Jutta Renkonen1,2
Sakari Joenväärä1,2
Ville Parviainen1,2
Pirkko Mattila1,2
Risto Renkonen1,2

1Transplantation Laboratory and Infection Biology Research Program, Haartman Institute, University of Helsinki, Helsinki; 2HUSLAB, Helsinki University Central Hospital, Helsinki, Finland

Background: Asthma is a chronic inflammatory disease of the airways with a complex genetic background. In this study, we carried out a meta-analysis of single nucleotide polymorphisms (SNPs) thought to be associated with asthma.

Methods: The literature (PubMed) was searched for SNPs within genes relevant in asthma. The SNP-modified genes were converted to corresponding proteins, and their protein–protein interactions were searched from six different databases. This interaction network was analyzed using annotated vocabularies (ontologies), such as the Gene Ontology and Nature pathway interaction databases.

Results: In total, 127 genes with SNPs related to asthma were found in the literature. The corresponding proteins were then entered into a large protein–protein interaction network with the help of various databases. Ninety-six SNP-related proteins had more than one interacting protein each, and a network containing 309 proteins and 644 connections was generated. This network was significantly enriched with a gene ontology entitled “protein binding” and several of its daughter categories, including receptor binding and cytokine binding, when compared with the background human proteome. In the detailed analysis, the chemokine network, including eight proteins and 13 toll-like receptors, were shown to interact with each other. Of great interest are the nonsynonymous SNPs which code for an alternative amino acid sequence of proteins and, of the toll-like receptor network, TLR1, TLR4, TLR5, TLR6, TLR10, IL4R, and IL13 are among these.

Conclusions: Protein binding, toll-like receptors, and chemokines dominated in the asthma-related protein interaction network. Systems level analysis of allergy-related mutations can provide new insights into the pathogenetic mechanisms of disease.

Keywords: asthma, network, pathway pathogenesis, single nucleotide polymorphisms

Introduction

Asthma is a chronic inflammatory disease of the airways characterized by infiltration and activation of inflammatory cells and by structural changes, including subepithelial fibrosis, smooth muscle cell hypertrophy/hyperplasia, epithelial cell metaplasia, and angiogenesis. These structural changes are believed to correlate with the severity of asthma and to some extent with the development of progressive lung function deterioration. The mechanism underlying airway angiogenesis in asthma and its precise clinical relevance has not yet been completely elucidated.1

Asthma may best be described as a loosely defined syndrome characterized by respiratory symptoms, airways narrowing, and inflammation. Asthma is a common pulmonary condition that involves heightened bronchial hyperresponsiveness and
reversible bronchoconstriction, together with acute-on-chronic inflammation that leads to airways remodeling. The most common causes predisposing for asthma include viral upper respiratory tract infections, cigarette smoke, cold temperatures, allergies, pets, and exercise. Symptoms of asthma include wheezing, intercostal and supraclavicular retraction, cough (worse at night), shortness of breath, chest pain, exercise intolerance, and limitation of daily activities, which should alert physicians to a diagnosis of possible asthma or an asthma exacerbation.2,3

Allergic asthma is characterized by a specific pattern of inflammatory attributes driven by IgE-dependent triggering of resident tissue mast cells and characterized by the influx of basophils and eosinophils in inflamed airways. The interaction between inflammatory cells and structural cells in asthmatic airways is complex. Several cytokines and growth factors released during allergic airway inflammation and remodeling are responsible for increasing basal levels of vascular endothelial growth factor in fibroblasts and smooth muscle cells.1,4,5

In spite of its great burden on public health care, our knowledge of the etiologic mechanisms underlying asthma, both genetic and environmental, is still very limited. One of the most promising approaches to expand further our understanding of the disease mechanisms involved is identification of the genetic variation that contributes to the risk of developing asthma.6

In recent years, research has mainly focused on detecting the genetic variations that predispose the individual to asthma. Three basic types of genetic study have been undertaken, ie, genetic linkage analysis, searches for focused candidate genes, and the modern genome-wide association studies performed with single nucleotide polymorphism (SNP) chips. Extensive epidemiologic studies have made little progress in determining the individual’s susceptibility to asthma. The molecular genetic studies of asthma offer the prospect of defining this susceptibility at a genetic level, and allow more precise studies on the etiology of asthma to be undertaken.7,9

Family studies using linkage methodologies conducted to date have not been very successful in identifying the genetic determinants of this complex disease.10 The revolution in genotyping technology with high-throughput methods now allows genotyping of greater numbers of SNPs in large cohort genome-wide association studies. Most of the genes uncovered during recent years with the genome-wide approach are novel, and were not even considered in the old candidate gene studies. Asthma is an example of a complex disease where several common susceptibility alleles affect the disease risk in varying combinations, but in a manner such that each gene contributes only a minor impact.11 The downstream biologic effects of the majority of these genes and their proteins are still unknown. Expression studies of these genes and proteins could allow us to uncover some of their effects.12

If some of the unexplained heritability in genome-wide association studies was due to interactions then, rather than discovering interactions per se, one goal might be to use these interactions in order to discover novel genes that act synergistically with other factors without having demonstrable marginal effects.13

Materials and methods

Literature search

A literature search in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) was carried out in September 2009 with the search phrase “asthma and SNP”. The list obtained was then manually annotated, and a list of SNP modifications in asthma was collected. Even though this is probably not a complete list, it gives a very good idea of the found and proposed SNP modifications linked to asthma. The list of SNPs claimed to be associated with asthma (Table 1) was used for further analyses.

Protein–protein interaction networks

We used a web-based protein interaction network analysis platform (PINA), which integrates protein–protein interaction data from six databases. The Cytoscape757 program provides a network construction tool, which uses our protein–protein interaction data as the baseline. The SwissProt names without the tag “__HUMAN” are used throughout this study, if not otherwise stated.

The gene ontology categories provide a controlled vocabulary to describe the gene and the gene product attributes of any organism. The enrichment analysis of proteins in the various gene ontology (GO Slim) categories was carried out essentially as described elsewhere.14–16

When the analysis within these annotated categories was carried out, we searched for enriched categories. The basic question was: Is the category “toll-like receptors” or the key phrase “cytokine–cytokine receptor interaction” within the observed SNP-related network enriched when compared with the background set of the whole human proteome? If such a phenomenon was observed, it could suggest that these enriched categories and keywords play a role within the SNP-related protein network found in asthma patients.
Results

We first searched the literature published for 2000–2009 on the association of SNPs with asthma. The search produced 251 articles, from which a list of 127 genes and their corresponding proteins linked to asthma was compiled (Table 1). Next, we generated a protein–protein interaction network for these 127 proteins. The interacting protein partners for each of these proteins were searched with PINA, which integrates six different protein–protein interaction databanks.¹⁷

We first pulled down all the protein–protein interactions of these putatively asthma-related proteins, which resulted in a very large network with 1073 proteins (nodes) and 1421 connections (edges) between them. This large data set was imported to Cytoscape and the network is displayed in Figure 1A.¹⁸ In order to facilitate data mining within the protein–protein interaction network, we created a more stringent query yielding a smaller subset of the original large network. This new limited network contained...
interacting proteins which were bound to at least two other proteins identified to carry asthma-related SNP modifications in their corresponding genes. This data set contained only 309 proteins (96 of which were SNP-related proteins, 213 interacting proteins, and 644 connections between them, Figure 1B).

Because such a network is far too large to be analyzed visually in a meaningful manner, we decided to perform a gene ontology enrichment analysis. When the analysis within the gene ontology-molecular function categories was performed, several strongly enriched classes were observed, as shown in Table 2. “Protein binding” and several of its daughter categories, including receptor binding, cytokine binding, growth factor binding, interleukin binding, transcription factor binding, chemokine binding, and pattern binding, were among the most significantly enriched categories. Furthermore, protein kinase activity, including tyrosine and serine/threonine kinase activity, as well as endopeptidase activity, was significantly enriched within the asthma-related SNP-modified genes and their corresponding proteins.

To utilize this asthma-related network further, we searched for connections between the asthma-related SNP proteins only (Figure 2). The results showed that while 31 of the 96 proteins do not interact with any other SNP protein, 14 interact with each other, 20 form a pair, two triplets, one a quartet, and two larger networks contain several connected asthma-related SNP proteins. The chemokine network includes eight proteins and the toll-like receptor network 13 proteins, all shown to carry asthma-related SNP modifications among their corresponding genes. We expanded to the most interesting networks with PINA and were able to create two new networks in Cytoscape. The chemokine network (green proteins, Figure 3) shows eight asthma-related SNP proteins binding to each other. Furthermore, a great number of other chemokines and their receptors also show an interaction within this network (yellow proteins, Figure 3).

Likewise, the other new subnetwork displays 13 asthma-related SNP proteins interacting with each other (Figure 4). This toll-like receptor/cytokine network also contains a large group of novel proteins, including toll-like receptors and cytokines, as well as signal transduction molecules (yellow proteins, Figure 4). Such an enlarged network of interacting proteins could putatively be used to search for novel proteins having a crucial role in the development of asthma-related inflammatory reactions. Of great interest are the nonsynonymous SNPs, which code for an alternative amino acid sequence of proteins. The red proteins within Figure 4, ie, TLR1, TLR4, TLR5, TLR6, TLR10, and IL4R and IL13, are among these. This small subnetwork of toll-like receptor-related proteins and their 40 interacting proteins was further analyzed. This analysis showed that these 40 interacting proteins have already been reported to carry almost 1000 nonsynonymous SNPs coding for alternative protein sequences (Table 3).

Finally, we manually annotated the whole set of 309 proteins in Figure 1B. A thorough analysis showed that the
**Table 2** Gene ontology (GO) enrichment on the SNP-related proteins and their first binding partners

<table>
<thead>
<tr>
<th>GO-ID</th>
<th>P-value</th>
<th>x</th>
<th>nX</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5515</td>
<td>6.507E-69</td>
<td>256</td>
<td>7023</td>
<td>Protein binding</td>
</tr>
<tr>
<td>4871</td>
<td>4.902E-44</td>
<td>130</td>
<td>2127</td>
<td>Signal transducer activity</td>
</tr>
<tr>
<td>6950</td>
<td>8.143E-63</td>
<td>120</td>
<td>1258</td>
<td>Response to stress</td>
</tr>
<tr>
<td>19538</td>
<td>2.190E-15</td>
<td>98</td>
<td>2768</td>
<td>Protein metabolic process</td>
</tr>
<tr>
<td>4872</td>
<td>2.927E-20</td>
<td>84</td>
<td>1792</td>
<td>Receptor activity</td>
</tr>
<tr>
<td>5102</td>
<td>3.345E-40</td>
<td>77</td>
<td>765</td>
<td>Receptor binding</td>
</tr>
<tr>
<td>166</td>
<td>3.810E-3</td>
<td>50</td>
<td>2146</td>
<td>Nucleotide binding</td>
</tr>
<tr>
<td>43234</td>
<td>8.197E-5</td>
<td>48</td>
<td>1712</td>
<td>Protein complex</td>
</tr>
<tr>
<td>7267</td>
<td>3.551E-16</td>
<td>40</td>
<td>535</td>
<td>Cell–cell signaling</td>
</tr>
<tr>
<td>5509</td>
<td>1.156E-8</td>
<td>40</td>
<td>932</td>
<td>Calcium ion binding</td>
</tr>
<tr>
<td>16740</td>
<td>7.737E-3</td>
<td>40</td>
<td>1693</td>
<td>Transferrase activity</td>
</tr>
<tr>
<td>30528</td>
<td>5.762E-3</td>
<td>35</td>
<td>1405</td>
<td>Transcription regulator activity</td>
</tr>
<tr>
<td>16265</td>
<td>2.280E-12</td>
<td>33</td>
<td>485</td>
<td>Death</td>
</tr>
<tr>
<td>8219</td>
<td>2.280E-12</td>
<td>33</td>
<td>485</td>
<td>Cell death</td>
</tr>
<tr>
<td>16301</td>
<td>4.330E-7</td>
<td>33</td>
<td>788</td>
<td>Kinase activity</td>
</tr>
<tr>
<td>30154</td>
<td>6.420E-5</td>
<td>33</td>
<td>1001</td>
<td>Cell differentiation</td>
</tr>
<tr>
<td>50222</td>
<td>4.232E-9</td>
<td>31</td>
<td>580</td>
<td>Protein kinase activity</td>
</tr>
<tr>
<td>8233</td>
<td>6.971E-7</td>
<td>28</td>
<td>616</td>
<td>Peptidase activity</td>
</tr>
<tr>
<td>16023</td>
<td>3.454E-7</td>
<td>21</td>
<td>359</td>
<td>Cytoplasmic membrane-bounded vesicle</td>
</tr>
<tr>
<td>6629</td>
<td>1.118E-2</td>
<td>20</td>
<td>716</td>
<td>Lipid metabolic process</td>
</tr>
<tr>
<td>5768</td>
<td>2.862E-8</td>
<td>18</td>
<td>228</td>
<td>Endosomie</td>
</tr>
<tr>
<td>8283</td>
<td>1.827E-5</td>
<td>16</td>
<td>290</td>
<td>Cell proliferation</td>
</tr>
<tr>
<td>8092</td>
<td>1.081E-2</td>
<td>14</td>
<td>437</td>
<td>Cytoskeletal protein binding</td>
</tr>
<tr>
<td>30246</td>
<td>8.720E-4</td>
<td>13</td>
<td>292</td>
<td>Carbohydrate binding</td>
</tr>
<tr>
<td>8289</td>
<td>8.302E-3</td>
<td>13</td>
<td>300</td>
<td>Mol func lipid binding</td>
</tr>
<tr>
<td>9719</td>
<td>5.236E-6</td>
<td>11</td>
<td>125</td>
<td>Response to endogenous stimulus</td>
</tr>
<tr>
<td>3779</td>
<td>2.507E-2</td>
<td>10</td>
<td>305</td>
<td>Actin binding</td>
</tr>
<tr>
<td>16032</td>
<td>3.273E-5</td>
<td>7</td>
<td>57</td>
<td>Viral reproduction</td>
</tr>
</tbody>
</table>

Notes: Go-ID is the GO class, x represents of the number of observed proteins in our data set, nX is the number of proteins in the background comparison dataset.

The second group (TLR3, TLR7, TLR8, and TLR9) resides intracellularly and recognizes nucleic acids. The reason for

most common class of annotations for these proteins was “cytokine–cytokine receptors”. We have generated a network of these proteins in Figure 5 (green proteins) and enlarged the pathway by also including their interacting proteins (yellow proteins, Figure 5). A strong input of other chemokines and signal transduction proteins is also seen here.

**Discussion**

Asthma is a major burden for health care worldwide. Although the pathophysiology of asthma has been studied intensively during recent years, much more work needs to be done before it will be possible to prevent the onset of symptoms of asthma and to cure patients. A great number of genetic analyses have been conducted with asthma patients describing SNPs in the coding sequences of several proteins and intergenic, nontranslated regions close to protein coding sequences. During recent years, more than 100 candidate genes harboring these SNP modifications have been associated with bronchial asthma.

Our aim in the present work was to start a systems level analysis of the putative pathogenetic mechanisms involved in asthma. Databases, and especially their integrated and merged data warehouses, allow rapid and convenient access to the data publicly available. Furthermore, it is possible to integrate one’s own data on top of the publicly available data and thus enhance the level of information. In this study, starting with 127 SNP-modified genes converted to the corresponding proteins obtained from published data, we could identify a protein–protein interaction network of over 1000 proteins.

Instead of focusing our analysis only on one or a few altered genes or their corresponding proteins, we attempted to generate larger protein–protein interaction networks from our data. No single database alone can provide such a connected network; in order to understand the new systems levels of diseases, we decided to build up an integrated data warehouse combining information from several databases. We are now able to show how the 127 proteins of asthma patients were connected together in a putative network. One of the most prominent observations in this network was that it was strongly enriched with protein binding, signal transduction, and peptidase functions. The network contained several subnetworks enriched with toll-like receptors or chemokines.

The innate immune system responds to invading pathogens by activating a proinflammatory cascade aiming at eradicating the invading agents. Pattern recognition receptors are a crucial part of this innate immune reaction. A variety of intra- and extracellular pattern recognition receptors are known today, of which toll-like receptors are involved in the recognition of molecular structures specific for microbial pathogens. Two main categories of toll-like receptors exist, ie, cell surface receptors and receptors localized in the endosome. It is important to make this distinction because surface toll-like receptors bind molecules on the bacterial cell wall, such as bacterial lipopeptides (TLR2) or lipopolysaccharide (TLR4), whereas endosomal toll-like receptors that are activated by microbial nucleic acids are less readily accessible.

Research has shown that toll-like receptors can now be divided into two groups on the basis of their subcellular localization. The first group (TLR1, TLR2, TLR4, TLR5, and TLR6) is present on the surface of the cell, and recognizes lipid structures and, in the case of TLR5, the protein flagellin. The second group (TLR3, TLR7, TLR8, and TLR9) resides intracellularly and recognizes nucleic acids. The reason for
Figure 2 Sixty-five asthma-related single nucleotide polymorphism proteins showed interconnectivity; 14 interact with themselves, 20 form a pair, two triplets, one a quartet, and two larger networks contain several connected asthma-related single nucleotide polymorphism proteins. The chemokine network includes eight proteins and the toll-like receptor network 13 proteins which are all shown to carry asthma-related single nucleotide polymorphism modifications among their corresponding genes. Synonymous nodes are marked as green and nonsynonymous nodes are marked as red.

Figure 3 The chemokine network shows eight asthma-related single nucleotide polymorphism proteins (green nodes) binding to each other and their receptors (yellow nodes) found in the integrated protein interaction network analysis database.
The different localization of toll-like receptors may be that TLR1, TLR2, TLR4, TLR5, and TLR6 recognize markers on the surface of pathogens, while TLR3, TLR7, TLR8, and TLR9 recognize nucleic acids derived from the genome of viruses and bacteria. It has become increasingly apparent that the localization and traffic of toll-like receptors within the cell is an important mechanism whereby toll-like receptors sense their ligands. Importantly, the traffic of certain toll-like receptors during signaling can also prevent overactivation of the toll-like receptor signaling pathways.22

A small proportion of SNP mutations can cause the altered amino acid sequence in the corresponding protein directly. Genes that have previously been shown to have a SNP mutation leading to a change in the actual protein structure and which have also been linked with asthma, are presented in red in Figure 4 (TLR1, TLR4, TLR5, TLR6, TLR10, and IL4R and IL13).23–26 Genes colored green have been shown to be asthma-related SNPs, but these SNP mutations do not have any effect on protein structures (IL4, STAT6, CD14, RIPK2, TLR2, TLR9). There are altogether 54 protein coding genes in the same pathway, with the aforementioned asthma-related genes having a SNP mutation (Figure 4).

The aberrant activation of toll-like receptor pathways, on the other hand, has been implicated in various chronic and autoimmune diseases affecting the gastrointestinal tract, central nervous system, kidneys, skin, lungs, and joints, whereby both exogenous and endogenous ligands have been suggested to act as toll-like receptor activators. The finding that intracellular proteins or the products of protein cleavage can act as endogenous ligands for toll-like receptors supports the hypothesis that toll-like receptors are important in mediating the response not only to infections but also to stress, damage, and death of cells in general.27

New developments in the fields of allergy and immunology have yielded a variety of novel therapeutic approaches in recent years, resulting in more agents at the clinical trial stage as well. Among the therapeutic approaches are the toll-like receptor agonists, immunostimulatory
oligodeoxynucleotides, oral and parenteral cytokine blockers, and specific cytokine receptor antagonists. However, a much better understanding of the “big picture of this systems inflammatory disease” must still be obtained before more target therapeutic approaches can be designed.28

Compared with the latest reports in which only one gene at a time has been in focus when analyzing the pathogenetic mechanisms of multifactorial diseases like asthma, we have now focused on the entire set of 127 genes and their corresponding proteins. Of these 127 genes, 96 could be connected to a same gene-mRNA-protein and protein–protein interaction network, and were found to be enriched significantly with protein binding, signal transduction, and endopeptidase activities. Taken together, we showed in this study using our in silico analysis framework and the outside databases that we can increase the level of knowledge by performing systems level analyses of previously characterized genes carrying SNPs related to asthma.

Acknowledgments
The work was supported in part by research grants from the Academy of Finland, Sigrid Juselius Foundation, Helsinki University Funds, Helsinki University Central Hospital, and Tampere University Hospital Research Funds.
Disclosure
The authors report no conflicts of interest in this work.

References


