Citation for published version (APA):

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Search for New Candidate Genes Involved in the Comorbidity of Asthma and Hypertension Based on Automatic Analysis of Scientific Literature

1 Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia, E-mail: saik@bionet.nsc.ru
2 Research Institute of Medical Genetics, Tomsk NRMC, Tomsk, Russia. https://orcid.org/0000-0002-1103-3073.
3 Bogomoletz Institute of Physiology, National Academy of Science, Kiev, Ukraine
4 Bielefeld University, International Research Training Group “Computational Methods for the Analysis of the Diversity and Dynamics of Genomes”, Bielefeld, Germany
5 Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, Russia
6 Bielefeld University, Technical Faculty, AG Bioinformatics and Medical Informatics, Bielefeld, Germany

Abstract:
Comorbid states of diseases significantly complicate diagnosis and treatment. Molecular mechanisms of comorbid states of asthma and hypertension are still poorly understood. Prioritization is a way for identifying genes involved in complex phenotypic traits. Existing methods of prioritization consider genetic, expression and evolutionary data, molecular-genetic networks and other. In the case of molecular-genetic networks, as a rule, protein-protein interactions and KEGG networks are used. ANDSystem allows reconstructing associative gene networks, which include more than 20 types of interactions, including protein-protein interactions, expression regulation, transport, catalysis, etc. In this work, a set of genes has been prioritized to find genes potentially involved in asthma and hypertension comorbidity. The prioritization was carried out using well-known methods (ToppGene and Endeavor) and a cross-talk centrality criterion, calculated by analysis of associative gene networks from ANDSystem. The identified genes, including IL1A, CD40LG, STAT3, IL15, FAS, APP, TLR2, C3, IL13 and CXCL10, may be involved in the molecular mechanisms of comorbid asthma/hypertension. An analysis of the dynamics of the frequency of mentioning the most priority genes in scientific publications revealed that the top 100 priority genes are significantly enriched with genes with increased positive dynamics, which may be a positive sign for further studies of these genes.

Keywords: ANDSystem, associative gene network, asthma, comorbid disease, dynamics of interest of genes in publications, gene prioritization, hypertension

DOI: 10.1515/jib-2018-0054

Received: July 30, 2018; Revised: October 16, 2018; Accepted: October 31, 2018

1 Introduction

Asthma is a chronic inflammatory disease of the respiratory tract, the main characteristics of which are the hypersensitivity of the airways to various stimuli and reversible airflow obstruction [1]. More than 300 million patients worldwide are suffering from this disease [2]. Another widespread disease in the world is hypertension, which is characterized by a persistent increase in blood pressure from 140/90 mmHg. Art. and higher [3]. Annually about nine million people die due to complications associated with hypertension [4]. A number of studies have shown that asthma is related to an increase in the incidence of cardiovascular diseases [5], [6]. In this regard, the combination of asthma and arterial hypertension (comorbidity) in a large percentage of clinical cases is not surprising. Prevalence of hypertension in asthma patients is high with odds ratio = 1.66 [147], [188], $p$-value < 0.00001 [7]. This may not be a simple coincidence, because standard asthma mechanisms have been implicated in key processes of initiation of hypertension. Good examples of this are the establishment of the role of arachidonic acid-leukotriene B4 production in spontaneously hypertensive rat [8] or the value of Th17 and
IL17 in arterial hypertension [9]. Therapeutic control of allergic inflammation leads to better control of blood pressure [10]. The important role of the transcription factor STAT3, which is involved in signal transduction for a variety of cytokines and takes an active part in allergic inflammation [11], [12] and vascular remodeling [13], is well-known. On the basis on this information, it can be assumed that the imbalance between the pro-inflammatory and anti-inflammatory factors in the body creates the optimal situation for the realization of a hereditary predisposition to both asthma and arterial hypertension. In addition to the important role played by immune responses and control of inflammation in the pathogenesis of asthma and hypertension, it is also possible to assume the involvement of other mechanisms underlying the comorbidity of these diseases. For example, β-adrenoblockers and ACE inhibitors, which for a long time were contraindicated for patients with bronchial asthma due to possible bronchospasm, are widely used in the treatment of hypertension. Polymorphism of β-adrenoreceptor genes is associated with the risk of hypertension and bronchial asthma [14], [15]. It is assumed that mutations in the SLC26A4 gene can play an important role in the pathogenesis of bronchial asthma and hypertension, and therefore in the comorbidity of these diseases [16], [17], [18]. However, despite the accumulated data on the participation of specific genes simultaneously in asthma and hypertension, the mechanisms of interaction of these two diseases and the formation of their comorbid state are still poorly understood. Previously, we performed the prioritization of genes that are associated with asthma and hypertension simultaneously, to find the candidates which can potentially be involved in the mechanisms of comorbidity of these two diseases [19].

In the present work, a prioritization of an extended set of genes, which are associated either with asthma or with hypertension, was carried out. In addition, a new criterion for prioritization, based on the assessment of cross-talk centrality, previously used by us for the prioritization of neuronal apoptosis genes in the associative gene network of Parkinson’s disease [20], was introduced. When assessing the quality of prioritization, this criterion was of the best significance. Among the most significant genes were cytokines (CXCL10, CD40LG, IL1A, IL13 and IL15), toll-like receptor TLR2, death domain receptor FAS, transcription factor STAT3, complement factor C3 and amyloid beta precursor protein APP. An analysis of the overrepresentation of Gene Ontology biological processes for the top 100 highest-priority genes has shown the potential importance of the response to hypoxia, the inflammatory and immune responses and regulation of the apoptotic process in the mechanism of interaction of asthma and hypertension. In addition, we analyzed the dynamics of interest in genes by the frequency of their mention in scientific publications. It appeared that genes simultaneously associated with asthma and hypertension are enriched with genes that have a high positive dynamic of interest in them from researchers. This fact may indicate the involvement of these genes in many mechanisms of manifestation of various phenotypic traits that can determine the importance of their role in the development of comorbidity of diseases. An analysis of the most priority genes also showed significant enrichment with genes that have a high dynamic of interest in them. This may indicate their involvement in multiple phenotypic signs and, accordingly, in the comorbidity of asthma and hypertension. The genes identified as a result of prioritization can be used in the planning of experiments on genotyping and search for associations with the comorbid state of asthma and hypertension.

2 Methods

The sets of genes associated with asthma and hypertension were extracted from the ANDSystem knowledge base [21], [22]. The set of genes associated with asthma included 749 genes. The set of genes associated with hypertension included 706 genes. The number of genes simultaneously presented in the first and second sets was 202. A test set of genes, used for prioritization, was obtained by combining sets of asthma and hypertension genes, minus the 202 genes gained by the intersection of these two sets. The total number of genes in the test set was 1051.

The over-represented Gene Ontology biological processes were identified using the DAVID 6.8 service [23] with the following parameters: the organism – “Homo sapiens”, Gene_Ontology – “GOTERM_BP_DIRECT”.

The reconstruction of associative gene networks was carried out with ANDSystem tool [21], [22].

To prioritize the genes we used four criteria, which we previously discussed in the analysis of asthma, hypertension and Parkinson’s disease [19], [20]. Criteria 1 and 2 were based on the well-known methods of prioritization: ToppGene [24] (https://toppgene.cchmc.org/prioritization.jsp) and Endeavor [25] version 3.71 (https://endeavour.esat.kuleuven.be/), respectively. The ToppGene and Endeavor programs allow one to perform a ranking of a test set of genes by a training set of genes, according to specific criteria characterizing the proximity of genes from the test set to the genes from the training sample. The genetic information (co-localization in the genome), the functional properties of genes (participation in the same GO categories), etc.,
can be used by these programs as properties of the vertices of the graph of protein-protein networks. As a training set for each of these programs, a list of genes associated with asthma and hypertension, described above, was provided, while a list of 1051 genes, which is also described above, was used as a test set.

For the ToppGene in the Training parameters section, all features were selected. The "Rank" indicator of the ToppGene program was used to rank the genes. In the case of the Endeavor system, all settings were set by default. The ranking of genes was carried out according to the "p-value" indicator of the Endeavor program. The lowest rank received genes with the lowest values of "p-value", and the largest rank - genes with the highest values of "p-value".

Criterion 3 reflects the specificity of association of biological processes in which the analyzed gene participates with comorbidity mechanisms. Criterion 3 was calculated by formula \( N_1/N_2 \), where \( N_1 \) is a number of the Gene Ontology biological processes, overrepresented for a set of genes associated with asthma and hypertension; \( N_2 \) is a number of all Gene Ontology biological processes in which the analyzed gene was involved. Information indicating that the gene is involved in the Gene Ontology biological process was obtained from the AmiGO system [26]. Ranks for genes were determined by sorting the list of genes by descending the number of over-represented Gene Ontology biological processes. Thus, the lowest rank was attributed to the gene for which the number of the over-represented Gene Ontology biological processes was the highest.

To calculate criterion 4, an associative gene network which describes the interactions between the analyzed gene/protein and the genes/proteins from the training set was used. The network was automatically reconstructed by ANDSystem. Criterion 4 was based on the use of the cross-talk centrality (CTC), calculated using the "Intelligent Filtration" function of the "ANDVisio" program. Within this criterion, the CTC value was calculated separately for vertices in the gene network corresponding to the gene and its product (protein). Thus, for the gene, two indices (CTCgene and CTCprotein) were determined.

The CTCgene was calculated by the formula:

\[
CTC_{\text{gene}} = \frac{N_i}{M},
\]

where \( N_i \) - is a number of links of the i-th gene with members of the associative gene network of asthma/hypertension; \( M \) - is a number of vertices of the associative gene network of asthma/hypertension. CTCprotein for the i-th protein was determined the same way. The CTCgene and CTCprotein grades were determined by sorting the values in descending order, like in the case of criterion 3.

In the case when several genes had the same value for the criterion under which the ranking was conducted, the rank of these genes was calculated the similar way as the Spearman rank correlation coefficient [27]. Such genes were assigned the same rank, equal to the average arithmetic ranks of these genes, according to their position, in the sorted list of genes.

As the fifth criterion, the average value of the ranks obtained according to the criteria 1-4 was taken like it was done in [19]. For the reconstruction of the dynamics, the information on investigated genes from PubMed publications for the period from 1990 to 2017 were considered. For each year from this period, the number of publications in which the analyzed gene was mentioned was calculated. The marking of gene names in the texts of publications was carried out by using the ANDSystem. The trend reflecting the dynamics of the number of gene references in the articles was calculated using the Pearson correlation coefficient between the year and the total number of articles published this year in which the gene was mentioned. The enrichment of the gene set with genes with increased interest dynamics was calculated using the hypergeometric distribution realized in the hypergeom.sf function of the scipy.stats package of the Python programming language [28]. Genes with high dynamics of interest were genes for which the Pearson correlation coefficient exceeded 0.9.

3 Results and Discussion

To prioritize the genes for their potential simultaneous involvement in asthma and hypertension 749 genes of asthma and 706 genes of hypertension were considered. The training set of genes consisted of 202 genes, which were presented both in asthma and hypertension lists. Prioritization was carried out by using 4 criteria (see Online Appendix 1). The ranking of genes was carried out according to the value calculated as the mean for the ranks of the genes obtained using each of the 4 criteria. The top 10 of the most significant genes is listed in Table 1. Among them were cytokines (CXCL10, CD40LG, IL1A, IL13 and IL15), toll-like receptor TLR2, death domain receptor FAS, transcription factor STAT3, complement component C3 and amyloid beta precursor protein APP.
It is interesting to note that all these genes are genes from the group of asthma. At the same time, the ratio of asthma genes to the genes of hypertension in the top 100 of the most significant genes appeared to be 2:1.

**Table 1:** Top 10 of the most significant genes.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>GeneID</th>
<th>Criterion 1</th>
<th>Criterion 2</th>
<th>Criterion 3</th>
<th>Criterion 4 A Gene CTC</th>
<th>Criterion 4 B Protein CTC</th>
<th>Average of criteria 1–4</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL1A</td>
<td>3552</td>
<td>17</td>
<td>102</td>
<td>80.00</td>
<td>2</td>
<td>11</td>
<td>42.4</td>
</tr>
<tr>
<td>CD40LG</td>
<td>959</td>
<td>3</td>
<td>72</td>
<td>113.50</td>
<td>23</td>
<td>4</td>
<td>43.1</td>
</tr>
<tr>
<td>STAT3</td>
<td>6774</td>
<td>50</td>
<td>4</td>
<td>143.00</td>
<td>18</td>
<td>2</td>
<td>43.4</td>
</tr>
<tr>
<td>IL15</td>
<td>3600</td>
<td>72</td>
<td>129</td>
<td>38.00</td>
<td>13.5</td>
<td>36.5</td>
<td>57.8</td>
</tr>
<tr>
<td>FAS</td>
<td>355</td>
<td>35</td>
<td>18</td>
<td>52.00</td>
<td>177.5</td>
<td>20</td>
<td>60.5</td>
</tr>
<tr>
<td>APP</td>
<td>351</td>
<td>25</td>
<td>16</td>
<td>263.00</td>
<td>5</td>
<td>19</td>
<td>65.6</td>
</tr>
<tr>
<td>TLR2</td>
<td>7097</td>
<td>6</td>
<td>75</td>
<td>84.00</td>
<td>146</td>
<td>22</td>
<td>66.6</td>
</tr>
<tr>
<td>C3</td>
<td>718</td>
<td>49</td>
<td>67</td>
<td>84.00</td>
<td>49.5</td>
<td>88.5</td>
<td>67.6</td>
</tr>
<tr>
<td>IL13</td>
<td>3596</td>
<td>5</td>
<td>110</td>
<td>33.50</td>
<td>177.5</td>
<td>14.5</td>
<td>68.1</td>
</tr>
<tr>
<td>CXCL10</td>
<td>3627</td>
<td>7</td>
<td>92</td>
<td>86.00</td>
<td>55</td>
<td>101.5</td>
<td>68.3</td>
</tr>
</tbody>
</table>

Figure 1 shows that the top 10 genes are relatively well connected to each other inside the associative gene network, reconstructed with ANDSystem (see Online Appendix 2). In particular, the transcription factor STAT3 regulates CXCL10 [29] and Fas [30]. In turn, TLR2, IL-13 and IL-15 regulate the activity of STAT3 [31], [32], [33].

For better understanding which biological processes can underlie the molecular mechanisms of interaction between asthma and hypertension, an analysis of the overrepresentation of Gene Ontology biological processes on two sets of data was conducted: the training set of genes and the top 100 of the highest priority genes from the table in Online Appendix 1 were analyzed. For the set of genes associated with asthma and hypertension simultaneously, there were found 69 significantly over-represented biological processes ($p$-value < 0.05 with FDR correction). For the top 100 of the most priority genes, there were 42 such processes. Twenty-two biological processes were over-represented simultaneously for these two sets of genes (see Online Appendix 3). For example, the response to hypoxia had the best relevance for the training set of genes ($p$-value < $10^{-12}$). In the second set of genes, the $p$-value for this process was $<10^{-8}$, and the most significant process appeared to be the inflammatory response ($p$-value < $10^{-19}$). The intersection of over-represented processes was more than 50% of the total number of all overrepresented processes for a set of top 100 priority genes. Such a high value was partly related to the 3rd criterion of prioritization, which was based on considering the overrepresented biological processes. But since the ranking was performed by using the rank averaged over all 4 criteria, such intersection indirectly confirms the correctness of the prioritization.

It should be noted that the manual analysis of the full-text articles of PubMed Central showed that for genes from the top 10 there was additional information that was not represented in ANDSystem that speaks in favor of their possible connection with hypertension. For example, for the IL1A gene, it is known that in Chinese
Han population genetic variations in IL1A are associated with the risk of preeclampsia, which is characterized by hypertension and proteinuria [34]. In addition in Chinese children with type 1 diabetes the prevalence of hypertension in carriers of T allele in the IL1A gene was 4.2-fold higher than for C allele carriers with \( p \)-value < 0.001 [35]. The article [36] stated that the down-regulation of CD40LG might be related to kidney dysfunction in an early phase of hypertension, and [37] discussed that CD40LG was included in pulmonary hypertension gene subnetwork. Gene STAT3 regulate angiogenesis and vasoconstriction and is involved in the development of portal hypertension [38]. In patients with essential hypertension, IL-15 might be involved in the development of cardiovascular diseases [39]. For gene FAS it was shown that its expression was significantly increased in spontaneously hypertensive rats compared to Wistar-Kyoto rats [40]. In [41] APP was identified to undergo differential ubiquitination during pulmonary hypertension pathogenesis. Hypomethylation of the TLR2 gene increased the risk of essential hypertension in the Chinese population [42]. It was shown that C3 is highly expressed in mesenchymal tissues in spontaneously hypertensive rats (SHR) [43] and increased C3 induces salt-sensitive hypertension in SHR [44]. In [45] was discussed that IL-13 was involved in the induction of pulmonary hypertension. CXCL10 appeared to be involved in the pathogenesis of hypertension [46] and its level in serum was elevated in untreated essential hypertension [47].

One of the interesting characteristics of a gene may be the dynamics of increasing researcher’s interest in this gene, which may reflect the prospect of its study. It turned out that the genes from the training set, simultaneously associated with asthma and hypertension, are statistically significantly enriched with genes that have a high positive dynamic of their mention in the scientific publications of PubMed \( (p\text{-value} < 10^{-15}) \). The total number of genes with a positive growing trend \( (r > 0.9) \) was 872 of the 17,868 human genes, presented in the ANDS system. Among the 202 genes of the training set, it appeared to be 20% of such genes (42 genes). The HMOX1 gene, which encodes heme oxygenase 1, possessed the greatest positive dynamics of interest \( (r = 0.99) \).

It is interesting to note that the top 100 of most priority genes contained 25% of genes possessing increased positive dynamics of interest. This enrichment was also statistically significant \( (p\text{-value} < 10^{-11}) \). Analysis of the top 5 genes showed that among them STAT3 belongs to the group of 25% of genes possessing increased positive dynamics of interest \( (r = 0.96) \). The differences in the dynamics of interest for genes from the top 5 group are shown in Figure 2.

Figure 2: The dynamics of researcher’s interest for the top 5 most priority genes, according to scientific publications of PubMed.

Gene STAT3 encodes a protein that in response to cytokines and growth factors is phosphorylated and translocated to the nucleus where it starts to act as a transcription activator [48]. STAT3 is involved in many cellular processes including cell growth and apoptosis [49].
4 Concluding Remarks

It is well known that many diseases are comorbid, which significantly complicates their treatment, and also prevents their proper diagnosis. The study of the molecular mechanisms of comorbid states is an urgent task of the modern biomedicine. Asthma and hypertension are examples of widespread comorbid diseases. In this study, we used prioritization methods, including existing well-known methods (ToppGene and Endeavor), and an approach (cross-talk centrality), based on the analysis of associative gene networks from ANDSystem, to find genes that could potentially be involved in asthma and hypertension simultaneously. Cytokines (CXCL10, CD40LG, IL1A, IL13 and IL15), toll-like receptor TLR2, death domain receptor FAS, transcription factor STAT3, complement factor C3 and amyloid beta precursor protein (APP), appeared to be the most significant. In addition, an analysis of the dynamics of the frequency of mentioning genes in scientific publications was carried out. It turned out that the top 100 of the priority genes are significantly enriched by genes with increased positive dynamics of researcher’s interest in them. This may indicate a promise for further researches on these genes. A further study of the predicted genes, simultaneously involved in asthma and hypertension, can help to illuminate the mechanisms of the comorbid state of these two diseases.

Acknowledgements

Gene prioritization was carried out with the support of the Volkswagen Stiftung Trilateral Partnerships – Cooperation Projects between Scholars and Scientists from Ukraine, Russia, and Germany “In silico screening and experimental validation of new drug targets for the treatment of co-morbid multifactorial diseases” (90.335).

Development of methods and analysis of the dynamics of interest in the area of genomics and medical genetics on the basis of frequency of mentioning of genes in scientific publications was supported by Federal target program “Research and development in priority areas of development of Russia’s scientific and technological complex for 2014–2020”. Agreement on granting a subsidy as of October 23, 2017 No. 14.601.21.0015 between the Ministry of Education and Science of the Russian Federation and Tomsk NRMC on conducting research project titled: Development of a forecast for the implementation of the priority of scientific and technological development defined in paragraph 20c “personalized medicine, high-tech health care and health saving technologies, including the rational use of medicines (primarily antibacterial)” Scientific and technological development strategy of the Russian Federation. The unique identifier of the project is RFMEFI60117X0015. The state agreement identifier is 000000007417PE1002.

Conflict of interest statement: Authors state no conflict of interest. All authors have read the journal’s Publication ethics and publication malpractice statement available at the journal’s website and hereby confirm that they comply with all its parts applicable to the present scientific work.

References


Chun HJ, Bonnet S, Chan SY. Translational advances in the field of pulmonary hypertension. Translating MicroRNA biology in pulmonary hypertension.


Saik et al.


Supplementary Material: The online version of this article offers supplementary material (DOI: https://doi.org/10.1515/jib-2018-0054).