BMC Systems Biology

BioMed Central

Research article

Open Access

# A module-based analytical strategy to identify novel disease-associated genes shows an inhibitory role for interleukin 7 Receptor in allergic inflammation

Reza Mobini*[1], Bengt A Andersson[2], Jonas Erjefält[3], Mirjana Hahn-Zoric[2], Michael A Langston[4], Andy D Perkins[4], Lars Olaf Cardell[5] and Mikael Benson[1]

Address: [1]Unit for Clinical Systems Biology, Department of Pediatrics, University of Gothenburg, Gothenburg, Sweden, [2]Department of Immunology, Sahlgrenska Academy, Gothenburg, Sweden, [3]Department of Experimental Medicine, Lund University, Lund, Sweden, [4]Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA and [5]Department of Oto-rhino-largyngeology, Karolinska Institute, Stockholm, Sweden

Email: Reza Mobini* - reza.mobini@pediat.gu.se; Bengt A Andersson - bengt.a.andersson@vgregion.se; Jonas Erjefält - jonas.Erjefalt@med.lu.se; Mirjana Hahn-Zoric - mirjana.hahn-zoric@immuno.gu.se; Michael A Langston - langston@eecs.utk.edu; Andy D Perkins - aperkins@cs.utk.edu; Lars Olaf Cardell - lars-olaf.cardell@ki.se; Mikael Benson - mikael.benson@vgregion.se

* Corresponding author

Published: 12 February 2009

*BMC Systems Biology* 2009, **3**:19 doi:10.1186/1752-0509-3-19

Received: 5 August 2008
Accepted: 12 February 2009

This article is available from: http://www.biomedcentral.com/1752-0509/3/19

© 2009 Mobini et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** The identification of novel genes by high-throughput studies of complex diseases is complicated by the large number of potential genes. However, since disease-associated genes tend to interact, one solution is to arrange them in modules based on co-expression data and known gene interactions. The hypothesis of this study was that such a module could be a) found and validated in allergic disease and b) used to find and validate one ore more novel disease-associated genes.

**Results:** To test these hypotheses integrated analysis of a large number of gene expression microarray experiments from different forms of allergy was performed. This led to the identification of an experimentally validated reference gene that was used to construct a module of co-expressed and interacting genes. This module was validated in an independent material, by replicating the expression changes in allergen-challenged $CD4^+$ cells. Moreover, the changes were reversed following treatment with corticosteroids. The module contained several novel disease-associated genes, of which the one with the highest number of interactions with known disease genes, *IL7R*, was selected for further validation. The expression levels of *IL7R* in allergen challenged $CD4^+$ cells decreased following challenge but increased after treatment. This suggested an inhibitory role, which was confirmed by functional studies.

**Conclusion:** We propose that a module-based analytical strategy is generally applicable to find novel genes in complex diseases.

## Background

Most common diseases, including allergy and diabetes, are complex. The genetic susceptibility of an individual to such a disease appears not to be the result of a single causative gene but rather arises from multiple interacting genes. The identification of novel disease-associated genes is complicated not only by the large number of potentially relevant genes as well as the heterogeneity of the diseases. One approach to address this complexity is to arrange disease-associated genes in networks, where the genes form nodes and the interactions between the genes are represented by links between the nodes. Since disease genes tend to interact [1,2] the investigation may be facilitated by searching for sub-networks of co-expressed and interacting genes (such sub-networks will henceforth be referred to as modules). The rationale behind this is that in gene interaction networks, functionally related genes tend to form modules of tightly interacting genes [1,2]. Thus, such modules give an overview of the main disease mechanisms and could thus be dissected to find novel genes [3,4]. It is not known if such a module-based analysis is generally applicable to complex diseases that are caused by altered interactions between many different cell types and tissues. The hypothesis of this study was that such a module could be a) found and validated in allergic disease and b) used to find and validate one or more novel disease-associated genes (i.e. genes that either show disease-associated changes in expression or whose alleles influence the risk of disease). Allergy was chosen as a model of complex diseases because the main disease process is known and readily examined in both clinical and experimental studies; an allergen causes increased release of cytokines from type 2 T helper (Th2) cells which activate B cells, eosinophils and mast cells to release inflammatory proteins [5]. This disease process is thought to be common to different forms of allergy, such as seasonal allergic rhinitis (SAR), atopic eczema and asthma. It involves many well-documented genes that can be used as references to validate the bioinformatic analyses as well as to construct modules of co-expressed and interacting genes. Moreover, since disease genes tend to interact, interactions with known disease genes can be used as a criterion to select novel genes for functional validation. However, several observations indicate more complex disease mechanisms than implied by the Th2 induced disease process. Many other cell types and tissues have been implicated and genomic high-throughout studies have shown the involvement of hundreds of genes. Several of these genes do not appear to overlap in different forms of allergy or to have roles that can be explained by the Th2 paradigm [6-11].

Based on the underlying principle that disease-associated genes tend to interact, we first identified a gene with a key regulatory role for the disease. This was used as a reference gene to construct a module and then dissected that to find the novel gene that had most interactions with known disease-associated genes. This was achieved by combining integrated analysis of gene expression microarray data with experimental studies.

We propose that these module-based analytical principles may be generally applicable to find novel genes in complex diseases.

## Results

The hypothesis of this study has been that a module of disease-associated genes could be found and validated in allergic disease, and then used to find and validate one or more novel disease genes. A step-wise analytical approach was employed (figure 1):

First, a gene with key regulatory role for the disease was identified based on converging data from integrated analysis of gene expression microarray data from allergic patients and healthy controls, information in the literature and experimental validation studies. Next, this gene was used to build a hypothetical module of genes co-regulated by the same transcription factors (TFs). The underlying assumption was that functionally related genes, such as those in a pathway or a module, are often co-regulated by the same TFs. The TFs and their target genes were identified and organized into a module by combining comparative promoter analysis, gene expression data and known gene interactions. This hypothetical module was tested in an independent material, by examining if the expression changes of the module genes could be replicated. Finally, the module was searched for a novel disease gene that was validated by functional studies.

### *1) Identification of disease-associated genes by integrated analysis of gene expression microarray data from cells and tissues from patients with different forms of allergy and healthy controls*

Combinatorial algorithms were used to extract groups of highly inter-connected genes (cliques) from graphs derived from transcriptomal correlation matrices containing millions of correlate pairs [12-14]. This resulted in the identification of 103 genes that will henceforth be referred to as disease-associated genes.

### *2) Identification of a common pathway among the disease-associated genes*

A manually curated pathway database (Ingenuity Pathway Analysis, IPA) was used to identify pathways among the disease-associated genes (see additional figure 1). IPA is a bioinformatics program that is based on manual review of more than 200,000 scientific publications. The resulting network of direct physical, transcriptional, and enzymatic interactions between mammalian orthologues was used

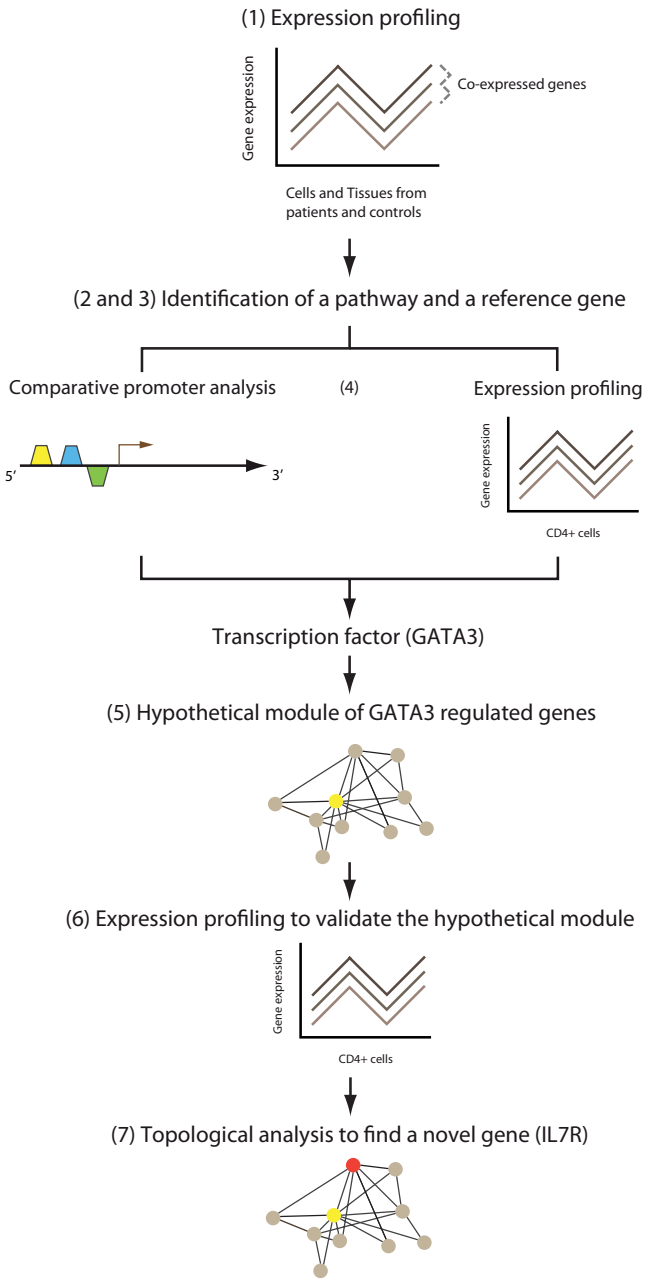



**Figure 1**

**An outline of the analytical process in the study**. 1) Integrated analysis of gene expression microarrays to find disease-associated genes, 2) Bioinformatic identification of a key pathway among those genes 3) Selection and experimental validation of a reference gene in the pathway, 4) Combined comparative promoter and gene expression microarray analysis to find a transcription factor that co-regulated the reference genes and other putative module genes, 5) Construction of a hypothetical module based on genes co-regulated by the transcription factor, 6) Independent gene expression microarray studies to validate the hypothetical module, 7) Topological analysis of module genes to find a novel gene. Experimental validation of the gene.

as a template to organize the disease-associated genes into statistically scored pathways. The pathway that received the highest statistical score was T cell receptor (TCR) signaling ($p < 10^{-6}$), which included genes belonging to the TCR complex (*TCRA*, *TCRB*, *CD3Z* and *CD3D*) as well as down-stream genes, *ZAP70*, *LAT*, *LCK* and *ITK*. In addition, the remaining disease-associated genes were searched for genes that had direct interactions with those generated by the pathway database. Three other disease-associated genes fulfilled this criterion, namely *IKBKG* [15], *LEF1* [16], and *MAP4K1* [17]. The above mentioned genes will henceforth be referred to as TCR pathway genes.

### *3) Identification and experimental validation of a reference gene in the TCR pathway*

This step aimed to find a reference gene in the TCR pathway that in the next step would be used to build a hypothetical module of functionally related disease genes. We proceeded by examining the TCR pathway in allergen-challenged $CD4^+$ cells from patients with seasonal allergic rhinitis (SAR). Gene expression microarray analyses of these cells showed a more than a two-fold increase in expression of six of the TCR pathway genes; *CD3D*, *CD3Z*, *LCK*, *ITK*, *ZAP70* and *MAP4K1*. Of these, *ITK* was selected as a putative reference gene because of its involvement in the polarization of Th2 cells and its location in a chromosomal susceptibility region for allergy, 5q31-33 [18]. The functional role of *ITK* was explored in an *in vivo* mouse model of allergic inflammation in the upper and lower airways. Allergen challenge resulted in a marked goblet cell hyperplasia in bronchi and bronchioles (figure 2A). This response was significantly reduced in similarly treated $ITK^{-/-}$ animals. In both upper and lower airways the allergen exposure led to eosinophilia that was significantly reduced in $ITK^{-/-}$ mice (figure 2B).

### *4) Identification of transcription factors that regulated* ITK

Functionally related genes, such as those in a pathway or a module, are often co-regulated by the same transcription factors (TFs). Thus, if the TF that regulated *ITK* and other genes of the TCR pathway could be identified this information could be used to find other disease-associated genes and organize them into a hypothetical disease module. Comparative promoter analysis revealed a framework of TF binding sites that was shared by *ITK* and two other TCR pathway genes, *CD3D* and *LEF1* (see additional file 1 for details). The framework contained TFBS for TF of the *GATA*, *MYB* and *MYT1* families (figure 3). Gene expression microarray analysis showed that the corresponding TF, namely *GATA3*, *MYB* and *MYBL1* increased in allergen-challenged $CD4^+$ cells (7-, 9- and 3-fold increases, respectively). To filter out the TF that was most relevant for Th2 polarization we analysed their correlations with *ITK* in gene expression microarray data from $CD4^+$ cells

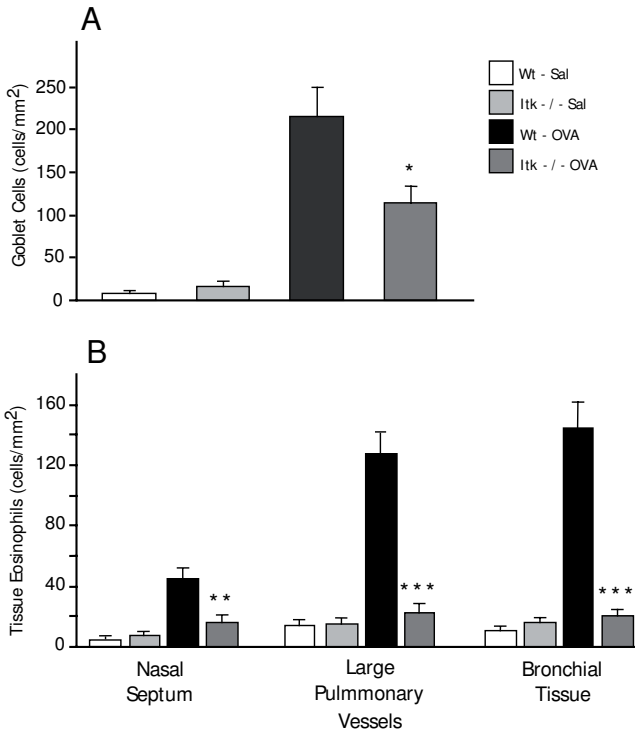



**Figure 2**
**The manifestation of an allergic inflammation is significantly reduced in mice lacking *ITK***. Numbers of PAS-positive bronchial goblet cells (A). EPO-histochemistry was used to quantify the tissue eosinophilia in the upper airways, bronchi, and around large pulmonary vessels (B). Data are presented as means ± SEM. Stars denote levels of statistical difference between OVA-treated wild-type and *ITK*-deficient animals (* = P < 0.05, ** = P < 0.01, *** = P < 0.001).

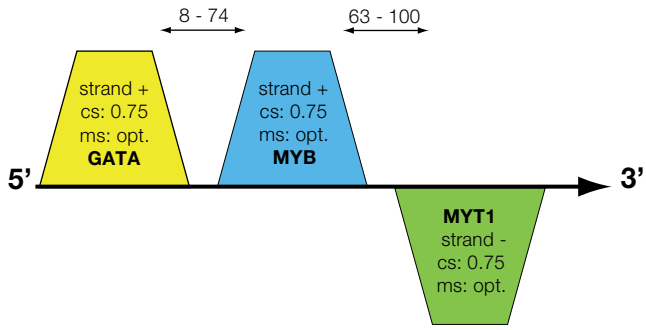



**Figure 3**
**Framework of three transcription factor binding sites**. The parameters and their abbreviations are: +, sense strand orientation; -, anti-sense strand orientation; cs, minimal core similarity; ms, minimal matrix similarity. Generally, cs was at default (0.75) and ms thresholds were the optimised default values for the respective matrices; distance, number of nucleotides to the next element (arrows in the upper part of the figure). *GATA*; strand, +; distance to next element, 8–74 bp. *MYBL*; strand, +; distance to next element, 63–100 bp. *MYT1*; strand, -.

from 73 atopic patients and 19 non-atopic controls. The correlations between *ITK* to *GATA3*, *MYB* and *MYBL1* were r = 0.68 (p < 0.0001), 0.41 (p < 0.001) and 0.25 (p < 0.01), respectively. Because *GATA3* had the highest correlation coefficient it was chosen to construct a hypothetical module as described below.

#### *5) Construction of a hypothetical module of genes co-regulated by GATA3*

Since functionally related genes are often co-regulated by the same TF, we used *GATA3* to construct a hypothetical module. We started by identifying 47 genes previously described to interact with *GATA3*. To filter out those of the interacting genes that were relevant for human CD4+ cells we again analyzed the gene expression microarray data from the CD4+ cells from the 73 atopic patients and 19 non-atopic controls. Genes whose expression values correlated with *GATA3* (r > 0.4 and p < 0.0001) were considered as hypothetical module genes. This resulted in a module that comprised 37 genes, several of which had key roles in allergic disease, for example *IFNG* and *IL4R* (table 1).

#### *6) Validation of the hypothetical module*

To validate the module we hypothesized that the module genes would show significant gene expression changes in an independent material of allergen-challenged CD4+ cells from 19 patients with SAR. Indeed, out of 33 genes that had detectable levels, 23 did change significantly. These genes formed a tightly connected module with an average of 3 ± 1 interactions/gene (figure 4). Since glucocorticoids are effective in treating allergic inflammation we hypothesized that this treatment would significantly reverse the expression levels of the 23 module genes. Indeed, 16 of the 23 genes satisfied this hypothesis.

#### *7) Identification and validation of a novel disease gene in the module*

Eleven of the 23 module genes that had been validated by the independent gene expression microarray study (figure 4) had not been previously been described in allergic diseases. These were *EED*, *ESR1*, *GRPR44*, *HSDP1*, *IL7R*, *LEF1*, *MAF*, *PTPRJ*, *RUNX3*, *TCF12*, and *SH2D1A*. Since disease-associated genes tend to interact [1,2] we computed the number of interactions between each of these eleven genes and the remaining, known disease-associated genes. The average number of interactions was 2.6 ± 0.4. *IL7R* had the highest number of interactions (n = 5) that included signature genes for Th2 activation like *IL4R* and *IFNG*. The expression levels of *IL7R* were lower in allergen-challenged CD4+ cells compared to diluent-

**Table 1: Genes included in the hypothetical module**

| Gene symbol | Gene Name |
|---|---|
| CCR4 | Chemokine (C-C motif) receptor 4 |
| CD28 | Antigen CD28 |
| CD3D | CD3 Antigen, delta subunit |
| CFLAR | CASP8- and fadd-like apoptosis regulator |
| CTLA4 | Cytotoxic t lymphocyte-associated 4 |
| EDG1 | Endothelial differentiation gene 1 |
| EED | Embryonic ectoderm development protein |
| ESR1 | Estrogen receptor 1 |
| GATA2 | GATA-binding protein 2 |
| GATA3 | GATA binding protein 3 |
| GPR44 | G protein-coupled receptor 44 |
| HIVEP2 | HUMAN immunodeficiency virus type 1 enhancer-binding protein 2 |
| HSPD1 | Heat-shock 60-kd protein 1 |
| IFNG | Interferon, gamma |
| IL4R | Interleukin 4 receptor |
| IL6 | Interleukin 6 |
| IL7 | interleukin 7 |
| IL7R | interleukin 7 receptor |
| IRF4 | Interferon regulatory factor 4 |
| IRS1 | Insulin receptor substrate 1 |
| ITK | IL2-inducible T-cell kinase |
| LEF1 | Lymphoid enhancer-binding factor 1 |
| LTA | Lymphotoxin-alpha |
| LTB | Lymphotoxin-beta |
| MAF | V-MAF avian musculoaponeurotic fibrosarcoma oncogene homolog |
| MAPK1 | Mitogen-activated protein kinase 1 |
| NFKB1 | Nuclear factor kappa-b, subunit 1 |
| NPPA | Natriuretic peptide precursor A |
| PP3CA | Protein Phosphatase 3 catalytic subunit, alpha isoform |
| PTCRA | Pre-t-cell receptor, alpha-chain precursor |
| PTPRJ | Protein-tyrosine phosphatase, receptor-type, J |
| RUNX3 | Runt-related transcription factor 3 |
| SH2D1A | SH2 domain protein 1A |
| STAT4 | Signal transducer and activator of transcription 4 |
| TCF12 | Transcription factor 12 |
| TCF3 | Transcription factor 3 |
| TNF | Tumor necrosis factor |

treated cells from patients, while treatment with corticosteroids increased the expression of *IL7R* (figure 5, both $p < 0.001$). Gene expression microarray analysis of an independent material of allergen-challenged $CD4^+$ cells from patients and healthy controls, also showed lower expression in patients, 3724 ± 512 vs. 5090 ± 519 ($p = 0.003$). These findings suggested the hypothesis that IL-7R signaling would inhibit the activation of Th2 cells. To test this hypothesis we examined if IL-7 would decrease production of the Th2 cytokine IL-4 from allergen-challenged PBMCs from allergic patients. However, IL-7 stimulation had the opposite effect; a 3-fold increased production of IL-4, 1.10 ± 0.45 compared to 3.22 ± 1.73 pg/mL, $p < 0.05$ (figure 6A). Similarly, the Th2 cytokines IL-5 and IL-13 also increased following IL-7 stimulation (see additional file 1). This suggested an alternative hypothesis, namely that IL-7R signaling also activated Th1 cells, but to a higher degree than that of Th2 cells, thereby causing a shift in the balance between Th1/Th2 cells towards a Th1 polarization. Indeed, IL-7 treatment caused a 7-fold increase of IFN-γ, 37 ± 13 compared to 253 ± 133 pg/mL, $p < 0.05$ (figure 6B). To measure the effect of IL-7 on the Th1/Th2 balance the ratio between IL-4/IFN-$\gamma^{-3}$ was computed; this ratio decreased significantly in allergen-challenged $CD4^+$ cells that were treated with IL-7 compared to untreated cells, 30 compared to 15, $p < 0.05$ (figure 6C). Thus, by a relatively higher Th1 than Th2 stimulation IL-7 signaling had a Th1 polarizing effect, which is consistent with an inhibitory role of IL-7R in allergic inflammation.

## Discussion

Identification of novel genes in high-throughput studies of complex diseases may require choosing between hundreds of genes. Since disease genes tend to interact with

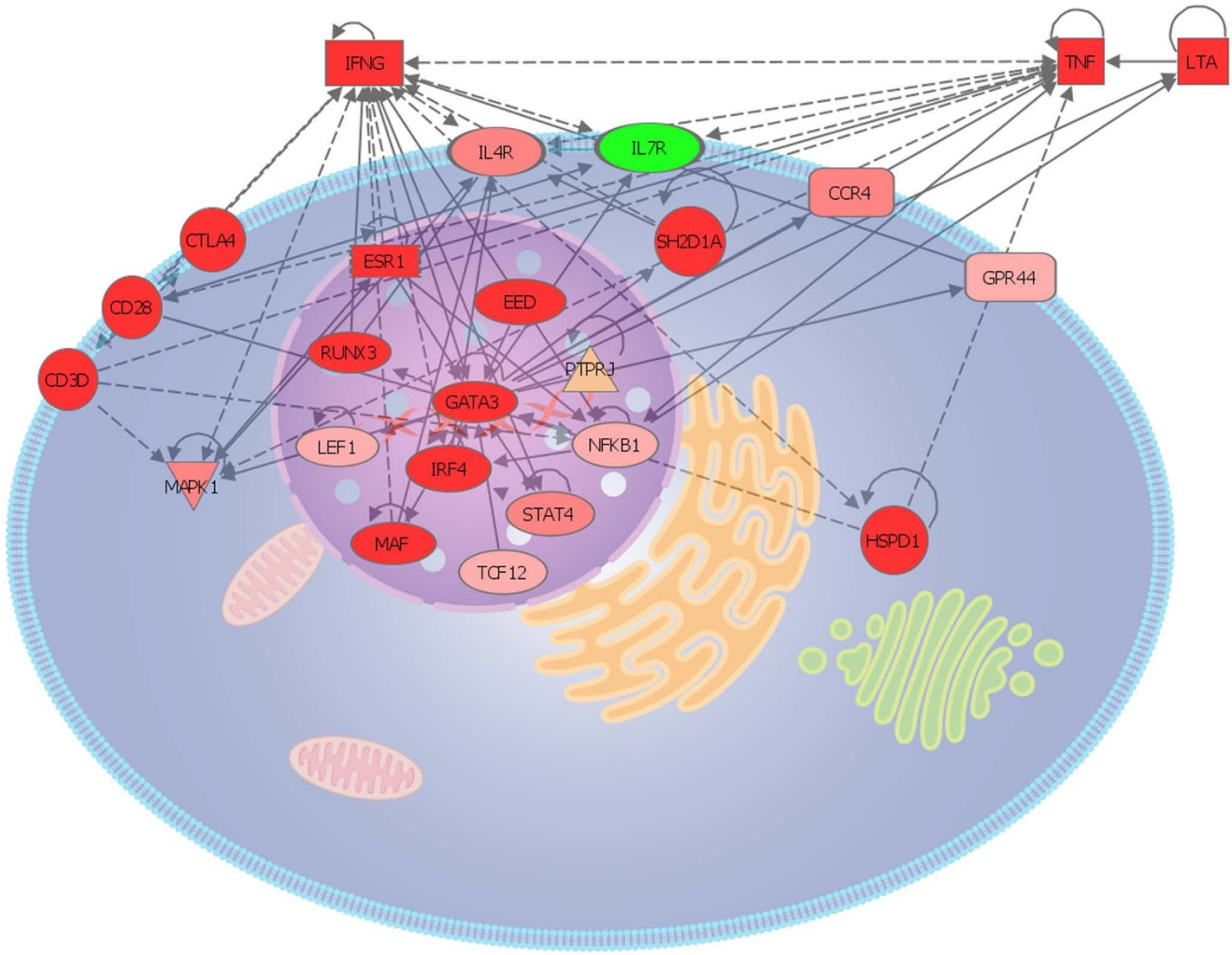



**Figure 4**

**Network representation of module formed by genes co-regulated by *GATA3* in allergen-challenged CD4⁺ cells from patients with seasonal allergic rhinitis**. The interaction network was generated using Ingenuity pathway analysis. Genes shaded in red were up-regulated whereas green were down-regulated. The color intensity reflects the significance value of the regulation. Solid and dotted lines represent direct and indirect interactions, respectively. Node shapes represent the class of molecule, for example donut shapes represent transcription factors (for detailed information see, http://www.ingenuity.com)

each other [1,2] the choice may be facilitated by searching for modules of disease genes that are co-expressed and interact [3,4].

It is not known if the module-based analytical strategy can also be applied to complex diseases other than cancer, which result from altered interactions between multiple cells and tissues. The hypothesis in this study was that we could find and validate a disease module in allergic disease and use this to find a novel disease gene.

In a recent study of breast cancer four known disease genes were used as references to construct a module of disease genes [3]. However, since allergy involves many different cells and tissues such an approach would result in a large number of reference genes to prioritize between. We therefore first performed an integrated analysis of gene expression microarray data from different forms of allergic diseases to find a common pathway. That pathway was used to find a reference gene for module construction. T cell receptor (TCR) signaling was the pathway with the

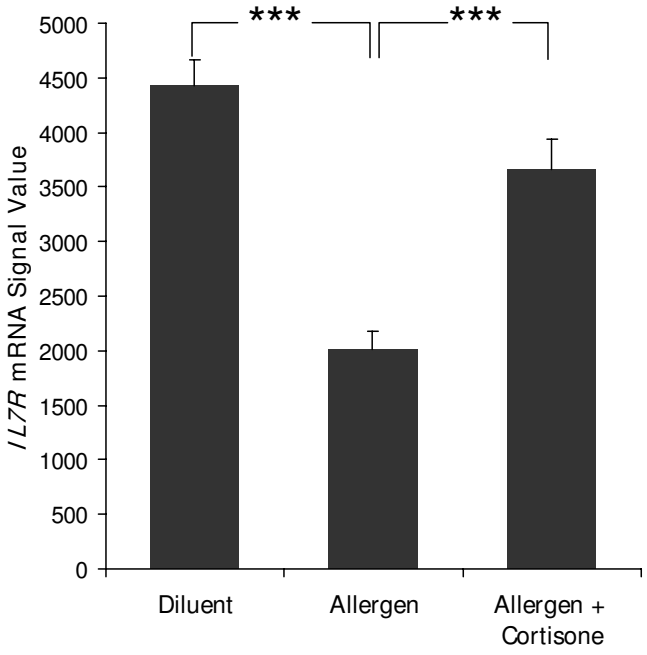



**Figure 5**
**Gene expression microarray analysis of *IL7R* expression in diluent-treated, and allergen-challenged CD4⁺ cells with and without treatment with corticosteroids**. Data are presented as means ± SEM (*** = P < 0.001).

highest statistical score. While many of the genes in the TCR pathway had not been previously described in human allergic disease, this pathway is in accordance with the current understanding that T cells play a key regulatory role in the disease [19]. On the other hand, it is important to emphasize the limitations of the study, for example that important low-abundance genes such as many cytokines are not detectable by gene expression microarray analyses. Only a small proportion of the different cells and tissues involved in allergic diseases were examined. Also, pathways other than TCR signaling may be important for the disease. Finally, the identification of pathways is restricted by current knowledge of gene functions and interactions.

We proceeded to study the TCR pathway genes in CD4⁺ cells from patients with seasonal allergic rhinitis (SAR). This lead to the identification of six differentially expressed genes, all of which could potentially be used as putative reference genes. However, in order to select one we focused on *ITK*, because of its involvement in the polarization of Th2 cells and its location in a chromosomal susceptibility region for allergy [18]. The relevance of *ITK* was validated in a mouse model of allergy. The number inflammatory cells in both the upper and lower airways was significantly decreased in allergen-challenged *ITK*-/- compared to wild type mice. Next, we identified

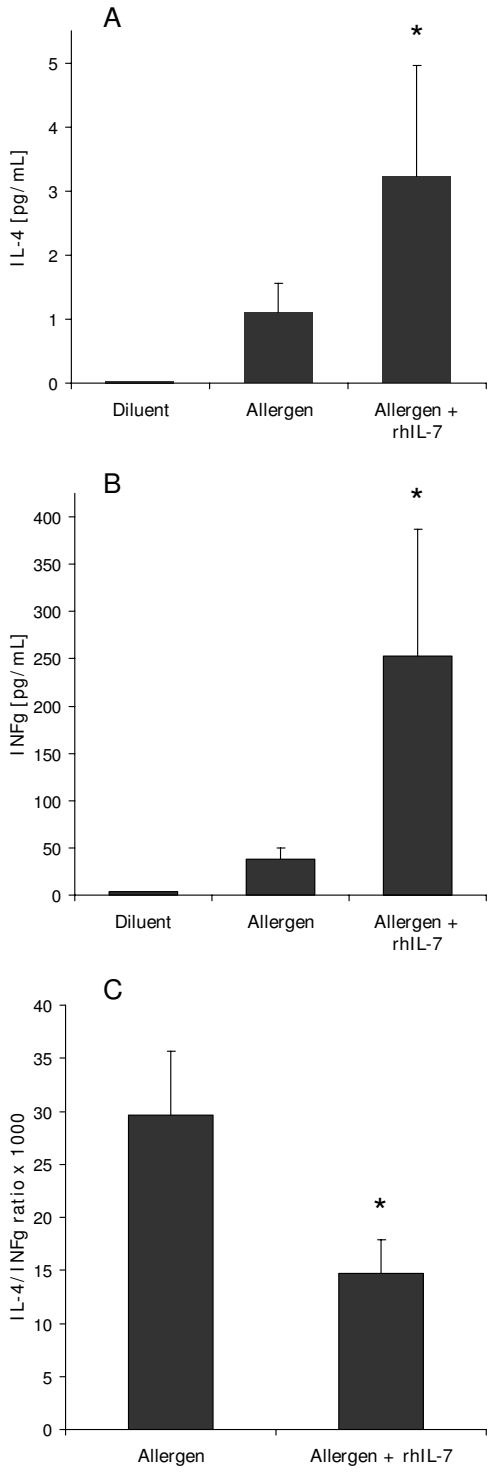



**Figure 6**
**A) IL-4, B) IFN-γ and C) the IL-4/IFN-γ[3] in the supernatants of allergen-challenged PBMC from patients with seasonal allergic rhinitis, treated with diluent, or allergen with and without rhIL-7**. Data are presented as means ± SEM (* = P < 0.05).

and validated a transcription factor (TF) that was shared by *ITK* and two other TCR pathway genes. This TF, *GATA3*, was used to construct a hypothetical module based on known gene interactions and co-expression data. This module contained a set of genes involved not only in TCR signaling, but also Th2 cell differentiation and homing. However, rather than being organized in distinct pathways these genes were highly interconnected. This suggests that complex cellular behavior like Th2 cell homing and activation is based on the co-regulation of genes from different pathways by common TF such as GATA3.

Validation studies showed that the module could be replicated in an independent material. Moreover, since corticosteroids are effective in treating allergic disease we tested if this treatment would also *reverse* the expression levels of the genes that had changed after allergen-challenge. Indeed, this was true for a significant proportion of the module genes. Taken together these findings lent strong support for our hypothesis that a relevant module could be found and validated in allergic disease.

We also hypothesized that the module could be used to identify a novel disease gene. Indeed, eleven of the 23 module genes had not been previously described in allergy, but may have important roles. For example, *SH2D1A* that promotes Th2 activation [20] increased in expression following allergen-challenge. However, while there are several options for functional studies of individual genes, it is less clear how to select one for such studies. Recent work indicates that genes with many interactions are functionally more important than those with few interactions [1,2]. We therefore reasoned that since disease genes tend to interact the novel gene that had most interactions with known disease genes would be most likely to be pathogenic. Analysis of the connectivity of the novel genes led to the identification of *IL7R*, whose relevance was supported by converging observations; *IL7R*'s interactions included two signature genes for Th1 and Th2 activation, namely *IFNG* and *IL4R* [21]; the expression levels of *IL7R* showed highly significant decreases in the allergen-challenged $CD4^+$ cells which were reversed by treatment with corticosteroids. Such expression changes are characteristic for genes that inhibit allergic inflammation [22,23]. These observations led to the hypothesis that *IL7R* inhibited Th2 cell activation. To test this hypothesis we treated allergen-challenged cells with IL-7. However, this resulted in increased rather than decreased IL-4 production, indicating that IL-7 activated Th2 cells. We therefore hypothesized that IL-7 stimulation would result in a relatively stronger activation of Th1 than Th2 cells. Indeed, IL-7 did cause an increased production of the Th1 cytokine IFN-γ, which significantly decreased the ratio between IL-4/IFN-γ. Thus, by inducing a relatively higher Th1 than Th2 activation, IL-7 signaling has a Th1 like effect, which is consistent with an inhibitory role of IL-7R in allergic inflammation. This has not been previously described, but increased expression of *IL7R* has been shown in Th1-like disorders like multiple sclerosis [24]. This suggests that this gene may have therapeutic potential in inflammatory diseases that depend on an altered balance between Th1 and Th2 cells.

The identification of *IL7R* in this study also supports the second hypothesis, namely that a module-based approach can be applied to find novel genes. It is also of note that although there is limited information about gene interactions, analyzing the number of interactions with known disease genes is a useful complementary method to prioritize between novel module genes for validation studies.

## Conclusion

The identification of novel genes by high-throughput studies of complex diseases is complicated by the large number of potential genes. Since disease genes tend to interact, one solution is to arrange them in modules based on co-expression data and known gene interactions. In this study we validated the hypothesis that such a module could be a) found and validated in allergic disease and b) used to find and validate one ore more novel disease genes. We propose that the analytical principles may be generally applicable to complex diseases.

## Methods

### *Subjects analyzed with DNA microarrays*

For the integrated analysis a total of 36 patient samples were analyzed with 40 DNA microarrays. A total of 27 control samples were analyzed with 31 DNA microarrays. These samples were used for steps 1–3 in the results. In addition, for step 4, coexpression studies were performed in T helper cells from 73 atopic patients and 19 non-atopic controls. These materials have been described in detail elsewhere [7-9,25]. The validation studies of hypothetical module genes in steps 5–7 were performed in $CD4^+$ cells from 19 patients, a) with and b) without allergen-challenge as well as c) with challenge and treatment with cortisone. Samples from each individual were analyzed with one microarray, unless otherwise stated. All gene expression microarray data, except those from the asthmatic patients, are derived from studies performed by the authors of this study. The materials can be summarized as follows:

#### *Seasonal allergic rhinitis (SAR)*

1) $CD4^+$ T cells from 22 patients with SAR were examined before and after stimulation with grass pollen extract. $CD4^+$ T cells will subsequently referred to as T helper cells. The cells were obtained outside of the pollen season. 2) Nasal fluid cells from 18 patients before the pollen season sampled before allergen challenge as well as one and six

hours after allergen challenge. In addition nasal fluid cells were obtained from symptomatic patients during season. The nasal fluid cells from each time point were pooled and analyzed with one DNA microarray GeneChip at each time point. 3) Skin biopsies from ten patients sampled one hour after subcutaneous allergen challenge 4) Nasal polyps from two patients with SAR analyzed before and after treatment with glucocorticoids. Each polyp was analyzed with duplicate microarrays on each occasion.

#### *Asthma*

T helper cells from seven patients with mild asthma and five patients with severe asthma.

#### *Eczema*

Skin biopsies from ten patients with atopic eczema who fulfilled the criteria of Williams.

#### *Controls*

1) T helper cells from four healthy controls. 2) Nasal fluid cells from twenty healthy controls. The nasal fluid cells were pooled and analyzed with one DNA microarray GeneChip. 3) Nasal polyps from two patients without allergy, before and after treatment with glucocorticoids. Each sample was analyzed with duplicate DNA microarrays. 4) Skin biopsies from ten healthy controls. 5) Allergen-challenged skin biopsies from ten healthy controls (these individuals were not the same as in study 4). The raw array data are available at Gene Expression Omnibus http://www.ncbi.nlm.nih.gov/geo, pending processing, accession numbers GSE6012, GDS266 and GSE473.

### *Allergen challenge of peripheral blood mononuclear cells*

Peripheral blood mononuclear cells were prepared and stimulated with grass pollen extract with or without hydrocortisone as described [7]. For gene expression studies, T helper cells were enriched from the allergen-challenged PBMC using anti-CD4-coated paramagnetic microbeads and a MACS (magnetic cell sorter) system according to the instructions of the manufacturer (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany). cRNA was extracted from 200 ng total RNA using Ambion's Illumina RNA TotalPrep Amplification kit (Ambion, Inc., U.S.A.). In vitro transcription (IVT) reaction and cRNA biotinylation was performed overnight (14 h). The RNA/cRNA concentrations where checked using Nanodrop ND-1000 before and after the amplifications. cRNA quality was controlled by BioRad's Experion electrophoresis station (Bio-Rad Laboratories, Inc., CA, U.S.A.).

### *Microarray hybridization and data analysis*

All gene expression profiling studies except the validation data in allergen-challenged $CD4^+$ cells were performed using DNA microarrays measuring the expression of 22,283 genes and variants (HuGe U133A GeneChip, Affymetrix, Santa Clara, CA) as previously described [2]. The validation studies of allergen-challenged $CD4^+$ cells from patients with SAR were performed DNA microarrays measuring the expression of 47293 genes (Illumina, San Diego, CA). The analyses were performed according to manufacturer's instructions and the MIAME guidelines. Identification of disease-associated genes was performed as described in the additional file 1.

### *Network-based analysis of disease-associated genes*

The Ingenuity Pathways Analysis application (IPA) was used to organize the disease-associated genes into networks of interaction and to find modules of functionally related genes that correspond to pathways as previously described [7]. A detailed description is given in the additional file 1.

### *Identification of a shared framework of transcription factor binding sites for the T cell receptor pathway*

A shared framework of transcription factor binding sites (TFBS) for genes of the T cell receptor pathway was identified using a set of *in silico* genomics tools (see additional file 1).

### *Animal Experiments*

An exploration of *ITK*-deficient mice in a validated mouse model of allergic airway inflammation was used to assess the involvement of ITK in manifesting characteristic pathogenic features of an allergic inflammation. Tissue eosinophilia was selected as key parameter since this feature is strongly associated with allergy. Furthermore, the development of an airway eosinophilia in the present model has been demonstrated to be strictly dependent on a functional Th2 response [26] (see additional file 1).

### *Analysis of the effect of IL-7 on Th1/Th2 polarization*

PBMCs from 5 patients with seasonal allergic rhinitis were purified as described above. $2.5 \times 10^6$ cells in 2 mL culture medium (cells/mL) were stimulated with 100 μg/mL grass or pollen extract in the presence or absence of 10 ng/mL rhIL-7 (R&D systems Ltd, UK). The culture supernatants were collected after 7 days. IL-5, IL-13 and IFNγ cytokines were analyzed in the supernatants using human interleukin Quantikin ELISA kits from R&D systems according to the manufacturer's protocol. IL-4 Quantikin HS kit was used to for quantitative determination of human IL-4 concentrations in cell culture supernatants.

### *Statistical analysis*

Paired comparisons were performed using the Wilcoxon signed-rank test, while unpaired ones were performed using the Mann-Whitney U-test. The Spearman rank correlation test was used to study correlations.

## Abbreviations

SAR: Seasonal Allergic Rhinitis; SNP: Single Nucleotide Polymorphism; TCR: T Cell Receptor; Th2: T helper cell type 2; TF: Transcription Factor; TFBS: Transcription Factor Binding Site.

## Authors' contributions

RM carried out the molecular studies and drafted the manuscript. BA, MHZ and JE carried out the immunological analyses. ML and AP performed the computational analyses. LOC participated in the design of the study. MB conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

## Additional material

### Additional file 1

***The document describes additional methods, analysis data of IL-5 and IL-13 in allergen-challenged CD4+ cells with and without IL-7 and additional figure 1 which describes the Pathway analysis of disease-associated genes.***

Click here for file

[http://www.biomedcentral.com/content/supplementary/1752-0509-3-19-S1.doc]

## Acknowledgements

We thank Dr. Bernward Klocke for valuable help with the comparative promotor analysis, Dr. Tomas Werner for expert comments on the manuscript, and Dr Gilles Guillot for statistical advice. Dr. Jörgen Nedergaard Larsen of ALK-Abello gave valuable advice on allergen extracts. This work was funded by grants from the Swedish Research Council, the EU project ComplexDis under the Sixth Framework Programme, the U.S. National Science Foundation under grant CCR-0311500, and the U.S. National Institutes of Health under grants 1-P01-DA-015027-01, 5-U01-AA-013512-02 and 1-R01-MH-074460-01.

## References

1. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL: **The human disease network.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104(21):**8685-8690.
2. Feldman I, Rzhetsky A, Vitkup D: **Network properties of genes harboring inherited disease mutations.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105(11):**4323-4328.
3. Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B, *et al.*: **Network modeling links breast cancer susceptibility and centrosome dysfunction.** *Nature genetics* 2007, **39(11):**1338-1349.
4. Segal E, Friedman N, Kaminski N, Regev A, Koller D: **From signatures to models: understanding cancer using microarrays.** *Nature genetics* 2005, **37(Suppl):**S38-45.
5. Bochner BS, Busse WW: **Allergy and asthma.** *The Journal of allergy and clinical immunology* 2005, **115(5):**953-959.
6. Akbari O, Faul JL, Hoyte EG, Berry GJ, Wahlstrom J, Kronenberg M, DeKruyff RH, Umetsu DT: **CD4+ invariant T-cell-receptor+ natural killer T cells in bronchial asthma.** *The New England journal of medicine* 2006, **354(11):**1117-1129.
7. Benson M, Langston MA, Adner M, Andersson B, Torinssson-Naluai A, Cardell LO: **A network-based analysis of the late-phase reaction of the skin.** *The Journal of allergy and clinical immunology* 2006, **118(1):**220-225.
8. Hansel NN, Hilmer SC, Georas SN, Cope LM, Guo J, Irizarry RA, Diette GB: **Oligonucleotide-microarray analysis of peripheral-blood lymphocytes in severe asthma.** *The Journal of laboratory and clinical medicine* 2005, **145(5):**263-274.
9. Olsson M, Broberg A, Jernas M, Carlsson L, Rudemo M, Suurkula M, Svensson PA, Benson M: **Increased expression of aquaporin 3 in atopic eczema.** *Allergy* 2006, **61(9):**1132-1137.
10. Wills-Karp M, Koehl J: **New insights into the role of the complement pathway in allergy and asthma.** *Current allergy and asthma reports* 2005, **5(5):**362-369.
11. Zimmermann N, King NE, Laporte J, Yang M, Mishra A, Pope SM, Muntel EE, Witte DP, Pegg AA, Foster PS, *et al.*: **Dissection of experimental asthma with DNA microarray analysis identifies arginase in asthma pathogenesis.** *The Journal of clinical investigation* 2003, **111(12):**1863-1874.
12. Abu-Khzam FN, Langston MA, Shanbhag P, Symons CT: **Scalable Parallel Algorithms for FPT Problems.** *Algorithmica* 2006, **45(3):**269-284.
13. Baldwin NE, Chesler EJ, Kirov S, Langston MA, Snoddy JR, Williams RW, Zhang B: **Computational, integrative, and comparative methods for the elucidation of genetic coexpression networks.** *Journal of biomedicine & biotechnology* 2005, **2005(2):**172-180.
14. Zhang Y, Abu-Khzam FN, Baldwin NE, Chesler EJ, Langston MA, Samatova NF: **Genome-Scale computational approaches to memory-intensive applications in systems biology.** *Proceedings of the 2005 ACM/IEEE conference on Supercomputing: November 12–18 2005, Washington, DC, USA* 2005:12-25.
15. Weil R, Schwamborn K, Alcover A, Bessia C, Di Bartolo V, Israel A: **Induction of the NF-kappaB cascade by recruitment of the scaffold molecule NEMO to the T cell receptor.** *Immunity* 2003, **18(1):**13-26.
16. Mayall TP, Sheridan PL, Montminy MR, Jones KA: **Distinct roles for P-CREB and LEF-1 in TCR alpha enhancer assembly and activation on chromatin templates in vitro.** *Genes & development* 1997, **11(7):**887-899.
17. Ling P, Meyer CF, Redmond LP, Shui JW, Davis B, Rich RR, Hu MC, Wange RL, Tan TH: **Involvement of hematopoietic progenitor kinase 1 in T cell receptor signaling.** *The Journal of biological chemistry* 2001, **276(22):**18908-18914.
18. Graves PE, Siroux V, Guerra S, Klimecki WT, Martinez FD: **Association of atopy and eczema with polymorphisms in T-cell immunoglobulin domain and mucin domain-IL-2-inducible T-cell kinase gene cluster in chromosome 5 q 33.** *The Journal of allergy and clinical immunology* 2005, **116(3):**650-656.
19. Rautajoki KJ, Kylaniemi MK, Raghav SK, Rao K, Lahesmaa R: **An insight into molecular mechanisms of human T helper cell differentiation.** *Annals of medicine* 2008, **40(5):**322-335.
20. Wu C, Nguyen KB, Pien GC, Wang N, Gullo C, Howie D, Sosa MR, Edwards MJ, Borrow P, Satoskar AR, *et al.*: **SAP controls T cell responses to virus and terminal differentiation of TH2 cells.** *Nature immunology* 2001, **2(5):**410-414.
21. Wenzel SE, Balzar S, Ampleford E, Hawkins GA, Busse WW, Calhoun WJ, Castro M, Chung KF, Erzurum S, Gaston B, *et al.*: **IL4R alpha mutations are associated with asthma exacerbations and mast cell/IgE expression.** *American journal of respiratory and critical care medicine* 2007, **175(6):**570-576.
22. Benson M, Carlsson L, Adner M, Jernas M, Rudemo M, Sjogren A, Svensson PA, Uddman R, Cardell LO: **Gene profiling reveals increased expression of uteroglobin and other anti-inflammatory genes in glucocorticoid-treated nasal polyps.** *The Journal of allergy and clinical immunology* 2004, **113(6):**1137-1143.
23. Benson M, Jansson L, Adner M, Luts A, Uddman R, Cardell LO: **Gene profiling reveals decreased expression of uteroglobin and other anti-inflammatory genes in nasal fluid cells from patients with intermittent allergic rhinitis.** *Clin Exp Allergy* 2005, **35(4):**473-478.
24. Booth DR, Arthur AT, Teutsch SM, Bye C, Rubio J, Armati PJ, Pollard JD, Heard RN, Stewart GJ: **Gene expression and genotyping studies implicate the interleukin 7 receptor in the pathogenesis of primary progressive multiple sclerosis.** *Journal of molecular medicine (Berlin, Germany)* 2005, **83(10):**822-830.
25. Benson M, Carlsson L, Guillot G, Jernas M, Langston MA, Rudemo M, Andersson B: **A network-based analysis of allergen-challenged**

**CD4+ T cells from patients with allergic rhinitis.** *Genes and immunity* 2006, **7(6):**514-521.

26. Foster PS, Mould AW, Yang M, Mackenzie J, Mattes J, Hogan SP, Mahalingam S, McKenzie AN, Rothenberg ME, Young IG, *et al.*: **Elemental signals regulating eosinophil accumulation in the lung.** *Immunological reviews* 2001, **179:**173-181.

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

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

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

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

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral

**ADDITIONAL FILE 1**

## METHODS

### Microarray Data Analysis

Control and patient data were handled separately. For each, the Affymetrix Micorarray Suite MAS5.0 was used to convert raw scanned images into intensity calls. Probe set values were averaged, log transformed and z-scored. Genes were eliminated from consideration if, under all conditions, they were either weakly expressed or their expression levels did not significantly change. Pearson's correlation coefficients were computed and used to populate symmetric correlation matrices with values in the range [-1, 1]. Negative and positive correlations are equally meaningful when measuring putative co-regulation, and so the absolute value of each correlation coefficient was taken. Thresholds were then used to transform matrices into correlation graphs. Vertices were joined by an edge if and only if the corresponding genes were correlated at or above the appropriate threshold. For each threshold considered, maximum clique computations were performed using codes synthesized from vertex cover duality and the theory of fixed parameter tractability [1]. For quality monitoring, Affymetrix control probe sets were retained. Their frequency was found to be in accordance with chance. To capture biological significance and balance edge densities, a threshold of .92 was selected for control data while .90 was chosen for patient data. Maximal cliques were extracted using the methodology and prioritized by gene ontology. Only genes that passed three differential analysis screens were selected for detailed scrutiny. First, individual transcripts had to be "differentially expressed," a well-known measure of significance. Next, pairs of transcripts had to be "differentially correlated" [2], which we define here as correlated at 0.85 or above in one treatment and at 0.25 or below in the other. Finally, genes had to be "differentially cliquified" [3] which we define in its simplest form as highly represented in the clique lists of one treatment and poorly represented in the other.

**Network-based analysis of gene expression microarray data**

The Ingenuity Pathway Analysis tool (IPA) was used to organize differentially expressed genes into networks of interacting genes and to find modules of functionally related genes that correspond to pathways. The IPAconsists of a global network that is based on reviewing and extracting mammalian gene interactions from review of more than 200,000 articles in 275 peer reviewed journals. The findings for the IPA were manually curated and assigned to ontologic classes by content and modeling experts. Interactions were also inferred algorithmically by combining functional and structural information about proteins. Every interaction can be referenced back to a MedLine ID within the application. Identification of networks formed by the differentially expressed genes was done in a stepwise manner:

1. Differentially expressed genes are scored on the basis of the number of specific interactions found within the IPA global network. Focus genes were identified as those having direct or indirect interactions with other genes in the IPA database.

2. The specificity of connections for each focus gene was calculated by the percentage of its connections to other differentially expressed genes. A network was then constructed, starting with the gene with the highest specificity of connections. Each network contained a maximum of 35 genes. This process was repeated for the remaining genes to construct new networks. As the number of focus genes is reduced as a result of their inclusion in previous networks, genes from the global network not in the original list but with significant specific connectivity to the growing network can be included into those new networks.

3. A statistical algorithm based on the Fischer exact test was used to rank the networks. Identification of pathways was based on using canonical pathways in the IPA as templates for the differentially expressed genes. Again using the Fischer exact test, the resulting pathways were ranked according to their correlation with the template in the IPA.

**Identification of a shared framework of transcription factor binding sites for the T cell receptor pathway**

A shared framework of transcription factor binding sites (TFBSs) for genes of the T cell receptor pathway was identified using a set of *in silico* genomics tools, i.e. Gene2Promoter, DiAlign, FrameWorker, and ModelInspector (Genomatix, Munich, Germany). Evolutionarily conserved frameworks containing 2 TFBSs were identified and searched to isolate a TCR pathway gene of importance for allergic inflammation. Based on experimental studies, *ITK* was chosen as such a gene. The initial set of conserved frameworks was subsequently used as a seed to find and refine matching frameworks among the other pathway genes. The initial, conserved frameworks were identified in proximal promoter regions that were defined as 500 nt upstream of and 100 nt downstream from the transcription start site (TSS). The sequences were retrieved using Gene2Promoter from three species (*H. sapiens* , *M. musculus* , and *R. norvegicus* ). Frameworks of different combinations of 2 TFBS conserved in TFBS distance and orientation were identified using FrameWorker. The default settings for position weight matrices were used. The conserved frameworks were then compared with the promoter regions of all genes of the TCR pathway to find a subset of pathway genes that shared frameworks using ModelInspector. Promoters from genes sharing frameworks were then extracted by Gene2promoter and subjected to FrameWorker to identify a generalized framework for this subset. The relevance of this framework was tested by examining the expression levels of the corresponding TF in allergen-challenged T lymphocytes that had been analyzed by DNA microarrays as described above.

**Animal experiments**

Briefly, wild type mice (C57/Bl6) and age-matched mice that carry targeted and $ITK^{-/-}$ alleles were actively immunized against ovalbumin (OVA, grade III, Sigma, St Louis, MO) at day 1

by an intraperitoneal injection (7.5 μg OVA + 1.5 mg $AlOH_3$). At 21-28 days after i.p., immunization animals were challenged daily with aerosolized allergen (30 minutes 1% OVA in saline). At 24 h after the last challenge and nasal and lung tissues were collected for histological analysis. Eosinophils were detected in cryo sections by histochemical staining of cyanide-resistant eosinophil peroxidase (EPO) and Normasky optics. Eosinophils were counted around the bronchi and large pulmonary vessels in a circular tissue region spanning from the epithelial/endothelial basement membrane to a tissue depth of 250 μm. Periodic Acid Schiff (PAS) staining of mucin granules was used to study the numbers and distribution of airway goblet cells. Briefly, the number of PAS-positive cells were counted along the bronchial epithelium and correlated to the length of the epithelial lining (which was determined using digital image analysis). All quantifications were performed in a blinded manner. The local ethics committee, Lund/Malmö, Sweden approved all animal procedures.

## RESULTS

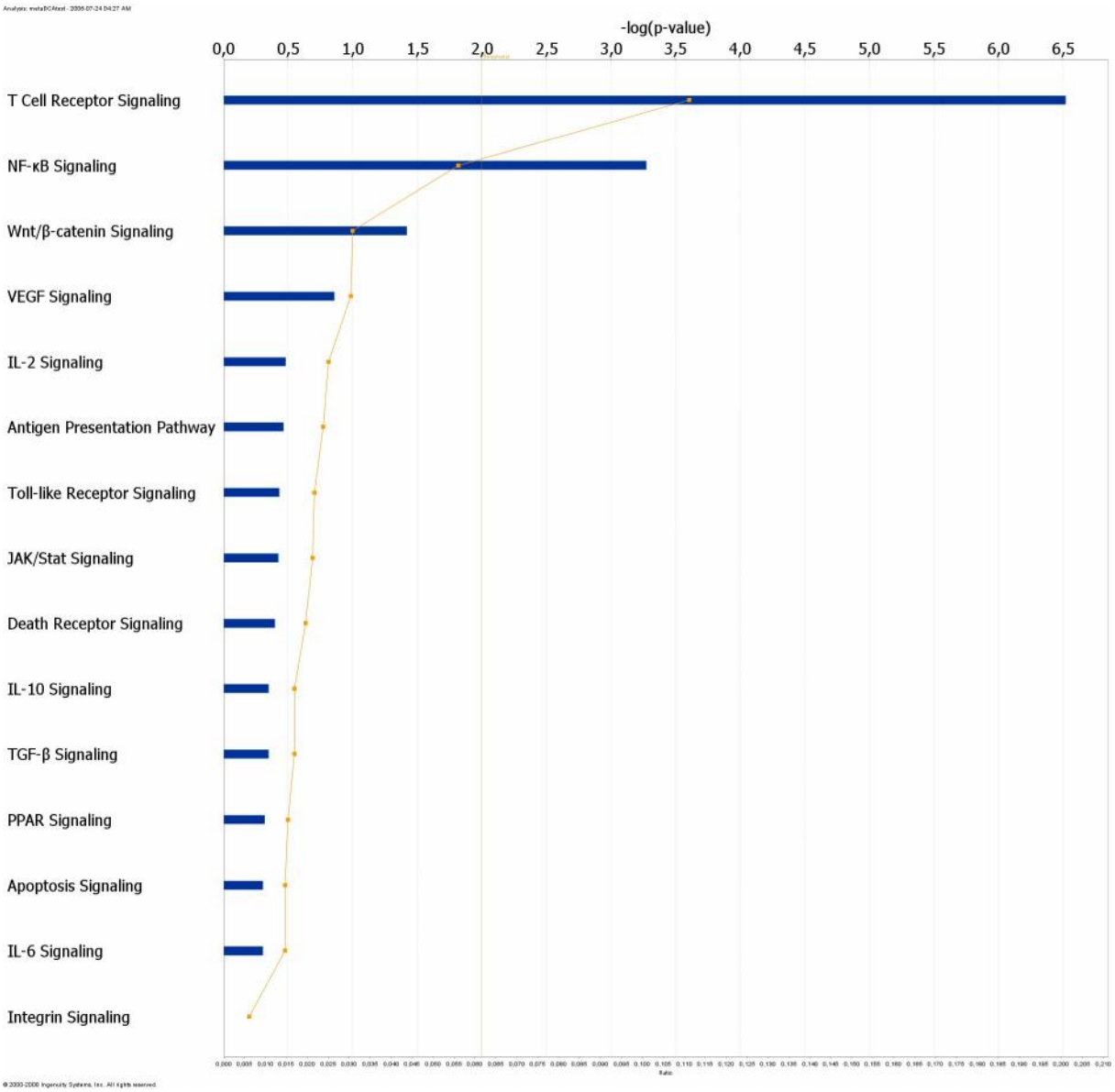
### Analysis of IL-5 and IL-13 in allergen-challenged $CD4^+$ cells with and without IL-7 stimulation

The IL-5 level in allergen-challenged $CD4^+$ cells (mean ± SEM) was 117 ± 47 compared to cells stimulated with allergen and IL-7, 326 ± 85 pg/mL, $p<0.05$. The corresponding figures for IL-13 were 134 ± 41 compared to 582 ±1 44 pg/mL ($p<0.05$).

## REFERENCES


1. Downey RG, Fellows MR: *Parameterized Complexity*. New York: Springer; 1999.
2. Voy BH, Scharff JA, Perkins AD, Saxton AM, Borate B, Chesler EJ, Branstetter LK, Langston MA: **Extracting gene networks for low-dose radiation using graph theoretical algorithms**. *PLoS computational biology* 2006, **2**(7):e89.
3. Lansgton MA, Perkins AD, Saxton AM, Scharff JA, Voy BH: **Innovative Computational Methods for Transcriptomic Data Analysis Proceedings**. In *ACM Symposium on Applied Computing: 2006; Dijon, France*; 2006.

**FIGURE 1**




## Legend to figure

**SI Figure 1.** Pathway analysis of disease-associated genes. The Fischer's exact test was used to rank the pathways statistically. The yellow line indicates the significance threshold of $p<0.01$.