

SCIENTIFIC REPORTS



OPEN

Revealing the acute asthma ignorome: characterization and validation of uninvestigated gene networks

Received: 07 December 2015

Accepted: 01 April 2016

Published: 21 April 2016

Michela Riba^{1,*}, Jose Manuel Garcia Manteiga^{1,*}, Berislav Bošnjak^{2,*}, Davide Cittaro¹, Pavol Mikolka^{2,†}, Connie Le^{2,‡}, Michelle M. Epstein^{2,#} & Elia Stupka^{1,#}

Systems biology provides opportunities to fully understand the genes and pathways in disease pathogenesis. We used literature knowledge and unbiased multiple data meta-analysis paradigms to analyze microarray datasets across different mouse strains and acute allergic asthma models. Our combined gene-driven and pathway-driven strategies generated a stringent signature list totaling 933 genes with 41% (440) asthma-annotated genes and 59% (493) ignorome genes, not previously associated with asthma. Within the list, we identified inflammation, circadian rhythm, lung-specific insult response, stem cell proliferation domains, hubs, peripheral genes, and super-connectors that link the biological domains (*Il6*, *Il1 β* , *Cd4*, *Cd44*, *Stat1*, *Traf6*, *Rela*, *Cadm1*, *Nr3c1*, *Prkcd*, *Vwf*, *ErbB2*). In conclusion, this novel bioinformatics approach will be a powerful strategy for clinical and across species data analysis that allows for the validation of experimental models and might lead to the discovery of novel mechanistic insights in asthma.

Allergen exposure causes a complex interaction of cellular and molecular networks leading to allergic asthma in susceptible individuals. Experimental mouse models of allergic asthma are widely used to understand disease pathogenesis and elucidate mechanisms underlying the initiation of allergic asthma¹. For example, gene profiling of lung tissue from experimental mice during the initiation of allergic asthma in experiments with different protocols^{2–7} validated well-known genes and identified new genes with roles in disease pathogenesis such as *C5*³, *Arg1*⁸, *Adam8*⁹, and *Pon1*⁷, and dissected pathways activated by *Il13*¹⁰ and *Stat6*¹¹. Despite these datasets revealing large numbers of genes expressed in lungs during the onset of allergic asthma, there is a skewed distribution of research and literature coverage on a small number of genes, as has been noted in other fields^{12–14}.

Previous meta-analysis of asthma datasets generated either very small¹⁵ or very large gene lists¹⁶ using approaches based on the analysis of gene lists. There is a need for new approaches that generate testable gene lists. To fully understand the genes and pathways in the lung that are involved in the initiation of allergic asthma, we reviewed all 23 published gene expression experiments performed on lungs from mice with allergic asthma to select those of sufficient robustness to be utilized for a multi-dataset bioinformatics analysis. We focused on 6 studies that share a common microarray platform, have a robust disease phenotype and were derived from experiments using different mouse strains and experimental protocols^{2–7}. We used a novel bioinformatics strategy combining gene-driven and pathway-driven approaches (see methods) to generate a stringent asthma signature list. This list included a large set of protein-coding genes that had no published function or literature connecting them to allergic asthma either at the level of the whole organism or for individual organs, tissues or cells, genes for which the term “ignorome” was recently coined¹⁷. Here, we present a gene signature list from mice consisting of 933 genes with 440 asthma-annotated and 493 “acute allergic asthma ignorome” genes. This novel bioinformatics

¹Center for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, Milan, Italy. ²Experimental Allergy Lab, Department of Dermatology, DIAID, Medical University of Vienna, Austria. *These authors contributed equally to this work. #These authors jointly supervised this work. †Present address: Comenius University in Bratislava, Jessenius Faculty of Medicine, Department of Physiology and Biomedical Centre, Martin, Slovakia. ‡Present address: The University of Alberta, Faculty of Medicine, Edmonton, Alberta, Canada. Correspondence and requests for materials should be addressed to B.B. (email: berislav.bošnjak@meduniwin.ac.at)

No.	Study				Affimerix Mouse Array	Mice	Antigen (Mice number)	No. of challenges	Sampling time (hr after last challenge)
	GSE	GPL	PMID	Comparison					
1	13032	1261	19491150	A	Genome 430 2.0	A/J	PBS (3)	3	3
				B			OVA (3)		
2	6858	1261	17437023	C	Genome 430 2.0	BALB/c	PBS (4)	7	24
				D			OVA (4)		
3	9465	1261	19057703	E	Genome 430 2.0	A/J	PBS (3)	1	96
				F			OVA (3)		
4	1301	339–340	–	G	U74 Version 2	BALB/c	PBS (3)	2	72
				H			HDM (3)		
5	3184	339	–	I	U74 Version 2	C3H	PBS (5)	1	6
				J			OVA (5)		
				K		C3H	PBS (5)	1	24
				L			OVA (5)		
6	18010	1261	19770271	M	Genome 430 2.0	A/J	PBS (5)	1	6
				N			OVA (5)		
6	18010	1261	19770271	O	Genome 430 2.0	B6.1l4raQ576/Q576	-(5)	7	24
				P		B6.1l4raQ576/Q576/IL-13tg	IL-13 (8)		

Table 1. Gene expression studies of asthma used for the bioinformatics analysis. Six GEO Datasets (<http://www.ncbi.nlm.nih.gov/gds/>) studies were selected for meta-analysis on the basis of the microarray platform and *in vivo* experimental protocol. From those 6 studies, 10 asthmatic vs. healthy control comparisons (A to J) have been extracted and further analyzed.

approach may also be powerful when used on large datasets from clinical studies, allowing for comparisons between mouse and human datasets thus, leading to the discovery of disease-specific genes and pathways.

Results

Acute allergic asthma gene signature list. To select the datasets for analysis, we analysed 23 microarray studies of lung RNA from mice at the initiation of allergic asthma [Ref. 18 and GEO database (<http://www.ncbi.nlm.nih.gov/geo/>)]. Our selection criteria included the phenotype elicited and data robustness (see methods). Six microarray datasets [Refs 2,10,19,20 and GSE1301 and GSE3184] from 5 different mouse strains, 2 allergens, 6 experimental protocols were subdivided into 10 comparisons of control vs. asthmatic mice (A–J) (Table 1 and Fig. 1).

To create an acute asthma signature list from these 6 microarray datasets, we used a multi-dataset bioinformatics analysis strategy including a pathway-driven and gene-driven approach (see methods and supplementary methods; Fig. 1) to ensure that both enriched pathways as well as robust genes not annotated to pathways were maintained in the final signature (see methods). The pathway-enrichment-driven approach yielded 493 genes, the gene-driven approach 602 genes, giving a final acute asthma signature list of 933 genes.

We validated the robustness of the signature by measuring 59 genes through qPCR in an independent experimental model of allergic asthma (Supplementary Fig. S1). Despite the difference in experimental setup and technology platform, we observed a significant positive correlation of 0.8660 and 0.8819 (both $p < 0.00001$) between the two datasets at 24 h and 72 h post-challenge, respectively (Fig. 2, Supplementary Fig. S3). These data indicate that stringent and multi-dataset bioinformatics analysis across independent experiments can yield robust gene signatures, which are found to be applicable regardless of experimental context and are representative of general features of the disease.

Allergic asthma ignorome. To determine the already known association of the 933 genes with allergic asthma, we enumerated publications associated with asthma, as well as asthma annotation in other sources (CTD and Malacards, see supplementary methods). Figure 3A illustrates the number of genes associated with asthma using these methods. There is an inverse correlation between these data sources and the number of genes, i.e. a large number of publications are associated with relatively few genes: 13 genes (*Ccl11*, *Ccl5*, *Cd14*, *Cysltr1*, *Gstm1*, *Ifng*, *Il10*, *Il13*, *Il33*, *Il4*, *Il4ra*, *Tlr2*, and *Tlr4*) account for approximately 1/3 of the articles published on asthma for these 933 genes (Fig. 3). In our dataset, there are 493 genes (Fig. 3B) that are as yet unrelated to asthma and belong to the acute allergic asthma “ignorome”. These data, combined with the robustness of the gene signature shown above, substantiate that the scientific community has so far focused only on a small subset of genes which are likely to play a role in asthma and that there is a substantial set of genes which warrant further investigation.

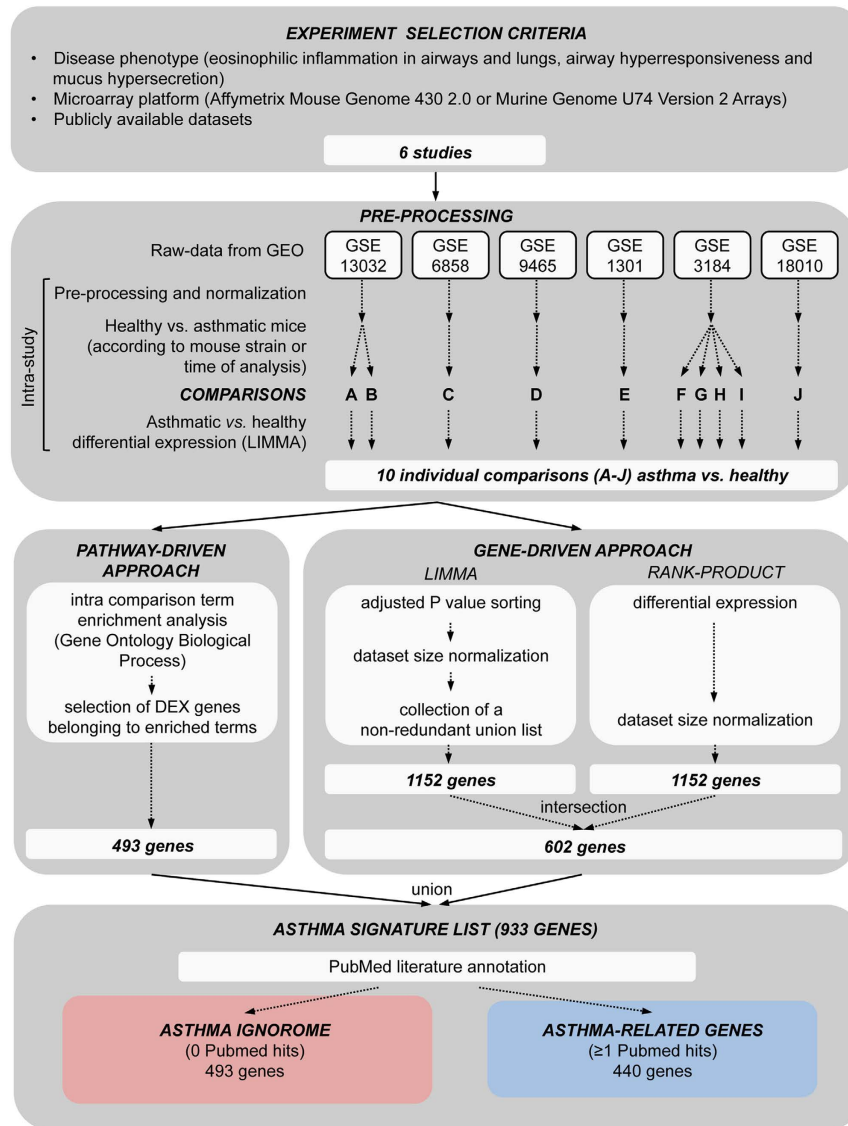


Figure 1. Schematic overview of analysis pipeline. Using experiment selection criteria, we selected six studies that were further subdivided into 10 comparisons of control and asthmatic mice according to mouse strain and time of analysis. After initial analysis with linear models to obtain differential gene expression, data from each comparison were re-analyzed using pathway- and gene-driven approaches (for details please refer to supplementary methods). Lists of differentially regulated genes generated in the 2 approaches were merged into a final asthma signature list of 933 genes, which was used for literature-coverage searches in PubMed.

Network connectivity analysis. To define the allergic asthma gene network, we analysed the connections of 897 human orthologs from 933 acute asthma mouse genes (see methods). We detected a network of 779 densely connected genes in which there are 7 topological clusters containing 763 genes (Fig. 4A), whereas 118 genes were outside of the network (Supplementary Table S1). In the 7 topological clusters, 409 genes are associated with asthma publications (Fig. 4B) and 645 are related to “inflammation or immunity” (Fig. 4C) in at least one relevant publication.

The 7 topological clusters were annotated for functional enrichments (see methods) (Table 2, Supplementary Fig. S4). The clusters fall into 4 major biological domains, of which two are well studied, and two novel: 1) inflammation, including clusters 1A,B: cytokine–cytokine receptor signaling, leucocyte endothelial migration (353 genes, 66.86% published in asthma) 2) circadian rhythm, cluster 2, including genes expressed in the lung as well as CD4⁺, CD8⁺, CD56⁺ and CD71⁺ cells (15 genes, 86.7% as yet unpublished in asthma), 3) lung-specific insult response including clusters 3A–C: extracellular matrix remodelling, adherens and tight junctions, and mucus secretion (309 genes, of which 48.5% are published in connection to asthma), and 4) stem cell proliferation in CD34⁺, CD71⁺, CD105⁺ cells (82 genes, of which 85.4% unpublished in asthma) (Table 2, Fig. 4A).

Investigating, in greater depth, the connectivity between the genes contained in these domains (see supplementary methods) enabled the characterization of the connectivity hierarchy amongst them (Fig. 5). As expected,

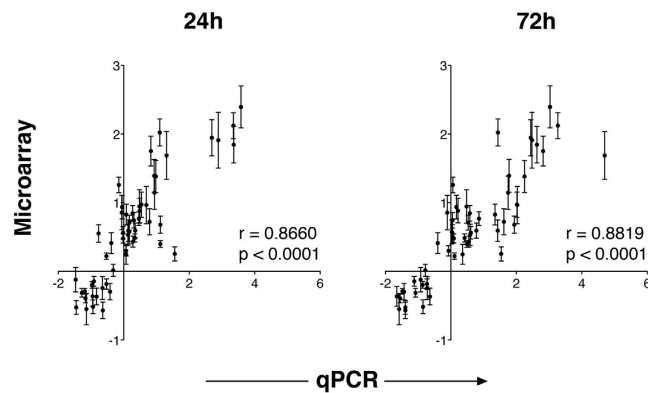


Figure 2. Microarray data for 59 randomly selected genes from 10 comparisons of control and asthmatic mice correlate to quantitative PCR data from our independent mouse asthma model. For qPCR, OVA-sensitized BALB/c mice received PBS (controls) or OVA challenge 24 h or 72 h before extraction of total lung RNA. Data are shown as mean \pm SEM. Pearson r coefficients and p values for each correlation are indicated.

clusters belonging to the inflammation domain have the strongest connectivity between themselves, as well as lung to inflammation domains. Interestingly, the genes within the 2 novel domains identified (stem cell proliferation and circadian rhythm) are strongly connected to a specific cluster within the lung-specific insult response, *i.e.* extracellular matrix remodelling. The same domains are also interconnected by a single gene, *Rora* that belongs to the circadian rhythm domain and couples the stem cell proliferation and the extracellular matrix remodelling cluster in the lung-specific insult response domain (Fig. 5).

When these novel clusters are annotated for publications in inflammation or immunity, they reveal a substantially larger body of literature (73.3% of circadian rhythm genes and 56.1% of the stem cell genes are associated), indicating that their role in the broader field of immunology is already being investigated, although not yet associated specifically to asthma.

The reliability of the identified functional gene clusters is corroborated by manual inspection of well established key genes in asthma, *e.g.*, the inflammation cluster contains lung allergic immune response genes (*e.g.*, cluster 1A: *Il4*, *Il4r*, *Il13*, *Il33*, *Ccl11*, *Ccl17* and *Ccl22*; and cluster 1B: *Cysltr1*, *Alox5ap*, *Arg1*, *Adam8* and *Fcer1g*), the lung-specific insult response clusters with known extracellular matrix remodelling, tight junctions, and mucus production genes (*e.g.*, *Gstm1*, *Vegfa*, *Chi3l1*, and *Pcna*, in cluster 3B, *e.g.*, *Pard3*, *Prkci*, and in cluster 3C, *e.g.*, *Muc4*, *Muc5b* and *Muc5ac*). In contrast, stem cell proliferation clusters contain genes that are essential for DNA replication and cell proliferation, which were not previously attributed to asthma. Enrichment analysis indicated that these genes are expressed mainly in cells that mediate leukocyte proliferation [CD34⁺ and CD71⁺ cells^{21–23}] and tissue regeneration [CD105⁺ cells²⁴] in allergic asthma. Cluster 2 contains well-known circadian rhythm genes *e.g.*, *Nr1d1* (or *Rev-Erb α*), *Nr1d2*, *Per2*, and *Per3*. Most of them have not been related to asthma yet, but growing evidence suggests they play a role in the immune system^{25,26}. This evidence suggests that our approach is able to identify valid functional clusters within the overall acute asthma gene signature and to identify genes that are less explored in the context of asthma.

We utilized the network connectivity information to prioritize specific genes that are likely to play an important role in the above domains. We concentrated on two categories of genes: hub genes, *i.e.* genes that are central, given the network structure and peripheral genes, *i.e.* genes which, despite not having a large number of connections themselves, are connected to a set of genes having full mutual interplay (*cliques*, see methods). The assumption is that hub genes play an important role in the overall network, while peripheral genes have reciprocal influence on specific *cliques*. This approach led to the identification of 34 hub genes and 8 peripheral genes in the inflammation domain, and 29 hub genes and 11 peripheral genes in the lung-specific insult response domain. The majority of the peripheral genes (14/22) have not yet been associated with asthma, but most of them have been associated with inflammation (17/22; $p = 0.0139$ vs. asthma-associated peripheral genes, Fisher's exact test). In contrast, only 11/61 hub genes have not been associated with asthma ($p = 0.0002$ vs. asthma-associated peripheral genes, Fisher's exact test), and 4/61 have not been associated with inflammation ($p = 0.051$ vs. inflammation-associated peripheral genes, Fisher's exact test). Interestingly, some well-studied genes are hubs, *e.g.*, as *Ifng*, *Tlr4*, and *Ptprc*, but several others are not, *e.g.*, *Il13*, *Il4*, and *Il4r*. The ignormes hubs and peripheral genes are summarized in Supplementary Table S2. They include *Igsf6* and *Slc15a3*, which are expressed on antigen presenting cells and involved in antigen recognition^{27,28}, as well as the peripheral gene *Clec5a* (also known as *Mdl-1*), which is involved in the innate immune response to microorganisms^{29–31}. Within the lung-specific insult response domain, a hub gene *Tgfb1* and a peripheral gene *Bambi* belong to the *Tgfb* signalling pathway, important for lung remodelling³².

Super-connector genes link functional domains. We then searched for “super-connectors”, *i.e.* genes that link at least 5 clusters, and thus, identified 12 genes (Fig. 6). There are 9 super-connectors in the inflammation domain that are cytokines *Il6* and *Il1 β* , cell surface receptors *Cd4* and *Cd44*, signaling molecules *Stat1*, *Traf6*, and *Rela*, as well as *Cadm1*, a receptor that mediates mast cell adhesion to lung structural cells³³ and the glucocorticoid receptor *Nr3c1*. The other 3 of 12 molecules belong to cluster 3 (extracellular matrix remodelling) and

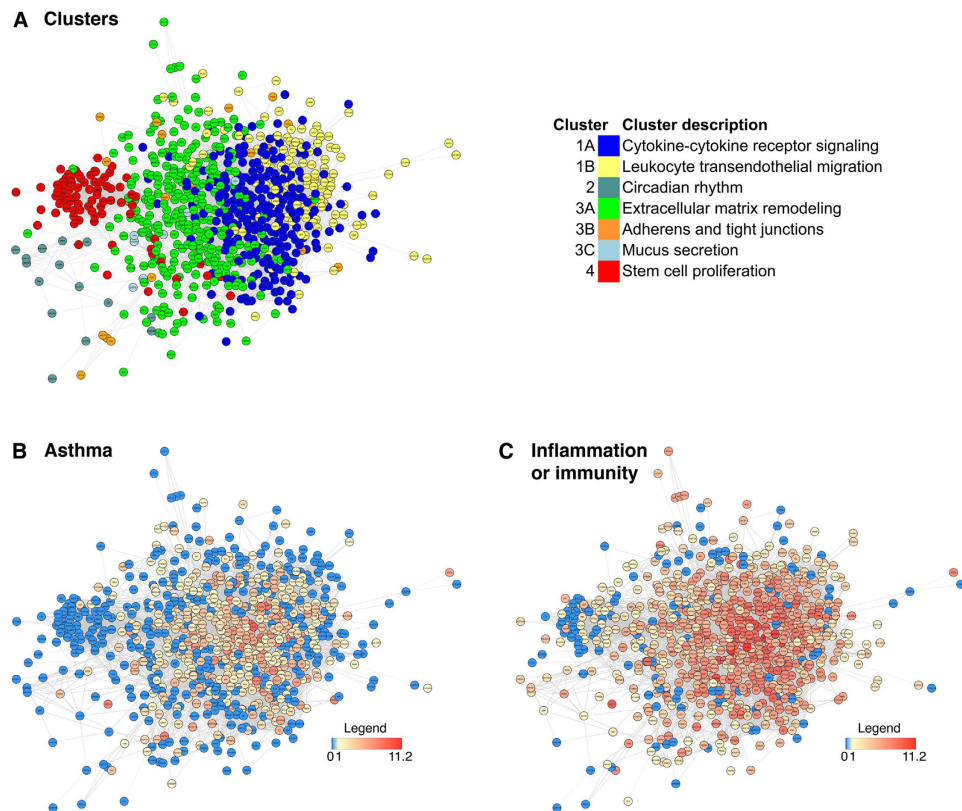


Figure 4. Network representation of 759 human ortholog genes in 7 clusters from our asthma-signature list. (A) Within the network, we have detected 7 main clusters using STRINGdb library in Bioconductor that were further functionally annotated to reveal biological functions with online tool EnrichR (for details please refer to supplementary material). Genes are coloured according to literature number associating each gene to (B) “asthma” or (C) “inflammation OR immunity”. For each gene, we have retrieved number of publications in PubMed in November 2014. Organic layout algorithm was used to produce this figure.

(Supplementary Fig. S5). These data indicate that genes, which are likely to play an important role, would often escape attention in a single experiment analysis due to their minimal fold-change, but can emerge from an approach that combines computational dataset analysis and gene network connectivity investigation.

Super-connectors, hubs and peripheral genes are affected by dexamethasone (DEX) treatment.

The glucocorticoid receptor *Nr3c1* is the only super-connector that linked all 4 major domains and robustly changed expression across microarray comparisons and our qPCR. Given existing knowledge of the role of *Nr3c1* in allergic asthma, we tested the effect of DEX, a steroid that binds to *Nr3c1*³⁸, in our animal model on both the disease phenotype and super-connector gene expression. As expected, DEX significantly suppressed airway and lung inflammation, mucus hypersecretion and lung Th2 cytokine gene expression (Supplementary Fig. S6). Out of the 9 super-connector genes that were significantly deregulated in our model, DEX partially or completely reverted 7 genes towards control values (Fig. 8). Although *Stat1* was not deregulated in our model, it was down-regulated significantly after DEX treatment, which is in agreement with a previous study³⁹. *Traf6* and *Vwf* were the only 2 super-connectors significantly down-regulated in our model that were unaffected by DEX treatment. Out of the 9 super-connectors belonging to the inflammation domain, 6 were partially or completely reverted by DEX treatment, whereas of the 3 super-connectors belonging to extracellular matrix remodelling, *Erbp2* was partially reverted by DEX, *Prkcd* was not deregulated, and *Vwf* was not reverted by the treatment. To determine whether genes other than super-connectors, were affected by DEX treatment, we randomly selected 12 hub genes and 4 peripheral genes from 6 large clusters in addition to the *Muc5ac* gene as a representative gene in cluster 3C, because this cluster does not contain peripheral or hub genes. We found that 15 of 17 analyzed genes (10 hub genes, 4 peripheral genes and *Muc5ac*) were differently regulated in our model and only one (hub gene *Ifng*) was not fully or partially reverted by DEX treatment (Supplementary Fig. S7). These data indicate that systemic treatment with glucocorticoids modulates genes regardless of their topological properties in the network.

Discussion

Here, we implement a novel computational approach used to analyse multiple microarray datasets derived from lungs of mice with acute onset allergic asthma and integrate it with independent experiments to unravel the asthma ignorome and dissect it into domains, clusters, hubs and peripheral genes.

Cluster	1A	1B	2	3A	3B	3C	4	
Description	Cytokine-cytokine receptor signaling	Leukocyte transendothelial migration	Circadian rhythm	Extracellular matrix remodeling	Adherens and tight junctions	Mucus secretion	Stem cell proliferation	
Gene number	239	114	15	282	19	8	82	
Gene Ontology Biological Processes ¹	Immune system process (GO:0002376); Defense response (GO:0006952); Immune response (GO:0006955); Chemotaxis (GO:0006935); Response to external stimulus (GO:0009605)	Immune system process (GO:0002376); Defense response (GO:0006952); Immune response (GO:0006955); Innate immune response (GO:0045087); Respiratory burst (GO:0045730)	Regulation of RNA metabolic process (GO:0051252); Regulation of transcription, DNA-dependent (GO:0006355); Heme biosynthetic process (GO:0006783); Regulation of transcription from RNA polymerase II promoter (GO:0006357); Regulation of transcription (GO:0045449)	Regulation of cell proliferation (GO:0042127); Negative regulation of apoptosis (GO:0043066); Negative regulation of programmed cell death (GO:0043069); Regulation of cell migration (GO:0030334); Regulation of cell motion (GO:0051270)	Establishment or maintenance of cell polarity (GO:0007163)	Monosaccharide metabolic process (GO:0005996); Cellular alcohol metabolic process (GO:0006066); Ventricular cardiac muscle morphogenesis (GO:0055010); Cardiac muscle contraction (GO:0060048); Carbohydrate metabolic process (GO:0005975)	Organelle organization (GO:0006996); Chromosome organization (GO:0051276); Regulation of cell cycle (GO:0051726); DNA replication (GO:0006260); DNA metabolic process (GO:0006259)	
KEGG Pathways ¹	Cytokine cytokine receptor interaction (HSA04060); Hematopoietic cell lineage (HSA04640); Toll like receptor signaling pathway (HSA04620); Cell adhesion molecules (HSA04514); JAK STAT signaling pathway (HSA04630)	Leukocyte transendothelial migration (HSA04670); Fc epsilon receptor signaling pathway (HSA04664); Phosphatidylinositol signaling system (HSA04070); Natural killer cell mediated cytotoxicity (HSA04650); Inositol phosphate metabolism (HSA00562)	Circadian rhythm (HSA04710); Glycine serine and threonine metabolism (HSA00260); Arginine and proline metabolism (HSA00330); Porphyrin and chlorophyll metabolism (HSA00860)	ECM receptor interaction (HSA04512); Focal adhesion (HSA04510); Complement and coagulation cascades (HSA04610); p53 signaling pathway (HSA04115); Arachidonic acid metabolism (HSA00590)	Tight junction (HSA04530)	Keratan sulfate biosynthesis (HSA00533); Glycan structures biosynthesis 1 (HSA01030); Olfactory transduction (HSA04740); Galactose metabolism (HSA00052)	Cell cycle (HSA04110); Purine metabolism (HSA00230); Pyrimidine metabolism (HSA00240)	
Human Atlas Enriched Tissues ^{2,3}	CD33 ⁺ Myeloid cells CD19 ⁺ B cells CD14 ⁺ Monocytes Whole blood Smooth muscle	CD14 ⁺ Monocytes CD33 ⁺ Myeloid cells Whole blood Lung	CD71 ⁺ early erythroid cells Lung CD56 ⁺ NK cells CD8 ⁺ T cells CD4 ⁺ T cells	Smooth muscle Lung	Bronchial epithelial cells CD105 ⁺ Endothelial cells CD34 ⁺ cells Whole blood CD71 ⁺ early erythroid cells	Trachea CD71 ⁺ early erythroid cells CD56 ⁺ NK cells CD14 ⁺ Monocytes Lung	B lymphocytes CD105 ⁺ Endothelial cells CD71 ⁺ early erythroid cells CD34 ⁺ cells	
Selected genes	Hubs ⁴	<i>Cd4, Il6, Stat1, Ifng, Cxcl10</i>	<i>Tyrobp, Plek, Itgb2, Clqa, Slc15a3</i>	<i>Alas1, Dbp, Gatm, Alas2, Nr1d1</i>	<i>Vegfa, Fn1, Erbb2, Itgb1, Cav1</i>	<i>Prkci, Sqstm1, Pard3, Pln, Grial</i>	<i>Muc5b, St3gal4, Clca1, Gaa, Tff2</i>	<i>Mad2l1, Nme1, Pttg1, TK1, Rrm2</i>
	Peripheral ⁵	<i>Mtpn, Hla-dbq2</i>	<i>Cotl1, Prickle1, Siglec8, Sdpr, Spon2, Clec5a</i>	–	<i>Tfpi2, Aqp5, Bambi, Myo1b, Efh2, Trpan7, Mmp19, Myc, Itgb1bp2, Tmem173, Bex1</i>	–	<i>Tff2, Slc26A4, Chst2</i>	<i>Adam19, Cd14</i>
	Super-connectors	<i>Nr3c1, cd4, Traf6, Rela, Stat1, Cadm1, Il1b, Cd44, Il6</i>	–	–	<i>Prckd, Vwf, Erbb2</i>	–	–	–
	Malacards Asthma ⁶	<i>Il13, Ccl11, Hhla-g, Il4r, Tlr6, Rnase3, Il4, Ccl17, Ccl5, Fcer2</i>	<i>Pla2g7, Alox5ap, Cysltr1, Adam8</i>	–	<i>Chia, Alox12, Alox15, Cat</i>	–	–	–
Percentage of genes in PubMed related to	“asthma”	75.7%	48.2%	13.3%	48.6%	42.1%	62.5%	14.6%
	“Inflammation or immunity”	96.7%	84.2%	73.3%	84.8%	63.2%	100.0%	56.1%

Table 2. Biological processes, pathways and tissue expression enrichment analysis of main clusters in asthma-signature genes list, examples of hubs, peripheral genes, super-connectors and genes present in Malacards, a gene-disease association database. ¹Maximum of 5 enriched categories ordered by Enrichr Combined score showing an adjusted P-value < 0.05. ²Maximum of 5 tissues within a list of 17 tissues potentially found in the whole lung tissue samples in datasets under study (See ³Supplementary Figure S4 for details on Human Atlas Gene expression for all genes in each cluster). ⁴Top 5 hub genes ordered by Betweenness Centrality. ⁵Peripheral genes defined as 1-degree nodes connected to a clique. ⁶Malacards overlap for Asthma (119 genes, as for Sept 2014).

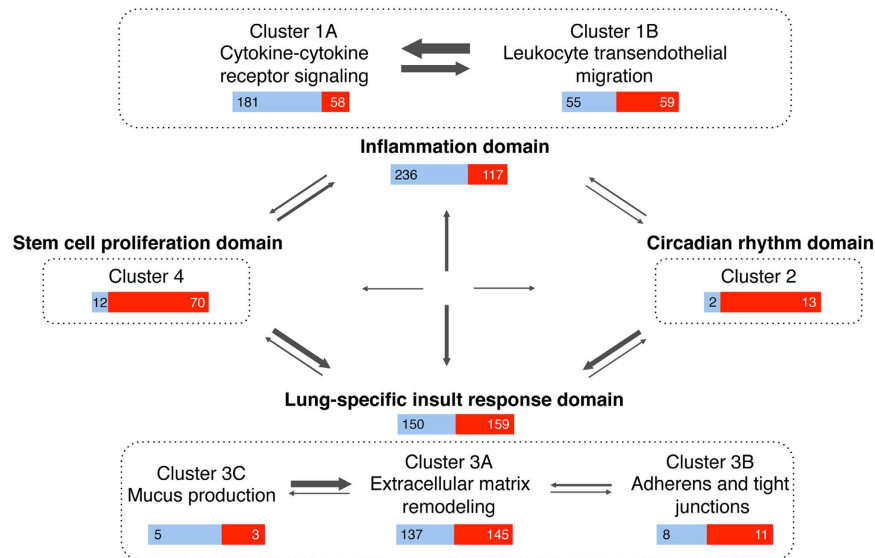


Figure 5. Schematic diagram depicting connectivity between 4 domains and individual clusters in acute asthma signature list. Bars indicate number of asthma-related (blue) and -ignorome (red) genes in each cluster or domain. Arrows indicating connections between clusters and domains are scaled according to relative connection strength between clusters and domains (see supplementary methods for details).

Technical limitations reduce the power of microarrays in detection of potentially important genes with limited variation in expression levels, especially when the sample size is small^{40,41}. Gene signals from small specific cell populations in complex tissues, such as lung, are hence diluted to low levels that are not accurately detected by microarray analyses^{40,41}. Diverse experimental protocols that use different allergens, mouse strains, sampling times may generate gene lists that reflect differences in the kinetics of the underlying biology of the allergic asthma response. Merging individual gene lists from experiments with varying conditions and different microarray platforms may exclude key experimental condition-specific genes or pathways. To address this issue, we utilized a pathway-driven approach, which only considers genes associated with significantly enriched GO:BP terms. The novelty of our overall strategy is the use of this pathway-driven approach, and, more importantly, the combination of both pathway- and gene-driven approaches. Our strategy reduced the biological noise and detected genes with significant < 1.5-fold changes that were excluded from the previous individual microarray studies^{2,10,19,20} and meta-analysis¹⁶.

Our signature gene list consists of 41% asthma-annotated genes and 59% of genes not previously associated with disease, defined as the ignorome¹⁷. Of the asthma-annotated genes, 13 genes associated with Th2 and inflammatory responses account for 1/3 of all asthma related publications and the remaining genes play roles in inflammation. Strikingly, there are a large number of ignorome genes involved in inflammatory pathways, DNA replication, cell cycle and wound healing, which may include important disease targets. Taken together, these data support the notion that biomedical research is skewed towards a relatively small number of 'successful' genes^{12,13,17,42}. We argue that novel insights into the pathogenesis of allergic asthma should be achieved irrespective of the current literature bias. The main ignorome clusters relate to stem cell proliferation and circadian rhythm domains. Both functional domains are well established in asthma^{21,22,24,43} and yet the individual genes identified in this study have as yet not been associated to it, thus providing interesting candidates to pursue these functional areas in the disease.

Understanding hubs (predominantly asthma-annotated) and peripheral genes (mostly ignorome) is essential for establishing functional interplay within the clusters and determining why key Th2 genes e.g., *Il13*, *Il4*, and *Il4r* are neither hub nor peripheral genes^{44,45}. These data support the finding that there is no correlation between a gene's degree of interaction, importance in the pathogenesis, and frequency of occurrence in the scientific literature¹⁴.

Notably, all super-connectors are asthma-annotated and belong to either inflammation or lung-specific insult response domains. In the DEX experiment, we found that only *Traf6* and *Vwf* had not changed expression levels. However, it is possible that they might have changed, if higher steroid doses were used, especially because steroids are known to increase *Vwf*^{46,47} and *Traf6*⁴⁸ expression. Although we focused on these super-connectors, there is evidence that 41% (314) of genes interconnect 3 or more clusters, 1/3 of them being an ignorome, further demonstrating a large constellation of interrelated clusters across tissues and biological domains.

There are caveats regarding the signature gene list and analyses presented here. Firstly, these analyses are based on the early events at the initiation of acute allergic asthma in the lungs of mice and may not reflect the same multiplicity of clusters and biological domains that would appear when testing chronic disease or disease relapse. Secondly, according to the current paradigm, genes in the inflammatory and immune response are expected, however, the remodelling response appears earlier than expected and these genes may reflect an early response of the lung to inflammatory signals and may represent the activation of healing mechanisms. Thirdly, it is not known

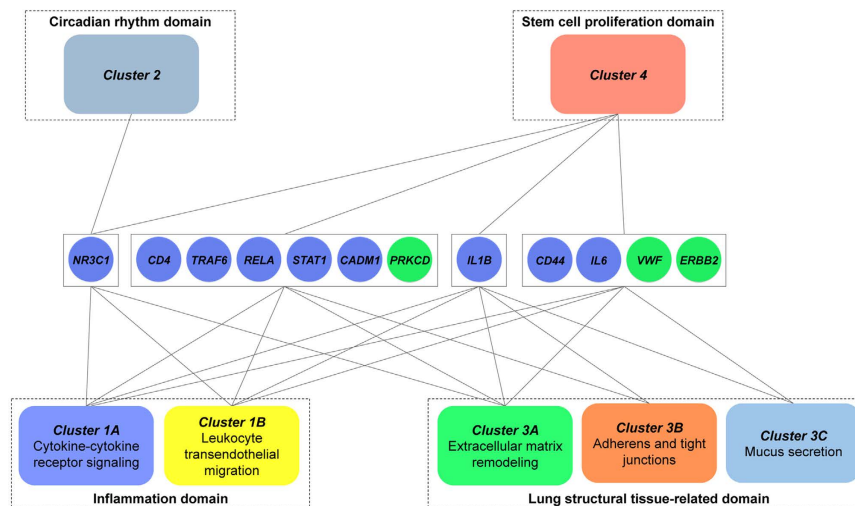


Figure 6. Schematic diagram depicting 12 super-connector genes in the asthma-signature gene list. Gene circle colour indicates the cluster of origin and connected to the pertaining topological cluster. Topological clusters are grouped into 4 biological domains by dotted rectangles.

Cluster	Gene symbol	qPCR at 72h	Microarray comparisons											
			A	B	C	D	E	F	G	H	I	J		
1A	CADM1	*												
	CD4	*												
	CD44	*												
	IL1B	*												
	IL6	*												
	NR3C1	*												
	RELA	*												
	STAT1	*												
3A	ERBB2	*												
	PRKCD	*												
	VWF	*												

Legend: nan, -4.5, -3, -1.5, 0, 1.5, 3, 4.5

Figure 7. Expression profiles of 8 from 12 super-connectors were confirmed with quantitative PCR. Data are presented as mean \log_2 fold changes of gene expression by quantitative real-time PCR and microarray relative to control mice. Quantitative real-time PCR data were determined in whole lung extracts and are pooled from 2 independent experiments ($n = 6$). Microarray data are from 6 publicly available datasets broken down into 10 direct comparisons of asthmatic and control mice (please refer to Supplementary data and Fig. 1 for details). * $p < 0.05$ compared with PBS challenged mice (unpaired t -test).

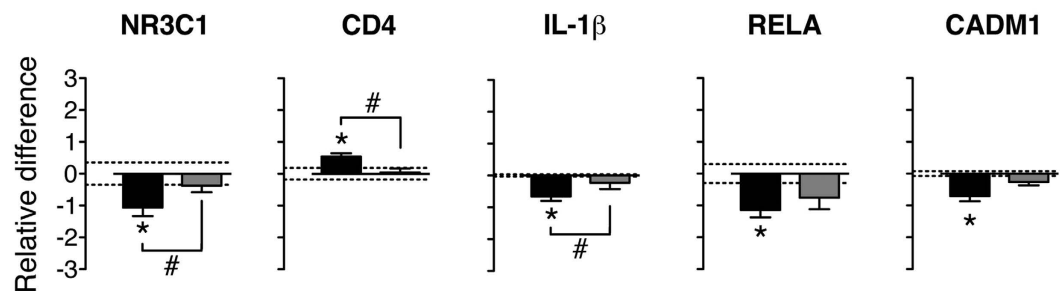
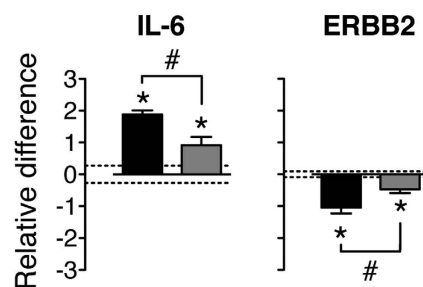
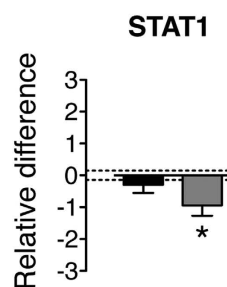
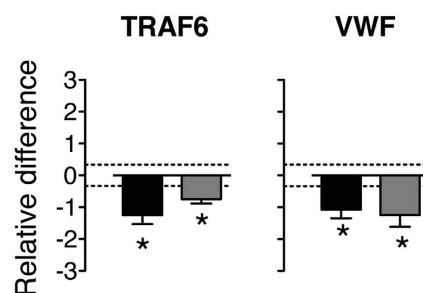
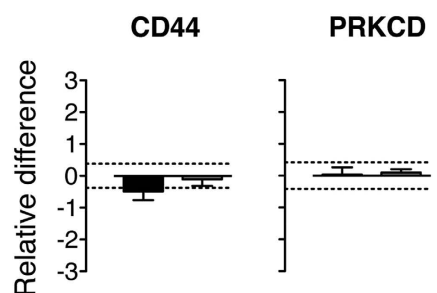
whether human gene expression profiles from patient lungs would produce a similar gene list and functional clusters, because there are only 2 publicly available human microarray datasets on lung biopsy samples [GEO database (<http://www.ncbi.nlm.nih.gov/geo/>)]. Importantly, this approach would enable comparisons between human and mouse datasets providing excellent opportunities to validate experimental models.

Asthma is a heterogeneous disease that includes several distinct disease phenotypes and endotypes⁴⁹ and is not usually studied at acute onset, but rather during ongoing disease. Multiple dataset analyses including more gene expression profiles from mouse and man and at different stages of disease and in distinct patient subgroups may result in the prioritization of essential genes and networks at the core of disease and simultaneously highlight molecular differences between phenotypes.

In conclusion, the integration of prior literature knowledge and unbiased multiple dataset analysis paradigms through the lens of systems biology provides important insights into asthma pathogenesis, allowing to bring novel biology and existing knowledge on a balanced plate, highlighting connections between the two and providing novel avenues to combat the disease.

Methods

Data selection. From the 23 experiments identified in the literature, we selected 6 datasets based on studies using the Affymetrix Mouse Genome 430 2.0 or Murine Genome U74 Version 2 Arrays microarray platform and mouse models of ovalbumin (OVA) and house dust mite induced acute asthma characterized by airway

A Completely reverted**B Partially reverted****C Down-regulated****D Not affected****E No change to PBS**

Legend: : : : PBS ■ Vehicle ■ DEX

Figure 8. Expression of super-connectors in acute allergic asthma after DEX treatment. Total lung RNA was extracted from mice intranasally challenged with PBS (controls) and mice that received vehicle or DEX before and after OVA challenge to induce allergic asthma. Samples were collected at 72 h after allergen challenge and super-connector expression was determined with quantitative PCR. Super-connectors are grouped according to their change to DEX treatment into (A) completely reverted, (B) partially reverted, (C) further down regulated, (D) not affected with DEX treatment, and (E) not changed in comparison to PBS controls. Data are presented as mean \pm SEM log₂ fold changes of gene expression relative to control (PBS) mice and are pooled from 2 independent experiments (n = 6). **p* < 0.05 compared with PBS challenged mice; #*p* < 0.05 compared with vehicle-treated group (unpaired *t*-test).

eosinophilia, airway hyperresponsiveness (AHR), and mucus hypersecretion. Within these independent datasets, we extracted in total 10 comparisons of asthmatic vs. control samples (see Table 1).

Multi-dataset bioinformatics analysis for selection of acute asthma signature list. We retrieved raw data from the GEO database¹ and analysed them using a common pipeline for signal normalization and extraction of differentially expressed genes (see supplementary methods and Fig. 1). We extracted differentially regulated genes for each experiment, which were then utilized for our multi-dataset bioinformatics analysis combining two approaches: a top-down, biological process enrichment-driven approach, from now on called

‘pathway-driven’, and a bottom-up, ‘gene-driven’ approach (see supplementary methods). The two approaches were merged to generate a final union list.

Gene annotation for acute asthma ignorome detection and its functional annotation. To assess the extent of the acute asthma ignorome, we performed an unambiguous literature search based on Entrez Gene and PubMed databases (see supplementary methods). To annotate the genes from the acute asthma signature list (including the asthma ignorome), we compiled a list of human orthologs, mapped it to the STRING interaction network, and performed a topological analysis for clusters of densely connected genes ($n \geq 5$; details in supplementary methods). For each cluster, we identified three sets of genes: (i) hub genes (the top 10% of the genes in the list ranked by betweenness centrality, defined as the ratio of shortest paths passing through a node, being a common measure of centrality of a node parameter in descending order); (ii) peripheral genes (1-degree nodes connected to a clique, *i.e.* a set of nodes completely connected to each other); and (iii) super-connectors (genes connecting more than 5 clusters). Each cluster was then further characterized in terms of functional enrichment, tissue specific expression and literature representation (see supplementary methods).

Biological validation experiments. We used 8–10 week old female BALB/c mice (Charles River, Sulzfeld, Germany). All experimental protocols were approved by the ethical committee of the Medical University of Vienna and the Animal Care Committee of the Austrian Ministry of Science and carried out in accordance with the approved guidelines. All mice were immunized with OVA (Sigma Chemical Co., St. Louis, MO) dissolved in PBS intraperitoneally on days 0 and 21 and intranasally on day 32 with OVA or PBS. In treatment experiments, we administered vehicle or DEX (Sigma) intraperitoneally at 1 mg/kg 30 min before and 24 and 48 h after challenge. Details on animal treatment and measurements of airway hyperresponsiveness, airway and lung inflammation, mucus hypersecretion and serum OVA-specific antibody titers are reported in supplementary methods. Separate groups of mice were used for lung mRNA extraction at 24 or 72 h after intranasal challenge with PBS or OVA. Methods for RNA extraction, reverse-transcription, quantitative polymerase chain reaction and statistical analysis are reported in supplementary information.

References

- Chapman, D. G., Tully, J. E., Nolin, J. D., Janssen-Heininger, Y. M. & Irvin, C. G. Animal models of allergic airways disease: where are we and where to next? *J. Cell. Biochem.* **115**, 2055–2064 (2014).
- Camateros, P. *et al.* Modulation of the allergic asthma transcriptome following resiquimod treatment. *Physiological Genomics* **38**, 303–318 (2009).
- Karp, C. L. *et al.* Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat Immunol* **1**, 221–226 (2000).
- Tumes, D. J., Connolly, A. & Dent, L. A. Expression of survivin in lung eosinophils is associated with pathology in a mouse model of allergic asthma. *International Immunology* **21**, 633–644 (2009).
- Di Valentin, E. *et al.* New asthma biomarkers: lessons from murine models of acute and chronic asthma. *AJP: Lung Cellular and Molecular Physiology* **296**, L185–L197 (2008).
- Stevens, T., Krantz, Q. T., Linak, W. P., Hester, S. & Gilmour, M. I. Increased Transcription of Immune and Metabolic Pathways in Naive and Allergic Mice Exposed to Diesel Exhaust. *Toxicological Sciences* **102**, 359–370 (2007).
- Tolgyesi, G. *et al.* Gene expression profiling of experimental asthma reveals a possible role of paraoxonase-1 in the disease. *International Immunology* **21**, 967–975 (2009).
- Zimmermann, N. *et al.* Dissection of experimental asthma with DNA microarray analysis identifies arginase in asthma pathogenesis. *J Clin Invest* **111**, 1863–1874 (2003).
- King, N. E. *et al.* Expression and regulation of a disintegrin and metalloproteinase (ADAM) 8 in experimental asthma. *American Journal of Respiratory Cell and Molecular Biology* **31**, 257–265 (2004).
- Tachdjian, R. *et al.* Pathogenicity of a disease-associated human IL-4 receptor allele in experimental asthma. *Journal of Experimental Medicine* **206**, 2191–2204 (2009).
- Zimmermann, N. *et al.* Transcript signatures in experimental asthma: identification of STAT6-dependent and -independent pathways. *J Immunol* **172**, 1815–1824 (2004).
- Pfeiffer, T. & Hoffmann, R. Temporal patterns of genes in scientific publications. *PNAS* **104**, 12052–12056 (2007).
- Su, A. I. & Hogenesch, J. B. Power-law-like distributions in biomedical publications and research funding. *Genome Biol.* **8**, 404 (2007).
- Hoffmann, R. & Valencia, A. Life cycles of successful genes. *Trends Genet.* **19**, 79–81 (2003).
- Freishtat, R. J. *et al.* Delineation of a gene network underlying the pulmonary response to oxidative stress in asthma. *J. Investig. Med.* **57**, 756–764 (2009).
- Novershtern, N., Itzhaki, Z., Manor, O., Friedman, N. & Kaminski, N. A Functional and Regulatory Map of Asthma. *Am J Respir Cell Mol Biol* **38**, 324–336 (2008).
- Pandey, A. K., Lu, L., Wang, X., Homayouni, R. & Williams, R. W. Functionally enigmatic genes: a case study of the brain ignorome. *PLoS ONE* **9**, e88889 (2014).
- Stupka, E. The emergence of genomics in mouse asthma models. *Drug Discovery Today-Disease Models* **6**, 119–127 (2009).
- Lu, X., Jain, V. V., Finn, P. W. & Perkins, D. L. Hubs in biological interaction networks exhibit low changes in expression in experimental asthma. *in* **3**, 98 (2007).
- Wang, T. *et al.* Murine Lung Responses to Ambient Particulate Matter: Genomic Analysis and Influence on Airway Hyperresponsiveness. *Environ Health Perspect* **116**, 1500–1508 (2008).
- Sehmi, R. *et al.* Allergen-induced increases in IL-5 receptor alpha-subunit expression on bone marrow-derived CD34+ cells from asthmatic subjects. A novel marker of progenitor cell commitment towards eosinophilic differentiation. *J Clin Invest* **100**, 2466–2475 (1997).
- Dorman, S. C. *et al.* Sputum CD34+IL-5Ralpha+ cells increase after allergen: evidence for *in situ* eosinophilopoiesis. *Am J Respir Crit Care Med* **169**, 573–577 (2004).
- Mazzarella, G. *et al.* Phenotypic features of alveolar monocytes/macrophages and IL-8 gene activation by IL-1 and TNF-alpha in asthmatic patients. *Allergy* **55** Suppl 61, 36–41 (2000).
- Bentley, J. K. *et al.* Ovalbumin sensitization and challenge increases the number of lung cells possessing a mesenchymal stromal cell phenotype. *Respir Res* **11**, 127 (2010).

25. Scheiermann, C., Kunisaki, Y. & Frenette, P. S. Circadian control of the immune system. *Nature Publishing Group* **13**, 190–198 (2013).
26. Curtis, A. M., Bellet, M. M., Sassone-Corsi, P. & O'Neill, L. A. J. Circadian clock proteins and immunity. *Immunity* **40**, 178–186 (2014).
27. Bates, E. E. *et al.* CD40L activation of dendritic cells down-regulates DORA, a novel member of the immunoglobulin superfamily. *Mol. Immunol.* **35**, 513–524 (1998).
28. Daniel, H. & Kottra, G. The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology. *Pflugers Arch.* **447**, 610–618 (2004).
29. Aoki, N. *et al.* Expression and functional role of MDL-1 (CLEC5A) in mouse myeloid lineage cells. *Journal of Leukocyte Biology* **85**, 508–517 (2009).
30. Chen, S.-T. *et al.* CLEC5A regulates Japanese encephalitis virus-induced neuroinflammation and lethality. *PLoS Pathog* **8**, e1002655 (2012).
31. Chen, S.-T. *et al.* CLEC5A is critical for dengue-virus-induced lethal disease. *Nature* **453**, 672–676 (2008).
32. Halwani, R., Al-Muhsen, S., Al-Jahdali, H. & Hamid, Q. Role of transforming growth factor- β in airway remodeling in asthma. *American Journal of Respiratory Cell and Molecular Biology* **44**, 127–133 (2011).
33. Moiseeva, E. P., Roach, K. M., Leyland, M. L. & Bradding, P. CADM1 is a key receptor mediating human mast cell adhesion to human lung fibroblasts and airway smooth muscle cells. *PLoS ONE* **8**, e61579 (2013).
34. Steinberg, S. F. Distinctive activation mechanisms and functions for protein kinase Cdelta. *Biochem. J.* **384**, 449–459 (2004).
35. Hillgruber, C. *et al.* Blocking von Willebrand factor for treatment of cutaneous inflammation. *J. Invest. Dermatol.* **134**, 77–86 (2014).
36. Gutierrez, C. & Schiff, R. HER2: biology, detection, and clinical implications. *Arch. Pathol. Lab. Med.* **135**, 55–62 (2011).
37. Lu, X., Jain, V. V., Finn, P. W. & Perkins, D. L. *Hubs in biological interaction networks exhibit low changes in expression in experimental asthma.* in **3**, 98 (2007).
38. Niu, N. *et al.* Human glucocorticoid receptor alpha gene (NR3C1) pharmacogenomics: gene resequencing and functional genomics. *J. Clin. Endocrinol. Metab.* **94**, 3072–3084 (2009).
39. Hu, X., Li, W.-P., Meng, C. & Ivashkiv, L. B. Inhibition of IFN-gamma signaling by glucocorticoids. *J Immunol* **170**, 4833–4839 (2003).
40. Wurmbach, E. *et al.* Validated genomic approach to study differentially expressed genes in complex tissues. *Neurochem. Res.* **27**, 1027–1033 (2002).
41. Draghici, S., Khatri, P., Eklund, A. C. & Szallasi, Z. Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet.* **22**, 101–109 (2006).
42. Hoffmann, R. & Valencia, A. Life cycles of successful genes. *Trends Genet.* **19**, 79–81 (2003).
43. Smolensky, M. H., Lemmer, B. & Reinberg, A. E. Chronobiology and chronotherapy of allergic rhinitis and bronchial asthma. *Adv. Drug Deliv. Rev.* **59**, 852–882 (2007).
44. Ingram, J. L. & Kraft, M. IL-13 in asthma and allergic disease: asthma phenotypes and targeted therapies. **130**, 829–42– quiz 843–4 (2012).
45. Kelly-Welch, A. E., Hanson, E. M., Boothby, M. R. & Keegan, A. D. Interleukin-4 and interleukin-13 signaling connections maps. *Science* **300**, 1527–1528 (2003).
46. Jilma, B. *et al.* High dose dexamethasone increases circulating P-selectin and von Willebrand factor levels in healthy men. *Thromb. Haemost.* **94**, 797–801 (2005).
47. Brotman, D. J. *et al.* Effects of short-term glucocorticoids on hemostatic factors in healthy volunteers. *Thromb. Res.* **118**, 247–252 (2006).
48. Sun, H. *et al.* TRAF6 inhibition rescues dexamethasone-induced muscle atrophy. *Int J Mol Sci* **15**, 11126–11141 (2014).
49. Martinez, F. D. & Vercelli, D. Asthma. *Lancet* **382**, 1360–1372 (2013).

Acknowledgements

The authors acknowledge excellent technical help of Drs. Rui-Yun Lee, Gordana Mokrovic and Daniela Reiner. This study was partially supported by the European Commission through an FP7- IAPP Marie Curie Action, P3AGI project (GA 230739).

Author Contributions

M.R. performed multi-dataset bioinformatics analysis, network analysis and contributed to the preparation of the manuscript; J.M.G.M. performed multi-dataset bioinformatics, network analysis and contributed to the preparation of the manuscript; B.B. designed and performed *in vivo* experiments, analysed the samples and to the preparation of the manuscript; D.C. performed, supervised computational analysis, and contributed to the preparation of the manuscript; P.M. contributed to *in vivo* experiments and sample analyses; C.L. participated in quantitative PCR analyses; M.M.E. designed *in vivo* experiments, contributed to the preparation of the manuscript; E.S. designed and supervised computational analysis, contributed to the preparation of the manuscript. All authors read and approved final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Riba, M. *et al.* Revealing the acute asthma ignorome: characterization and validation of uninvestigated gene networks. *Sci. Rep.* **6**, 24647; doi: 10.1038/srep24647 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>