues suggest an important role of the gene as a first line regulator of immune response, perhaps by preventing excessive Th1 response. Interestingly, IFNg was not decreased in PBMC-derived T cells, but IFNGR1 was, suggesting that in T cells, SERPINB2 may exert its role on the expression of the receptor rather than on IFNg itself. In Schroder et al (2010), stimulation of SERPINB2 -/- cells with antiCD40/IFNg resulted in greater Th1 cytokine production, thus supporting the idea that SERPINB2 affects IFNGR (36).

However, certain genes are found to be differently expressed between nasal epithelium or PBMC-derived T cells. An explanation is that the activation of Th1 immune response may be different in the target tissue and peripheral circulation. For example, IL1B, a Th1 promoting cytokine is found to be increased in CD4+ T cells in allergic patients (Suppl. Figure 1C, from GSE70760), but in the nasal epithelium, the gene is not upregulated in patients with severe case of allergic rhinitis ("uncontrol asthma") compared against healthy control (Suppl. Figure 1D, from GSE19190). Further hypothesis-generating comparisons can be made by comparing the top differently expressed genes in both datasets as listed in Suppl. Table 1.

Conclusion

The dataset collection may be useful for exploring specific gene signatures in response to natural antigen/allergen exposure (i.e. allergic patients vs control) and delineate the major genetic drivers associated with increased level of IgE (i.e. cellular responses to specific antigen/allergen). Furthermore, comparative investigation of similarities and differences in expression of genes between datasets can highlight key mechanistic differences of immune

signaling and/or provide insights for tissue-targeted intervention. In compiling the present dataset collection, we hope to offer a resource that may improve accessibility of public omics data to researchers in this field.

DATA CITATION

1. IgE_GXB_Database *Figshare* http://doi.org/10.6084/m9.figshare.7176851 (2018).

AUTHOR CONTRIBUTIONS

MG and SSYH contributed to conceptualization. MG, SSYH, and FA contributed to data curation and validation. MG and SSYH led, and FA supported, investigation and visualization. MG and SSYH performed formal analyses. SB and MT contributed to the maintenance of software. MG contributed writing – original draft, methodology, and project administration. MG and SSYH led, and FA, SB, MT, DC supported, writing – review & editing. MG lead and SSYH supported supervision. DC contributed funding acquisition and resources. The contributor's roles listed above (underlined) follow the Contributor Roles Taxonomy (CRediT) described in Nature Communication 2014 (37) and managed by The Consortia Advancing Standards in Research Administration Information (CASRAI) (https://casrai.org/credit/).

COMPETING INTERESTS

No competing interests were disclosed.

FUNDING

All authors listed on this publication are affiliated with Sidra Medicine and the work is supported by the Qatar Foundation and a Qatar National Research Fund grant NPRP10-0205-170348.

ACKNOWLEDGEMENT

We would like to thank all the investigators who decided to make their datasets publicly available by depositing them into the NCBI GEO repository.

REFERENCES

1. Pawankar, R., Holgate, S., Canonica, G. W., et al. *WAO White Book on Allergy 2013 Update*; World Health Organization: Malwaukee, Wisconsin, USA, 2013.
2. Nyan, O. A., Walraven, G. E., Banya, W. A., et al. (2001) Atopy, intestinal helminth infection and total serum IgE in rural and urban adult Gambian communities, *Clin. Exp. Allergy*, **31**, 1672–1678.
3. Gurish, M. F., Bryce, P. J., Tao, H., et al. (2004) IgE Enhances Parasite Clearance and Regulates Mast Cell Responses in Mice Infected with Trichinella spiralis, *The Journal of Immunology*, **172**, 1139–1145.

4. Galli, S. J. and Tsai, M. (2012) IgE and mast cells in allergic disease, *Nat Med*, **18**, 693–704.
5. Takhar, P., Corrigan, C. J., Smurthwaite, L., et al. (2007) Class switch recombination to IgE in the bronchial mucosa of atopic and nonatopic patients with asthma, *J. Allergy Clin. Immunol.*, **119**, 213–218.
6. Hamilton, R. G. and Oppenheimer, J. (2015) Serological IgE Analyses in the Diagnostic Algorithm for Allergic Disease, *The Journal of Allergy and Clinical Immunology: In Practice*, **3**, 833–840.
7. Hopp, R. J., Bewtra, A. K., Watt, G. D., et al. (1984) Genetic analysis of allergic disease in twins, *J. Allergy Clin. Immunol.*, **73**, 265–270.
8. Lichtenstein, P. and Svartengren, M. (1997) Genes, environments, and sex: factors of importance in atopic diseases in 7-9-year-old Swedish twins, *Allergy*, **52**, 1079–1086.
9. Hong, X., Tsai, H.-J. and Wang, X. (2009) Genetics of Food allergy, *Curr Opin Pediatr*, **21**, 770–776.
10. Madore, A.-M. and Laprise, C. (2010) Immunological and genetic aspects of asthma and allergy, *J Asthma Allergy*, **3**, 107–121.
11. Ober Carole and Yao Tsung-Chieh (2011) The genetics of asthma and allergic disease: a 21st century perspective, *Immunological Reviews*, **242**, 10–30.
12. Hong, X. and Wang, X. (2012) Early life precursors, epigenetics, and the development of food allergy, *Semin Immunopathol*, **34**, 655–669.
13. Jiang, P., Liu, J., Yan, X.-B., et al. (2009) Several interleukin-4 and interleukin-13 gene single nucleotide polymorphisms among Chinese asthmatic patients, *Allergy Asthma Proc*, **30**, 413–418.
14. Bønnelykke, K., Sleiman, P., Nielsen, K., et al. (2014) A genome-wide association study identifies *CDHR3* as a susceptibility locus for early childhood asthma with severe exacerbations, *Nature Genetics*, **46**, 51–55.
15. Gharagozlou, M., Behniafard, N., Amirzargar, A. A., et al. (2015) Association between single nucleotide polymorphisms of the interleukin-4 gene and atopic dermatitis, *Acta Dermatovenerol Croat*, **23**, 96–100.
16. Hong, X., Hao, K., Ladd-Acosta, C., et al. (2015) Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children, *Nature Communications*, **6**, 6304.
17. Tamari, M., Tanaka, S. and Hirota, T. (2013) Genome-wide association studies of allergic diseases, *Allergology International*, **62**, 21–28.
18. Bønnelykke, K., Matheson, M. C., Pers, T. H., et al. (2013) Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization, *Nature Genetics*, **45**, 902–906.
19. Portelli, M. A., Hodge, E. and Sayers, I. Genetic risk factors for the development of allergic disease identified by genome-wide association, *Clinical & Experimental Allergy*, **45**, 21–31.
20. Yazdanbakhsh, M., Kremsner, P. G. and Ree, R. van (2002) Allergy, Parasites, and the Hygiene Hypothesis, *Science*, **296**, 490–494.
21. Fuleihan, R., Ramesh, N., Loh, R., et al. (1993) Defective expression of the CD40 ligand in X chromosome-linked immunoglobulin deficiency with normal or elevated IgM., *PNAS*, **90**, 2170–2173.

22. Speake, C., Presnell, S., Domico, K., et al. (2015) An interactive web application for the dissemination of human systems immunology data, *Journal of Translational Medicine*, **13**, 196.
23. Rinchai, D., Boughorbel, S., Presnell, S., et al. (2016) A curated compendium of monocyte transcriptome datasets of relevance to human monocyte immunobiology research, *F1000Research*, **5**, 291.
24. Blazkova, J., Boughorbel, S., Presnell, S., et al. (2016) A curated transcriptome dataset collection to investigate the immunobiology of HIV infection, *F1000Research*, **5**, 327.
25. Marr, A. K., Boughorbel, S., Presnell, S., et al. (2016) A curated transcriptome dataset collection to investigate the development and differentiation of the human placenta and its associated pathologies, *F1000Research*, **5**, 305.
26. Rahman, M., Boughorbel, S., Presnell, S., et al. (2016) A curated transcriptome dataset collection to investigate the functional programming of human hematopoietic cells in early life, *F1000Research*, **5**, 414.
27. Roelands, J., Decock, J., Boughorbel, S., et al. (2017) A collection of annotated and harmonized human breast cancer transcriptome datasets, including immunologic classification, *F1000Res*, **6**, 296.
28. Mackeh, R., Boughorbel, S., Chaussabel, D., et al. (2017) -A curated transcriptomic dataset collection relevant to embryonic development associated with in vitro fertilization in healthy individuals and patients with polycystic ovary syndrome, *F1000Research*, **6**, 181.
29. Rinchai, D., Kewcharoenwong, C., Kessler, B., et al. (2016) Increased abundance of ADAM9 transcripts in the blood is associated with tissue damage, *F1000Res*, **4**.
30. Rinchai, D., Presnell, S., Vidal, M., et al. (2015) Blood Interferon Signatures Putatively Link Lack of Protection Conferred by the RTS,S Recombinant Malaria Vaccine to an Antigen-specific IgE Response, *F1000Res*, **4**, 919.
31. Chaussabel, D. and Rinchai, D. (2018) Using "collective omics data" for biomedical research training, *Immunology*, **155**, 18–23.
32. Giovannini-Chami, L., Marcet, B., Moreilhon, C., et al. (2012) Distinct epithelial gene expression phenotypes in childhood respiratory allergy, *Eur. Respir. J.*, **39**, 1197–1205.
33. Holt, P. G., Strickland, D., Bosco, A., et al. (2016) Distinguishing benign from pathologic TH2 immunity in atopic children, *J. Allergy Clin. Immunol.*, **137**, 379–387.
34. Baos, S., Calzada, D., Cremades-Jimeno, L., et al. (2018) Nonallergic Asthma and Its Severity: Biomarkers for Its Discrimination in Peripheral Samples, *Front Immunol*, **9**, 1416.
35. ELBadawy, N. E., Abdel-Latif, R. S. and El-Hady, H. A. (2017) Association between SERPINB2 Gene Expression by Real Time PCR in Respiratory Epithelial Cells and Atopic Bronchial Asthma Severity, *Egypt J Immunol*, **24**, 165–181.
36. Schroder, W. A., Le, T. T. T., Major, L., et al. (2010) A physiological function of inflammation-associated SerpinB2 is regulation of adaptive immunity, *J. Immunol.*, **184**, 2663–2670.
37. Allen, L., Scott, J., Brand, A., et al. (2014) Publishing: Credit where credit is due, *Nature*, **508**, 312–313.

38. Holland, S. M., DeLeo, F. R., Elloumi, H. Z., et al. (2007) STAT3 mutations in the hyper-IgE syndrome, *N. Engl. J. Med.*, **357**, 1608–1619.

39. Kosoy, R., Agashe, C., Grishin, A., et al. (2016) Transcriptional Profiling of Egg Allergy and Relationship to Disease Phenotype, *PLoS ONE*, **11**, e0163831.

LIST OF TABLES AND FIGURES

LIST OF SUPPLEMENT TABLES AND FIGURES

LIST OF SUPPLEMENT INFORMATION

FIGURE LEGENDS

Figure 1. Characteristic of the curated dataset collection. Proportion and number of the study type, sample type, sample size, disease categories, and type of comparison performed for fold-change calculation on the GXB are depicted.

Figure 2. Word clouds depicting the frequency of (A) the terms retrieved from the associated GEO descriptions of the datasets and (B) those derived from the MeSH/keywords of the associated published articles.

Figure 3. Examples of dataset validation. Differently expressed genes from 2 datasets were compared with the results presented in the respective publications. A) When comparing between Job's syndrome patients (HIES_control_0) and healthy controls (Healthy_control_0), the mean fold-change of CD151 was 1.7 on the GXB (GSE8507-PBMC). The reported value in Holland et al (2010) was 2.0 (38). B) When comparing between PBMC samples from egg allergic patients (Allergic_Egg) and allergic egg-tolerant controls (Allergic_Control), the mean fold-change of CEACAM1 was 1.69 on the GXB (GSE88796). Kosoy et al. (2016) reported a mean fold-change of 1.6 for the same gene (39). The overall trends of gene expression are conserved between GXB and published data.

Supplement Figure 1. Comparison of gene expression profile in different tissue from house dust mite (HDM)-sensitized individuals. Gene expression data from two datasets present in our collection are shown: 1) Gene expression patterns in house dust mite stimulated CD4 T cells and IgG to IgE ratios - GSE70760, and 2) Distinct epithelial gene expression phenotypes in childhood respiratory allergy - GSE19190 - Disease State. SERPINB2 (A and B) and IL1β (C and D) gene expression. HDM = house dust mite; Control = Healthy = healthy individuals; AND Uncontrolled asthma = individuals with rhinitis and uncontrolled asthma (further definition can be found in the original study description (32).

Supplement Figure 2. GXB interfaces. The end-users interact with 3 main interactive pages: the Landing Page (A), Dataset Browser (B) and Graphical (C) interfaces. A detailed description of the interface and functionalities have been previously described (PMID 26088622).

SUPPLEMENTAL INFORMATION

Supplement Information 1. Calculation of FC expression.

GXB uses the geometric means of replicates of each sample and calculates FC via the difference (for data in log2 scale) or the ratio (for data in linear scale) of the means. The data displayed in GXB are linear scale FC. The following example illustrates the calculation:

GSE ID: GSE54336

Type of file: *.soft file (The data deposited by the contributor were analyzed with Partek Genomic Suite 6.6 using Affymetrix default analysis settings, quantile normalization and RMA background correction)

Number of samples: 6

Number of groups: 3 (G1TEPP, G1V, and Mock)

Groups compared (for this example)**:** G1TEPP and G1V

Gene symbol: DUSP2

Probe set ID: 204794_at

Sample Information:

GSM1313408 A2EN cells_Chlamydia G1TEPP_4h_rep1

GSM1313409 A2EN cells_Chlamydia G1TEPP_4h_rep2

GSM1313410 A2EN cells_Chlamydia G1V_4h_rep1

GSM1313411 A2EN cells_Chlamydia G1V_4h_rep2

GSM1313412 A2EN cells_Mock infected_repl1

GSM1313413 A2EN cells_Mock infected_repl2

Condition 1: G1TEPP (signal value in log_2 scale)

GSM1313408: 6.31727

GSM1313409: 6.26554

Geometric Mean = sqrt (6.31727 * 6.26554)

→ Geometric Mean = 6.29135183214

Condition 2: G1V (signal value in log_2 scale)

GSM1313410: 6.93182

GSM1313411: 6.88052

Geometric Mean = sqrt (6.93182 * 6.88052)

→ Geometric Mean = 6.90612236689

Calculation of FC expression (G1TEPP/G1V) and transformation to linear scale FC:

log_2 FC = Condition 1 - Condition 2

log_2 FC = 6.29135183214 - 6.90612236689 = -0.61

linear FC = Antilog (-0.61) = $2^{(-0.61)}$ = 0.65

Mathematical transformation when linear FC is less than 1: -1/(FC)

→ -1/0.65 = -1.52 (i.e. down regulation in linear scale)

Reference:

The dataset used in this example can be accessed via these links:

GXB: http://cd2k.gxbsidra.org/dm3/geneBrowser/show/4000099

GEO2R: https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE54336

TABLE 1

A. Search strategy applied for retrieval of relevant datasets from GEO

Description	Search strategy	# of dataset
Dataset relevant to hypersensitivity, hyper-IgE syndrome, and IgE in syndromes, disorders, diseases, and the regulation/phenotype in general of IgE.	("Hypersensitivity" OR "hyper ige syndrome" OR HIES OR hyper-IgE OR IgE OR "immunoglobulin e") AND (syndrome OR syndromes OR disorder OR disorders OR "job's syndrome" OR disease OR diseases OR regulation OR modulation OR change OR changes OR elevated OR decreased OR deficiency OR phenotype) AND "gse"[Filter]	130
Dataset relevant to hyper-IgE and IgE.	("hyper ige syndrome"[All Fields] OR HIES[All Fields] OR hyper-IgE[All Fields] OR IgE[All Fields] OR "immunoglobulin e"[All Fields]) AND "gse"[Filter] = 75	75
Dataset relevant to hyper-IgE and IgE in allergenic/antigenic reactions, syndromes, disorders, diseases, and the regulation/phenotype in general of IgE.	("Hyperimmunoglobulinemia E syndrome" OR "hyper ige syndrome" OR HIES OR hyper-IgE OR IgE OR "immunoglobulin e") AND ("Allergens"[Mesh] OR "Antigens"[Mesh] OR syndrome OR syndromes OR disorder OR disorders OR "job's syndrome" OR disease OR diseases OR regulation OR modulation OR change OR elevated OR decreased OR deficiency OR phenotype) AND ("gse"[Filter])	62
Dataset relevant to IgE receptor, hyper-IgE, IgE and mast cells in hypersensitivity and allergy.	(FceR OR "IgE receptor" OR "hyper ige syndrome"[All Fields] OR HIES[All Fields] OR hyper-IgE[All Fields] OR IgE[All Fields] OR "immunoglobulin e"[All Fields] OR "mast cells"[All Fields] OR "Mast Cells"[Mesh]) AND ("hypersensitivity"[MeSH Terms] OR "allergy and immunology"[MeSH Terms] OR ("hypersensitivity"[MeSH Terms] OR "allergy and immunology"[MeSH Terms] OR ("hypersensitivity"[MeSH Terms] OR "allergy and immunology"[MeSH Terms] OR ("hypersensitivity"[MeSH Terms] OR "allergy and immunology"[MeSH Terms] OR allergy[All Fields])))) AND "gse"[Filter]	48
Dataset relevant to hyper-IgE and IgE in hypersensitivity and allergy.	("hyper ige syndrome"[All Fields] OR HIES[All Fields] OR hyper-IgE[All Fields] OR IgE[All Fields] OR "immunoglobulin e"[All Fields]) AND ("hypersensitivity"[MeSH Terms] OR "allergy and immunology"[MeSH Terms] OR allergy[All Fields]) AND "gse"[Filter]	33
Dataset relevant to IgE in dermatitis, hypersensitivity, allergy, and mast cells.	((("dermatitis"[MeSH Terms] OR dermatitis[All Fields]) OR ("hypersensitivity"[MeSH Terms] OR "allergy and immunology"[MeSH Terms] OR allergy[All Fields]) OR ("mast cells"[MeSH Terms] OR mast cells[All Fields])) AND (IgE[DESC] OR ("immunoglobulin e"[All Fields]))) AND "gse"[Filter]	50
Dataset relevant to hyper-IgE and IgE in immune system diseases and study about allergens.	("hyper ige syndrome"[All Fields] OR HIES[All Fields] OR hyper-IgE[All Fields] OR IgE[All Fields] OR "immunoglobulin e"[All Fields]) AND "gse"[Filter] AND ("Immune System Diseases"[Mesh] OR "Allergens"[Mesh])	37

B. Merged of the result from the search strategy described in A.

Description	Search strategy	# of dataset
Merging the results of all 7 collections mentioned above.	Merge saved collections in "My NCBI"(https://www.ncbi.nlm.nih.gov/myncbi/)	196
---> performed on human samples	Filtering	143
---> performed on mouse samples	Filtering	52

C. Merged of the terms used in the search strategy described in A.

Description	Search strategy	# of dataset
Merging all terms used in the 7 search strategy above.	(FceR[All Fields] OR "IgE receptor"[All Fields] OR "mast cells"[All Fields] OR "Mast Cells"[Mesh] OR "hyper ige syndrome"[All Fields] OR HIES[All Fields] OR hyper-IgE[All Fields] OR IgE[All Fields] OR "immunoglobulin e"[All Fields] OR "Immunoglobulin E"[Mesh]) **AND** ("Immune System Diseases"[Mesh] OR dermatitis[All Fields] OR allergy[All Fields] OR ("Allergens"[Mesh] OR "Antigens"[Mesh]) syndromes[All Fields] OR disease[All Fields] OR "disorder"[All Fields] OR "job's syndrome"[All Fields] OR "regulation"[All Fields] OR modulation[All Fields] OR (change[All Fields] OR changes[All Fields]) OR elevated[All Fields] OR decreased[All Fields] OR deficiency[All Fields] OR phenotype[All Fields]) **AND** "gse"[Filter]	117
---> performed on human samples	Filtering	70
---> performed on mouse samples	Filtering	37

D. Final assembly and curation of search strategy

Description	Search strategy	# of dataset
Merging of "B"and "C" results, only gene expression profiling by microarrays or high-throughput sequencing.	Assembly of the final list of dataset by merging collections B (196) and C (117) in "My NCBI"(https://www.ncbi.nlm.nih.gov/myncbi/)	203
---> performed on human samples	Filtering	115
---> performed on mouse samples	Filtering	76
For human samples only:		
Manual inspection of dataset to confirm implication to IgE-related pathogenesis in human only.	Inspect all human dataset descriptions using the "GEO accession display tool" (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi)	53
Select gene expression by arrays only.	Filter platform type	49
Manual inspection of the type of study and design, the type of samples, and the disease groups.	Read full description and protocol of each dataset available on "GEO accession display tool" (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi) and inspect associated published article when necessary	30

	Title	GEO ID	Platform name	GPL	Disease/ treatment/ therapy	Sample type	Number of sample	Expt design	Pubmed ID*	Trend validation	FC validation	Comments for validation
1	Gene expression patterns in house dust mite stimulated CD4 T cells and IgG to IgE ratios	GSE70760	Affymetrix Human Gene 1.0 ST Array [hugene10st_Hs_ENTREZG_ 19.0.0]	GPL20171	House dust mite	T cells	90	In vitro	26518094	IL4, IL13, IL17RB		Data not available or reported differently in associated publication
2	* NEW Gene expression patterns in PBMC associated with asthma exacerbation attack	GSE19301	Affymetrix Human Genome U133A Array	GPL96	Asthma	PBMC	685	Ex vivo	21779351			Data reported differently
3	Comparison of two sets of microarray experiments to define allergic asthma expression pattern	GSE41649	Affymetrix Human Genome U133A Array	GPL96	Allergic asthma	Bronchial biopsy	8	Ex vivo	19842841		SERPINB2, CX3CR1, C7	Strong validation
4	Distinct epithelial gene expression phenotypes in childhood respiratory allergy - Disease State	GSE19190	Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	GPL6244	Rhinitis, allergy, asthma	Human nasal epithelium cells	38	In vitro	22005912			Validated with stimulation
5	Distinct epithelial gene expression phenotypes in childhood respiratory allergy - Stimulation	GSE19190	Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	GPL6244	Rhinitis, allergy, asthma	Human nasal epithelium cells	21	In vitro	22005912		DHX58, PRIC285, STARD5	Good validaton
6	Effect of intradermal immunotherapy (IDIT) injections on gene expression profiles of activated T cells derived from skin biopsy explants	GSE72324	Illumina HumanHT-12 V4.0	GPL10558	Grass pollen allergy	T cells	15	In vitro	27773851		TNFSF8, TNIP3, HDAC1	Strong validation
7	Expression data for human epithelium from subjects with atopic dermatitis, psoriasis and nonatopic controls	GSE26952	Sentrix HumanRef-8 Expression BeadChip	GPL2700	Atopic dermatitis, psoriasis	Epidermis	16	Ex vivo	21163515		GJA1, TGM1, OCLN	Strong validation
8	Expression data from IBS patients before and after treatment	GSE14842	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	Diarrhea-irritable bowel syndrome	Epidermis	14	Ex vivo	N/A			No associated publication
9	Functional classes of bronchial mucosa genes that are differentially expressed in asthma	GSE15823	Affymetrix Human Genome U95 Version 2 Array	GPL8300	Asthma	Bronchial biopsy	12	Ex vivo	15038835		SFRP1, Alox15, LDB1	Strong validation
10	Gene expression analysis related to olive pollen allergy	GSE37157	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	Olive pollen allergy	PBMC	28	Ex vivo	23830385			No associated publication
11	Gene expression changes in early phase of venom immunotherapy	GSE92866	Affymetrix Human Gene Expression Array [Brainarray ENTREZG Version 20]	GPL22841	Venom allergy	Blood	59	Ex vivo	N/A		GATA3, FoxP3	No associated publication, but some articles mentioned up-regulated genes that matches with GXB data
12	Gene expression pattern of alveolar macrophages of allergic asthmatics in comparison with control subjects	GSE22528	Affymetrix Human Genome U133A Array	GPL96	Asthma	Alveolar macrophage	10	Ex vivo	19913588		CCR1, HSPD1, SERPINH1	Strong validation
13	Gene expression profile of patients with moderate and severe chronic spontaneous urticaria	GSE72542	Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 (Feature Number version)	GPL16699	Chronic spontaneous urticaria	Skin/whole blood	61	Ex vivo	28407332	FOSB, S100A9, ADAMTS4		Data not available or reported differently in associated publication
14	Gene expression profiling of patients with allergy to latex and/or vegetable food	GSE13619	Affymetrix Human HG-Focus Target Array	GPL201	Latex/plant-derived food allergy	PBMC	21	Ex vivo	N/A	IFNG, STAT4, IL10RA		No associated publication, but some articles mentioned up-regulated genes that matches with GXB data
15	Genome-wide expression analysis demonstrates a dominant role of TLR4 for activation of human phagocytes by the alarmin MRP8	GSE56681	Affymetrix Human Genome U133 Plus 2.0 Array	GPL96 GPL570	Alarmins myeloid-related protein 8 and 14 signalling	Monocytes	19	In vitro	25505274		TRIP10, NFKB1, RNASE6	Good validaton
16	Human basophil expression profiles - Atopic vs Non-atopic	GSE64639	Illumina HumanHT-12 V4.0	GPL10558	Healthy subject	Pheripheral blood	16	In vitro	25962139			Data not available or reported differently in associated publication

17	Identification of IL-21-induced STAT3 dependent genes in human B cells	GSE51587	Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	GPL6244	Job's syndrome	B cells	14	In vitro	24159173		PRDM1, IL2RA, SGK1	Gene expression values not available in associated publication
18	Illumina Bead expression array data from Human IgE+ and IgG+ B cell subsets	GSE99948	Illumina HumanHT-12 V4.0	GPL10558	Tonsillectomy/ Heathly subject	B cells	24	In vitro	N/A	CD99, SDC1		No associated publication, but some articles mentioned up-regulated genes that matches with GXB data
19	Influence of olive pollen stimuli on the gene-expression profile in healthy controls and allergic patients	GSE54522	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	Olive pollen allergy	PBMC	46	In vitro	25553522			Data not available or reported differently in associated publication
20	Integrated genomic and prospective clinical studies show the importance of modular pleiotropy for disease susceptibility, diagnosis and treatment (dataset 1)	GSE44956	Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (Probe Name Version)	GPL14550	Seasonal allergic rhinitis	T cells	48	In vitro	24571673			Data not available or reported differently in associated publication
21	Leukotriene E4 is a full functional agonist for human cysteinyl leukotriene type 1 receptor - Cell Comparison	GSE75603	Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	GPL6244	Leukotriene E4 response	Mast cells	6	In vitro	26830450			Validated with stimulation
22	Leukotriene E4 is a full functional agonist for human cysteinyl leukotriene type 1 receptor - Stimulation	GSE75603	Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	GPL6244	Leukotriene E4 response	Mast cells	9	In vitro	26830450		STX3, GGR3, NFKBID	Good validaton
23	Neutrophil and PBMC gene expression data from Job's Syndrome - Neutrophils	GSE8507	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	Job's syndrome	Neutrophils	90	In vitro	17881745			Validated with PBMC
24	Neutrophil and PBMC gene expression data from Job's Syndrome - PBMC	GSE8507	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	Job's syndrome	PBMC	51	In vitro	17881745		CD151, CLEC12A, CoL6A2	Good validaton
25	Novel mediators of eicosanoid and epithelial nitric oxide production in asthma	GSE13785	Affymetrix Human Genome U133 Plus 2.0 Array	GPL96 + GPL570	Asthma	Sputum	22	Ex vivo	20052409		TGM2, THBS1, CD24	Good validaton
26	Progressive activation of Th2/Th22 characterizes acute and chronic atopic dermatitis	GSE36842	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	Atopic dermatitis	Skin	39	Ex vivo	22951056		PI3, S100A8, S100A7	Strong validation
27	The effect of a dexamethasone and a FK506 on the induction of chemokines in human mast cells	GSE15174	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	Allergy	Mast cells	5	In vitro	19454720		CCL1, CCL2, CCL4	Good validaton
28	Transcriptional profiling of egg allergy and relationship to disease phenotype	GSE88796	Illumina HumanHT-12 V4.0	GPL10558	Egg allergy	PBMC	132	In vitro	27788149		CEACAM1, CLC, CCL17	Good validaton
29	Transcriptome profiling of 1) nasal polyp derived human IL17RB positive and negative T-helper cells and 2) T-helper cells from normal nasal mucosa and matched peripheral blood	GSE70898 GSE70900	Illumina HumanHT-12 V4.0	GPL10558	Chronic rhinosinusitis with nasal polyposis	T cells	24	In vitro	26684290		PCID2, VPS13C, GPR87	Strong validation
30	Upper airway gene expression is an effective surrogate biomarker for Th2-driven inflammation in the lower airway	GSE41861	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	Allergic asthma	Human nasal epithelium/ Bronchial biopsies	138	Ex vivo	N/A			No associated publication

* N/A = Not available.

FIGURE 1

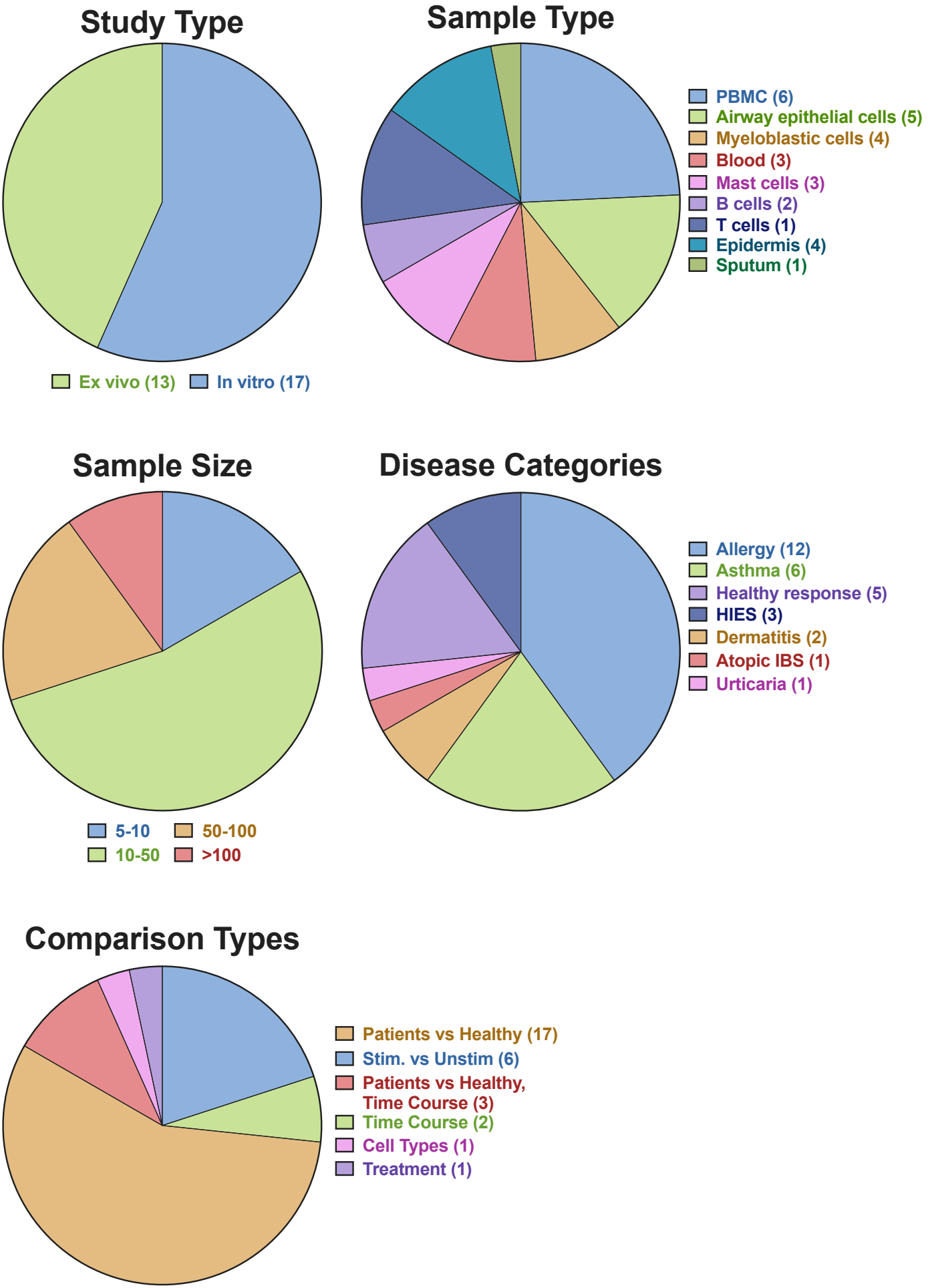
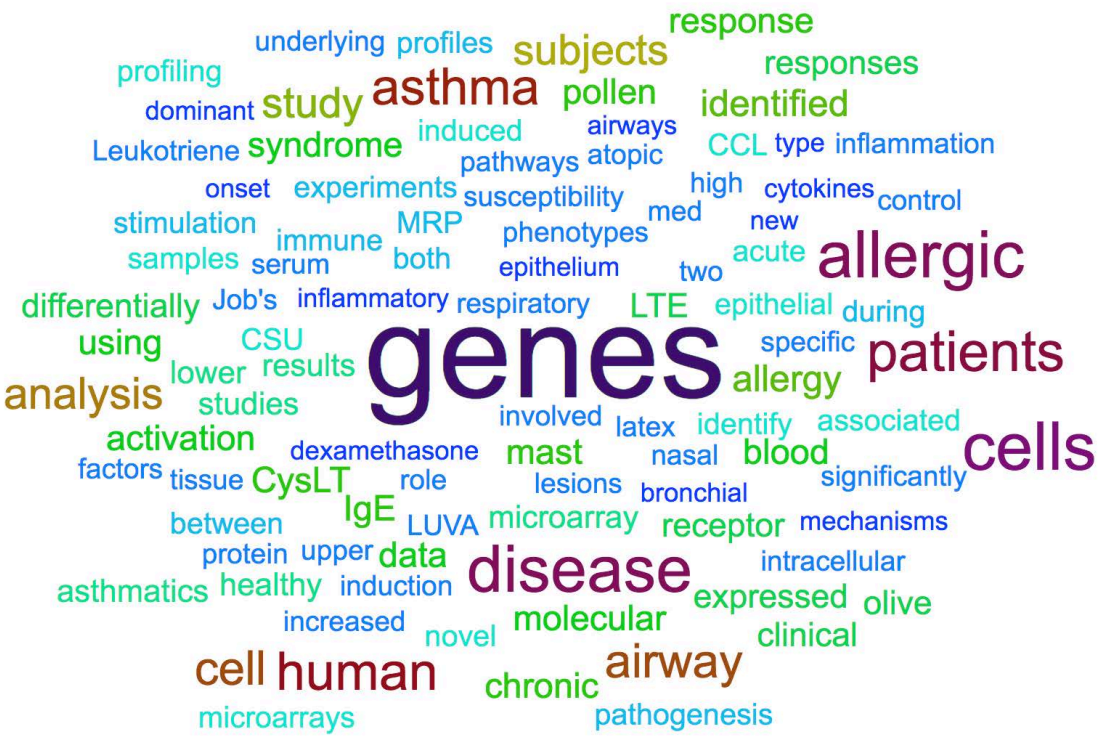

Study Type
Ex vivo (13)
In vitro (17)
Sample Type
PBMC (6)
Airway epithelial cells (5)
Myeloblastic cells (4)
Blood (3)
Mast cells (3)
B cells (2)
T cells (1)
Epidermis (4)
Sputum (1)
Sample Size
5-10
50-100
10-50
>100
Disease Categories
Allergy (12)
Asthma (6)
Healthy response (5)
HIES (3)
Dermatitis (2)
Atopic IBS (1)
Urticaria (1)
Comparison Types
Patients vs Healthy (17)
Stim. vs Unstim (6)
Patients vs Healthy, Time Course (3)
Time Course (2)
Cell Types (1)
Treatment (1)

FIGURE 2

A


response
underlying
profiles
subjects
responses
profiling
asthma
pollen
identified
dominant
study
induced
airways
Leukotriene
syndrome
pathways
atopic
CCL
type
inflammation
onset
experiments
susceptibility
high
cytokines
control
stimulation
MRP
med
new
immune
phenotypes
samples
serum
both
epithelium
two
acute
allergic
differentially
Job's
inflammatory
respiratory
LTE
epithelial
during
using
CSU
genes
specific
patients
lower
results
allergy
analysis
studies
involved
latex
identify
associated
activation
dexamethasone
mast
nasal
blood
cells
factors
tissue
CysLT
role
lesions
bronchial
significantly
IgE
between
LUVA
microarray
receptor
mechanisms
protein
upper
data
disease
intracellular
asthmatics
healthy
induction
expressed
olive
increased
molecular
novel
clinical
cell
human
airway
chronic
microarrays
pathogenesis

B

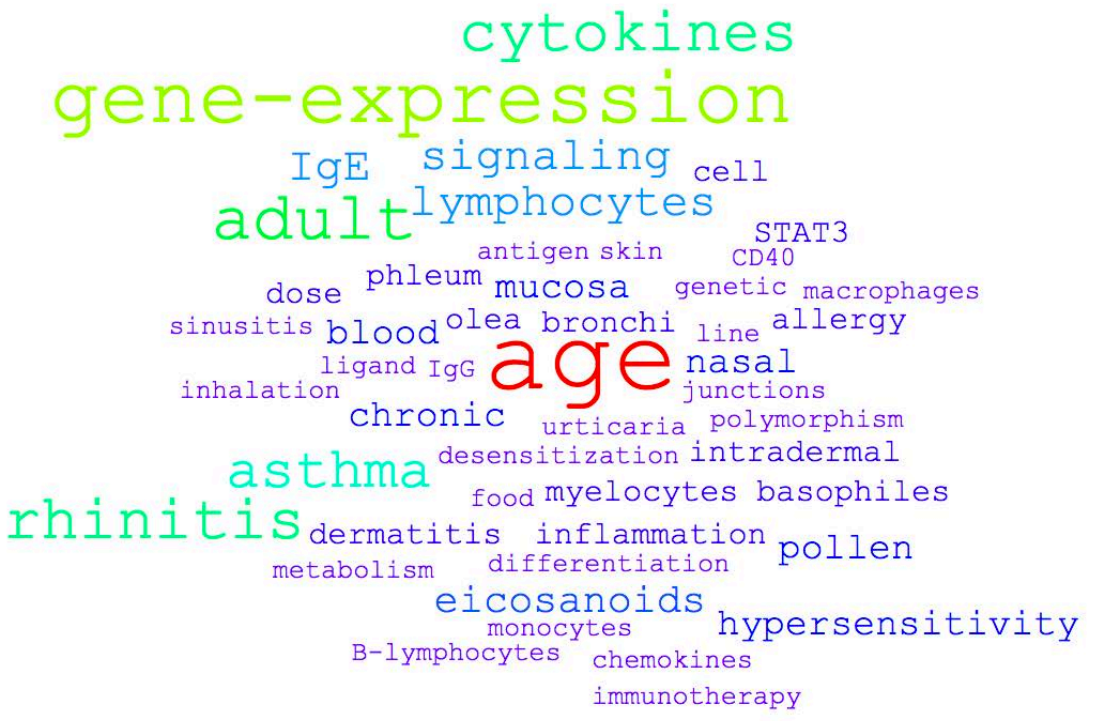
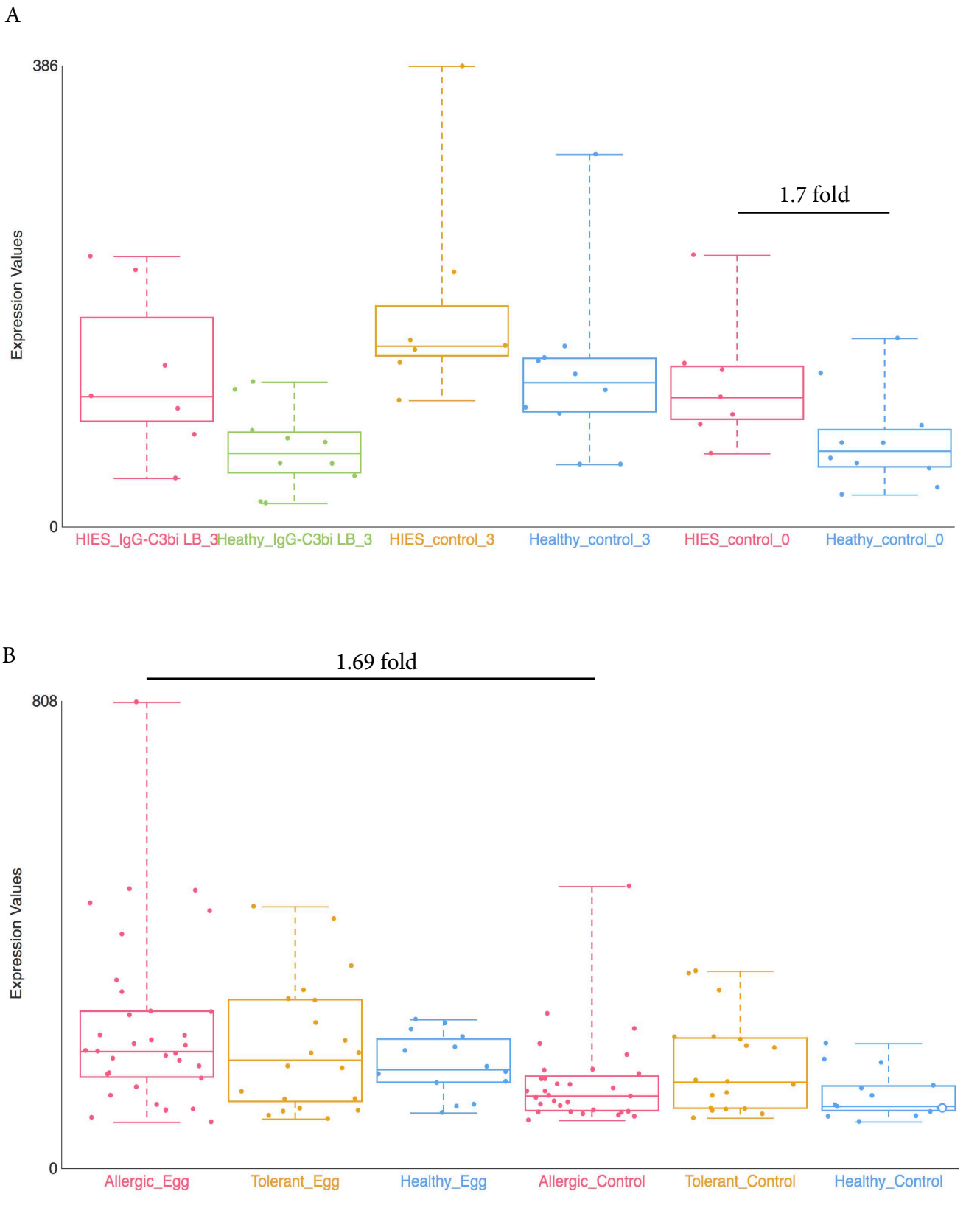

cytokines
gene-expression
IgE
signaling
cell
adult
lymphocytes
STAT3
antigen
skin
CD40
phleum
mucosa
genetic
macrophages
dose
sinusitis
blood
olea
bronchi
line
allergy
ligand
IgG
age
nasal
inhalation
junctions
chronic
urticaria
polymorphism
desensitization
intradermal
asthma
food
myelocytes
basophiles
rhinitis
dermatitis
inflammation
pollen
metabolism
differentiation
eicosanoids
monocytes
hypersensitivity
B-lymphocytes
chemokines
immunotherapy

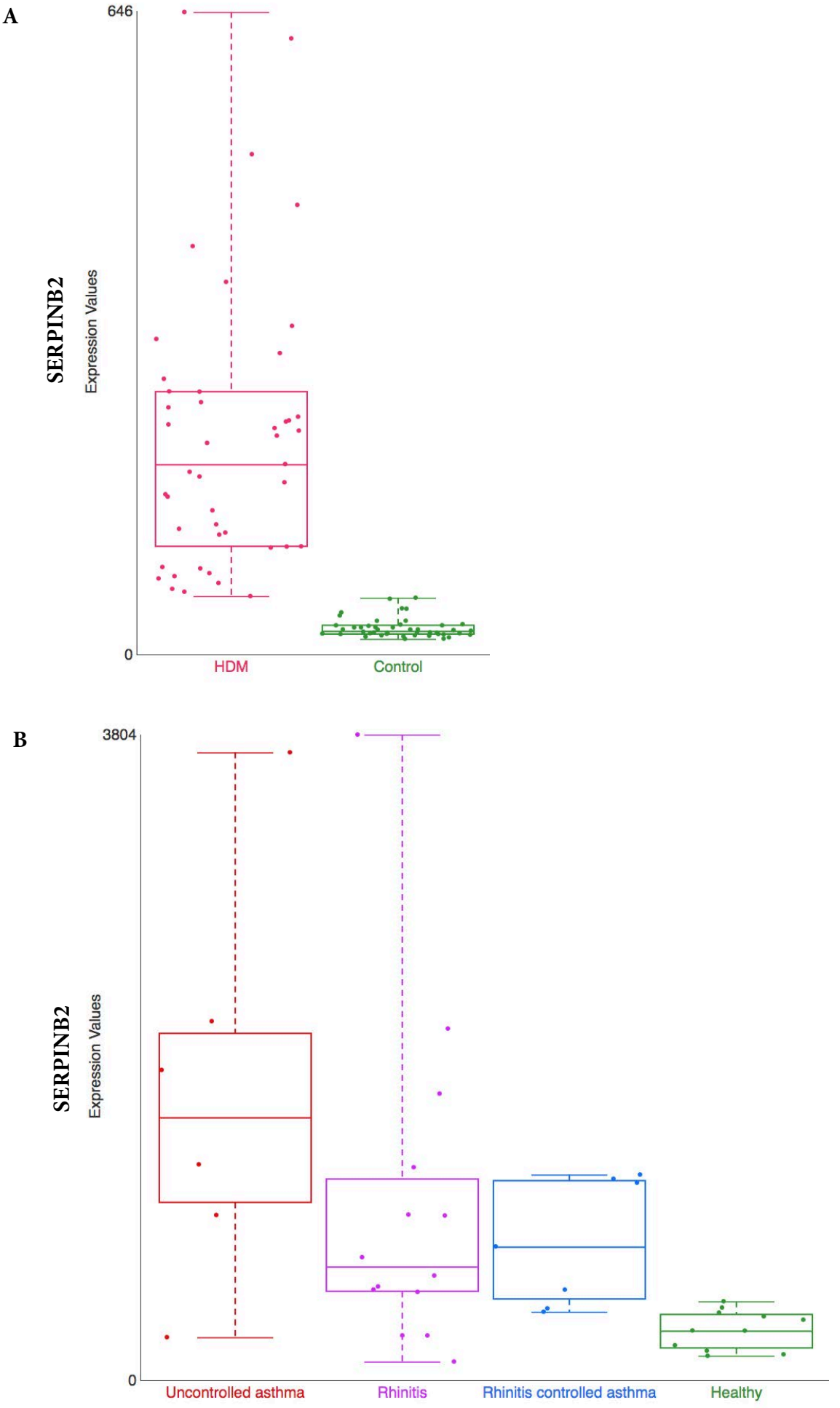
FIGURE 3


A
386
0
Expression Values
1.7 fold
HIES_IgG-C3bi LB_3
Heathy_IgG-C3bi LB_3
HIES_control_3
Healthy_control_3
HIES_control_0
Heathy_control_0
B
1.69 fold
808
0
Expression Values
Allergic_Egg
Tolerant_Egg
Healthy_Egg
Allergic_Control
Tolerant_Control
Healthy_Control

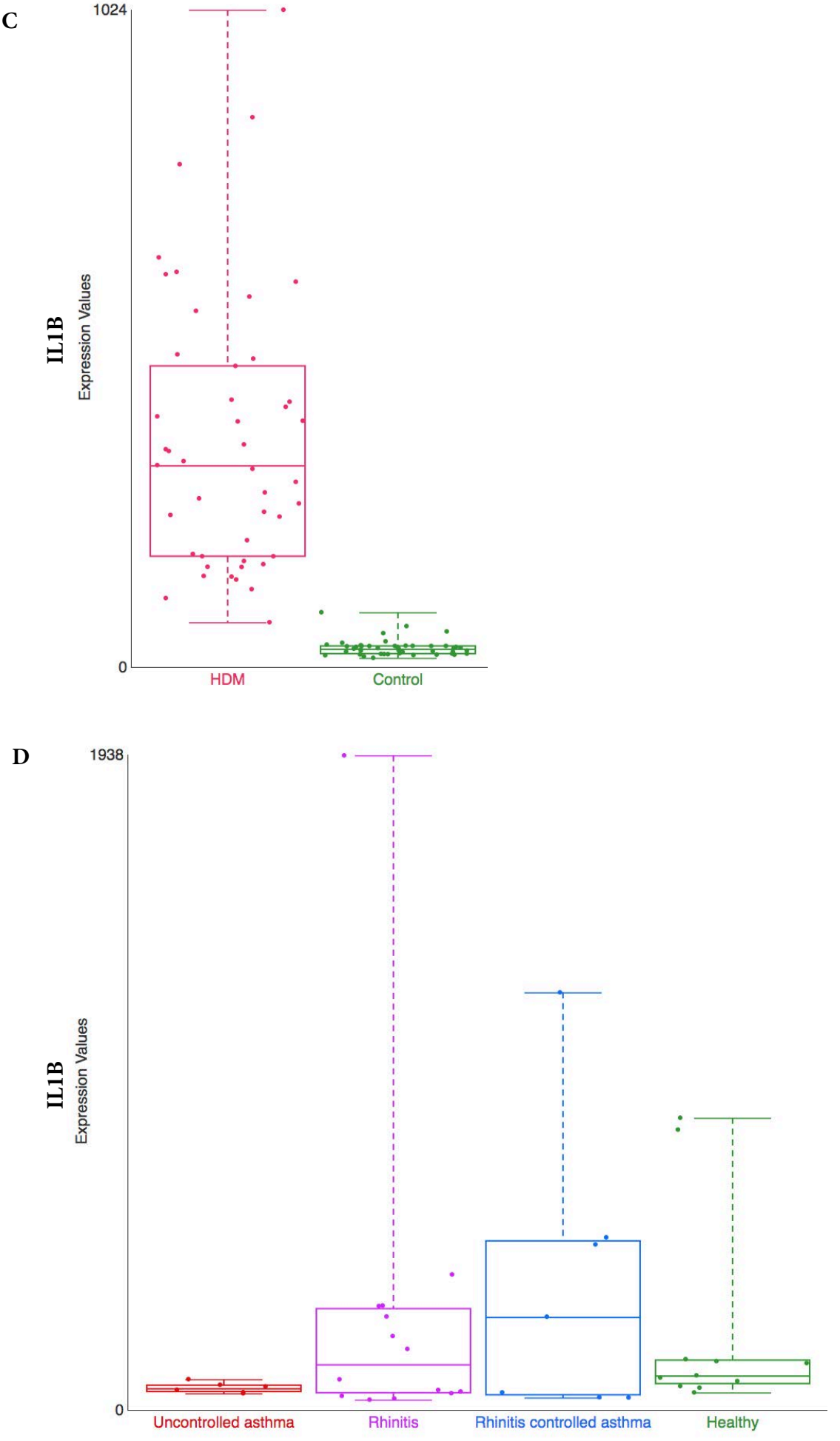
Symbol	Name	Fold-Change
IL1B	interleukin 1, beta	FC: 8.55
SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	FC: 5.57
CXCL5	chemokine (C-X-C motif) ligand 5	FC: 4.89
CCL2	chemokine (C-C motif) ligand 2	FC: 4.79
DACT1	dapper, antagonist of beta-catenin, homolog 1 (Xenopus laevis)	FC: 4.65
IL2RA	interleukin 2 receptor, alpha	FC: 4.26
CISH	cytokine inducible SH2-containing protein	FC: 3.98
IL4R	interleukin 4 receptor	FC: 3.96
GPNMB	glycoprotein (transmembrane) nmb	FC: 3.65
PLBD1	phospholipase B domain containing 1	FC: 3.55
FGL2	fibrinogen-like 2	FC: 3.51
ME1	malic enzyme 1, NADP(+)-dependent, cytosolic	FC: 3.42
TGFBI	transforming growth factor, beta-induced, 68kDa	FC: 3.37
IFITM3	interferon induced transmembrane protein 3	FC: 3.27
RNASE1	ribonuclease, RNase A family, 1 (pancreatic)	FC: 3.15
LYZ	lysozyme	FC: 3.02
CD36	CD36 molecule (thrombospondin receptor)	FC: 3
CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	FC: 2.99
FPR3	formyl peptide receptor 3	FC: 2.93
CCL22	chemokine (C-C motif) ligand 22	FC: 2.93
RAB27B	RAB27B, member RAS oncogene family	FC: 2.87
CLEC4E	C-type lectin domain family 4, member E	FC: 2.81
RNASE6	ribonuclease, RNase A family, k6	FC: 2.76
RAB19	RAB19, member RAS oncogene family	FC: 2.71
PTGER2	prostaglandin E receptor 2 (subtype EP2), 53kDa	FC: 2.67
PPBP	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)	FC: 2.62
NDFIP2	Nedd4 family interacting protein 2	FC: 2.55
TFPI2	tissue factor pathway inhibitor 2	FC: 2.52
C3	complement component 3	FC: 2.52
PLA2G7	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	FC: 2.51
LRRK2	leucine-rich repeat kinase 2	FC: 2.48
SLC37A3	solute carrier family 37 (glycerol-3-phosphate transporter), member 3	FC: 2.48
PDK4	pyruvate dehydrogenase kinase, isozyme 4	FC: 2.46
CXCL10	chemokine (C-X-C motif) ligand 10	FC: 2.44
CAMK2D	calcium/calmodulin-dependent protein kinase II delta	FC: 2.39
SOCS3	suppressor of cytokine signaling 3	FC: 2.35
IL13RA1	interleukin 13 receptor, alpha 1	FC: 2.32
MAL	mal, T-cell differentiation protein	FC: 2.31
MT1G	metallothionein 1G	FC: 2.3
GZMB	granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1)	FC: 2.29
RCBTB2	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 2	FC: 2.26
TAF4B	TAF4b RNA polymerase II, TATA box binding protein (TBP)-associated factor, 105kDa	FC: 2.21
LRFN2	leucine rich repeat and fibronectin type III domain containing 2	FC: 2.21
CD180	CD180 molecule	FC: 2.2
TLR4	toll-like receptor 4	FC: 2.19
ACP5	acid phosphatase 5, tartrate resistant	FC: 2.18
ADAM28	ADAM metallopeptidase domain 28	FC: 2.16
CD9	CD9 molecule	FC: 2.16
LGALS2	lectin, galactoside-binding, soluble, 2	FC: 2.11
SLC11A1	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 1	FC: 2.09

Symbol	Name	Fold-Change
CST1	cystatin SN	FC: 17.45
POSTN	periostin, osteoblast specific factor	FC: 9.51
ITLN1	intelectin 1 (galactofuranose binding)	FC: 7.65
CLCA1	chloride channel accessory 1	FC: 6.33
FETUB	fetuin B	FC: 4.19
GPR128	G protein-coupled receptor 128	FC: 3.96
SLC5A5	solute carrier family 5 (sodium iodide symporter), member 5	FC: 3.67
CST2	cystatin SA	FC: 3.5
SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	FC: 3.34
DPP4	dipeptidyl-peptidase 4	FC: 3.19
CA2	carbonic anhydrase II	FC: 2.88
CLC	Charcot-Leyden crystal protein	FC: 2.87
CPA3	carboxypeptidase A3 (mast cell)	FC: 2.83
CDH26	cadherin 26	FC: 2.82
SH2D1B	SH2 domain containing 1B	FC: 2.73
ANO1	anoctamin 1, calcium activated chloride channel	FC: 2.72
EMP1	epithelial membrane protein 1	FC: 2.68
CD200R1	CD200 receptor 1	FC: 2.52
CPA4	carboxypeptidase A4	FC: 2.5
GCNT4	glucosaminyl (N-acetyl) transferase 4, core 2	FC: 2.49
NTS	neurotensin	FC: 2.46
KRT13	keratin 13	FC: 2.44
CCL26	chemokine (C-C motif) ligand 26	FC: 2.38
FAM177B	family with sequence similarity 177, member B	FC: 2.3
IGHM	immunoglobulin heavy constant mu	FC: 2.29
SPINK5	serine peptidase inhibitor, Kazal type 5	FC: 2.28
IGKC	immunoglobulin kappa constant	FC: 2.26
CD274	CD274 molecule	FC: 2.25
GCNT3	glucosaminyl (N-acetyl) transferase 3, mucin type	FC: 2.19
TFF3	trefoil factor 3 (intestinal)	FC: 2.18
CR2	complement component (3d/Epstein Barr virus) receptor 2	FC: 2.17
CD69	CD69 molecule	FC: 2.14
NTRK2	neurotrophic tyrosine kinase, receptor, type 2	FC: 2.14
FGFBP1	fibroblast growth factor binding protein 1	FC: 2.13
TNIP3	TNFAIP3 interacting protein 3	FC: 2.11
KRT6A	keratin 6A	FC: 2.1
KRT24	keratin 24	FC: 2.08
SPRR1B	small proline-rich protein 1B	FC: 2.07
PTHLH	parathyroid hormone-like hormone	FC: 2.06
SERPINB4	serpin peptidase inhibitor, clade B (ovalbumin), member 4	FC: 2.06
HGD	homogentisate 1,2-dioxygenase	FC: 2.06
HGD	homogentisate 1,2-dioxygenase	FC: 2.04
CYP24A1	cytochrome P450, family 24, subfamily A, polypeptide 1	FC: 2.04
S100A2	S100 calcium binding protein A2	FC: 2
SERPINB13	serpin peptidase inhibitor, clade B (ovalbumin), member 13	FC: 2
MS4A1	membrane-spanning 4-domains, subfamily A, member 1	FC: 1.96
LGALS7	lectin, galactoside-binding, soluble, 7	FC: 1.96
FN1	fibronectin 1	FC: 1.94
SERPINB10	serpin peptidase inhibitor, clade B (ovalbumin), member 10	FC: 1.93
DSG3	desmoglein 3	FC: 1.91

Suppl. Figure 1

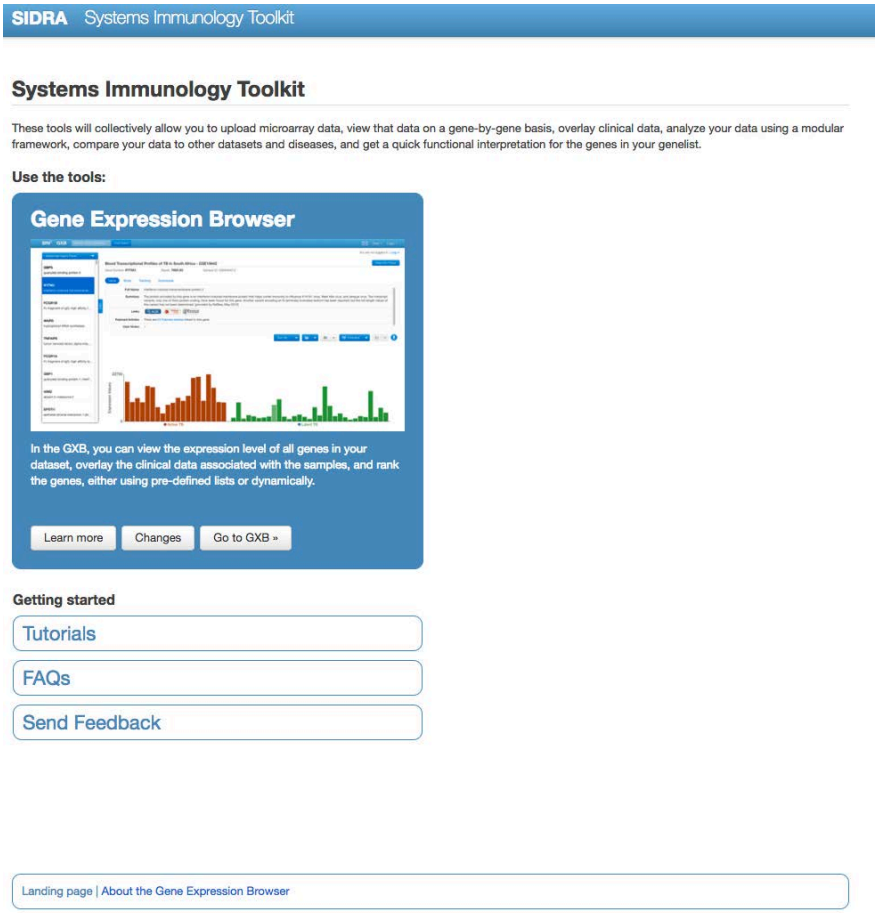


Suppl. Figure 1

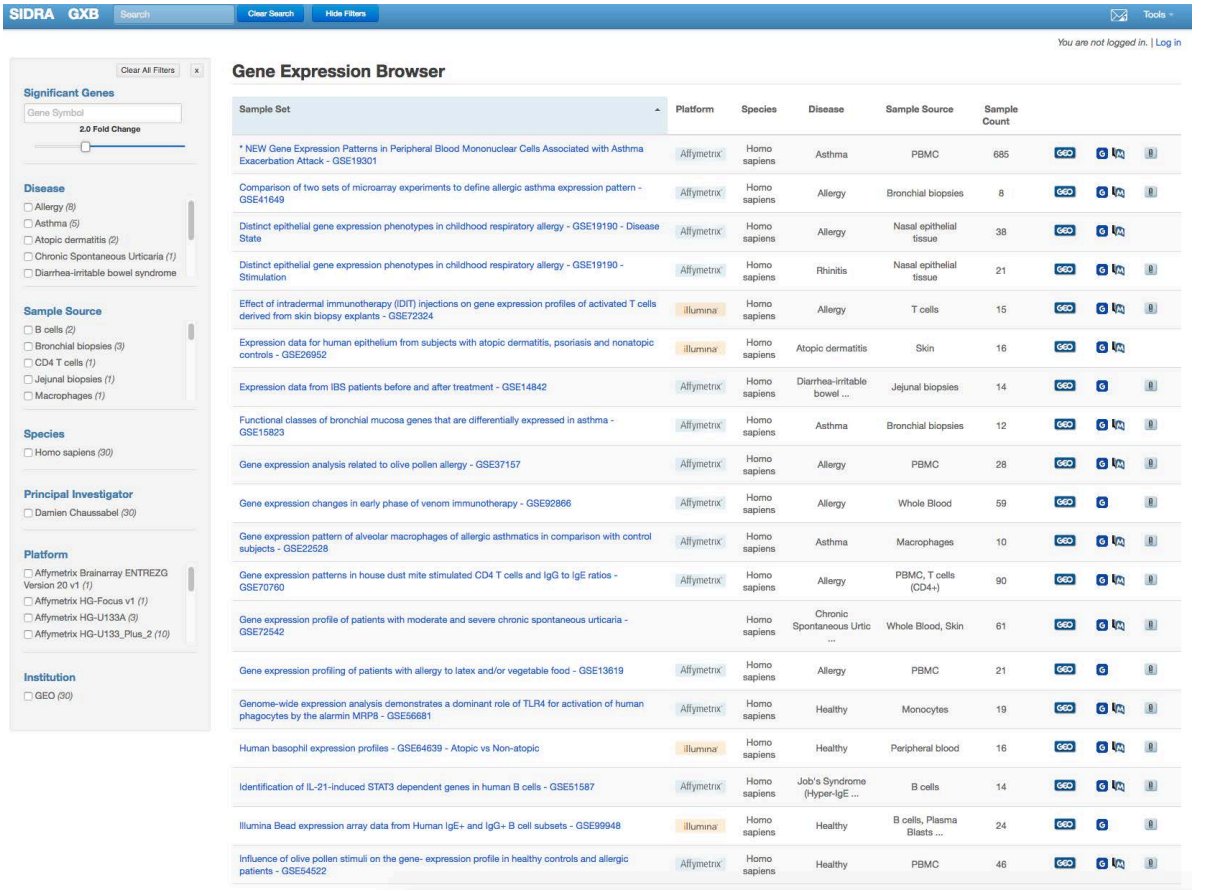


SUPPL. FIGURE 2

A



B



SUPPL. FIGURE 2

C