

RESEARCH ARTICLE

Open Access



Knowledge-based analysis of genetic associations of rheumatoid arthritis to inform studies searching for pleiotropic genes: a literature review and network analysis

Weiyang Zheng¹ and Shaoqi Rao^{1,2*}

Abstract

Introduction: Pleiotropy describes the genetic effect of a single gene on multiple phenotypic traits. Gene variants directly affect the normal processes of a series of physiological and biochemical reactions, and therefore cause a variety of diseases traits to be changed accordingly. Moreover, a shared genetic susceptibility mechanism may exist between different diseases. Therefore, shared genes, with pleiotropic effects, are important to understand the sharing pathogenesis and hence the mechanisms underlying comorbidity.

Methods: In this study, we proposed combining genome-wide association studies (GWAS) and public knowledge databases to search for potential pleiotropic genes associated with rheumatoid arthritis (RA) and eight other related diseases. Here, a GWAS-based network analysis is used to recognize risk genes significantly associated with RA. These RA risk genes are re-extracted as potential pleiotropic genes if they have been proved to be susceptible genes for at least one of eight other diseases in the OMIM or PubMed databases.

Results: In total, we extracted 116 potential functional pleiotropic genes for RA and eight other diseases, including five hub pleiotropic genes, *BTNL2*, *HLA-DRA*, *NOTCH4*, *TNXB*, and *C6orf10*, where *BTNL2*, *NOTCH4*, and *C6orf10* are novel pleiotropic genes identified by our analysis.

Conclusions: This study demonstrates that pleiotropy is a common property of genes associated with disease traits. Our results ascertained the shared genetic risk profiles that predisposed individuals to RA and other diseases, which could have implications for identification of molecular targets for drug development, and classification of diseases.

Introduction

Rheumatoid arthritis (RA) is one of the most common systemic autoimmune diseases, characterized by synovial inflammation and hyperplasia, autoantibody production, cartilage and bone destruction, and systemic features, including cardiovascular, pulmonary, psychological, and skeletal disorders [1]. Parikh-Patel et al. [2] reported that among RA patients in California, males had significantly higher risks of lung neoplasm, but a lower risk of

prostate neoplasm; while females had a significantly reduced risk for breast neoplasm. In addition, these diseases were race-specific and Hispanics had increased risks of leukemia and lung neoplasms. It is also recognized that some patients with systemic lupus erythematosus (SLE) will develop a symmetrical polyarthritis while others might not experience any arthritis at all [3]. Lewder [4] reckoned that 15–20 % of patients with psoriasis would develop arthropathies, which shared some common features with reactive arthritis. Several recent studies indicated that autoimmune disorders increase the likelihood for prostate neoplasm [5] and lung neoplasm [6, 7]. Khurana et al. [8] showed a significant positive association between RA and the development of

* Correspondence: raoshaq@gdmc.edu.cn

¹College of Biomedical Engineering, Capital Medical University, 10 Xitoutiao Youanmen Fengtai, Beijing 100069, People's Republic of China

²Institute of Medical Systems Biology and School of Public Health, Guangdong Medical College, 1 Xin Cheng Avenue, Songshan Lake, Dongguan 523808, Guangdong, People's Republic of China

lung neoplasm in the veteran population. Isomäki et al. [9] reported that the incidence of leukemia was significantly higher in patients with rheumatoid arthritis than in the general male population.

The chronic, debilitating, autoimmune nature of RA affects the patient directly or indirectly in almost all organ systems, from cardiovascular problems and infections to depression and gastrointestinal ulcers. On average, the established RA patient has two or more comorbid conditions [10]. Causes of death in RA patients are cardiovascular disease in 31 %, respiratory disease in 22 %, solid tumors in 20 %, cerebrovascular disease in 10 % and other reasons in 17 % [11, 12]. Thus, studies to examine the molecular bridge linking RA and numerous related disease (for examples, breast neoplasm, Alzheimer disease, diabetes mellitus, type 1 (T1D), prostate neoplasm, lung neoplasm, psoriasis, SLE and leukemia, investigated by the present study) are in high demand.

Genomic variation can influence disease susceptibility, disease progression, the risk of specific outcomes, or the individual's response to therapy. Most molecular genetic studies focus on the identification of genes involved in a single human disease. To date, approximately 1,800 genes have been identified that are mutated in human diseases [13]. However, there are limited attempts to study multiple diseases together to discover the molecular bridges between them. In fact, the molecular mechanisms for complex human diseases could be more sophisticated than what we are imaging. On the one hand, a single disease phenotype can be the result of mutations in many different genes, the so-called genetic heterogeneity, and on the other hand one gene also can affect several different phenotypes, namely genetic pleiotropy. In definition, pleiotropy describes the genetic effect of a single gene on multiple phenotypic traits, which occurs when a new mutation in the gene may have an effect on some or all traits simultaneously or may cause different pathological effects in complex human diseases [14–16].

Pleiotropy in diseases often occurs when an impaired gene generates pathological effects on some molecular-associated diseases simultaneously. For example, the gene *PTPN22* showed pleiotropic effects in multiple autoimmune diseases including T1D, SLE and RA [17]. Furthermore, one clear finding to emerge from the published genetic studies of autoimmunity was that different autoimmune diseases shared common susceptibility loci. The *HLA* region was well known for being associated with several autoimmune diseases including T1D, multiple sclerosis (MS), RA as well as others [18–20].

Recently, Chavali et al. [16] described a method for estimating the pleiotropic effects of human disease genes from network properties. Sivakumaran et al. [21] presented a systematic review of pleiotropy among single

nucleotide polymorphisms (SNPs) and genes and found that pleiotropic links between common complex diseases and traits occurred more often than expected. Goh et al. [13] performed a comprehensive analysis of Online Mendelian Inheritance in Man (OMIM) for human traits and found several hundred genes of pleiotropic nature. And this gene list was expected to grow fast with increasing applications of the large-scale genome-wide association studies (GWAS) approach [22].

The GWAS approach analyzes SNPs across the whole genome to robustly identify key inherited genetic variations that have critical but as yet largely uncharacteristic roles in development of human diseases. The success of GWAS has opened a wide new horizon for exploration and highlighted the complicated genomic architecture of disease susceptibility. However, there are few attempts to use the GWAS approach to explore the shared (pleiotropic) genetic factors, which may give clues to underlying etiological links between these diseases and pinpoint potential directions and practical recommendations for future research in this field.

In the present study, we provided an integrated approach to combine GWAS and public knowledge databases, seeking for pleiotropic genes that link RA to eight other complex diseases. Here, the GWAS approach was used to obtain single risk SNPs from the Wellcome Trust Case Control Consortium (WTCCC) dataset, and the acquired SNPs were mapped to genes to extract risk genes to RA. Such genes were defined as pleiotropic genes if they had been proved to be also the susceptible genes for at least one of eight other diseases by using public knowledge databases PubMed or OMIM. To extract the disease-related pleiotropic functional modules and hub genes, we further used an interactive tree-model and logistic regression model to construct a SNP-SNP network for RA. In total, we extracted 116 pleiotropic genes, including five hub pleiotropic genes: *BTNL2*, *HLA-DRA*, *NOTCH4*, *TNXB* and *C6orf10*, where *BTNL2*, *NOTCH4*, and *C6orf10* were novel ones identified by this study. This exploratory study provided promising clues to inform experimental studies searching for pleiotropic genes.

Materials and methods

Data source

Genome-wide SNP data were provided by WTCCC, which contains SNP genotypic data for seven common diseases and the shared control. The present study focused on the analysis of the GWAS data for RA, which included 1,860 cases and 2,938 controls. Prior to association analysis, some SNPs were excluded due to significant deviations from the Hardy-Weinberg equilibrium ($P < 5 \times 10^{-5}$), or having a high rate of missing data (genotyping efficiency <95 %), or minor allele frequency (MAF) <1 %.

Extracting risk SNPs associated with RA

First, genome-wide single-point SNP association analysis was performed by two methods, the interactive tree-model analysis and conventional logistic regression. Multiple testing corrections were performed, and the P values were adjusted by using the false discovery rate (FDR) method, with the family-wise error rate of $P < 0.05$. This analysis was implemented by using Helix-Tree software (see [23]).

Identifying pleiotropic genes for RA and other diseases

To mine high-risk genes linked with RA, SNPs were mapped to genes using the SNPnexus tool (see [24]). To extract pleiotropic actions of the identified genes for RA, we manually selected those genes that had been proved to be associated with at least one of eight other diseases by literature reviews of the PubMed or OMIM database.

Identifying hub SNPs

To understand the functional mechanisms involved in RA pathogenesis, we further analyzed genetic factors along with biological network information [25]. Significant SNP-SNP interactions (nominal $P < 0.01$) were detected by using the interactive tree-based method. Then, a logistic regression model was used to estimate their odds ratios and 95 % CIs (confidence intervals). All the significant SNP-SNP interaction pairs contributing to RA were used to construct a disease-specific SNP-SNP network for RA. In the network, the nodes represented SNPs, and the links between SNPs suggest their synergistic actions contributing to RA. As a network measure, connectivity (the number of links) was used to measure importance of a hub node. To obtain significant hub nodes, we assumed that in a random network, connectivity followed a Poisson distribution in a random network [26, 27]. We used the following formula to assess whether a node could be categorized as a hub node. Suppose that p was the probability of connecting any two nodes in a random network with n nodes, the probability of connectivity of equal or larger than t was as follows:

$$p(x \geq t) = 1 - p(x < t) = 1 - \sum_{k=0}^{t-1} \frac{\lambda^k e^{-\lambda}}{k!}, \quad (\lambda = nP_1, P_1 = m/C_n^2),$$

Where m is the number of interacting SNP pairs contained in the disease-specific SNP network. We considered a SNP with >5 connections ($P = 0.01$) in a random network as a rare event under the null hypothesis that n nodes (SNPs) were connected randomly. The probability of this rare event was taken as a threshold, and a node was considered a hub SNP when its P value was smaller than the threshold.

Mining hub pleiotropic genes

To mine the disease-related pleiotropic functional modules, the SNP network was first turned into a gene network by mapping SNPs to genes using the dbSNP database. We used the following rule for mapping SNPs onto genes, i.e., to see whether a SNP is located within a gene or the untranslated regions (UTRs) of this gene. SNPs that were mapped onto multiple genes were assigned to a single gene according to the following hierarchy: coding $>$ intronic $>$ 5' UTR $>$ 3' UTR $>$ 5' upstream $>$ 3' upstream. This strategy can avoid issues with a weighted inflation induced by genes having different numbers of SNPs [28]. All the significant gene-gene interaction was utilized to construct a gene network. In both SNP-SNP and gene-gene interaction assays, we chose to use a nominal level of $P < 0.01$ as the cutoff to identify important SNP pairs or gene pairs because no suitable methodologies for adjusting for multiple correlated tests of massive SNP pairs (or gene pairs) were available, which may have led to false positives. However, in the following step for identifying hub pleiotropic genes, we may alleviate this issue by using evidence from the network analysis and literature reviews.

Results

Identifying pleiotropic genes for RA and related diseases

In single-point genome-wide association analysis of 459,236 autosomal SNP markers, 2046 SNPs were found significantly associated with RA. These SNPs were mapped to 502 RA risk genes (See the Table S1 in Additional file 1 and Table S2 in Additional file 2).

To confirm their pleiotropic effects, we tried to find evidence from OMIM or PubMed by literature reviews to support that these genes were associated with at least one of eight other diseases (i.e., breast neoplasm, Alzheimer disease, T1D, prostate neoplasm, lung neoplasm, psoriasis, SLE and leukemia). We extracted a total of 116 pleiotropic genes (corresponding to 205 SNPs), see Additional file 3 for details. Their pleiotropic associations with multiple diseases were summarized in Additional file 4.

In these 116 genes, *HLA-DRB1* was discovered first and is still by far the strongest genetic link to RA. It was estimated that this *HLA* locus contributed about 30 % of the overall familial RA risk [29]; *PTPN22*, a lymphocyte-specific nonreceptor tyrosine phosphatase involved in regulation of activation threshold of lymphocytes, was the second most contributed genetic link to RA. *PTPN22* also represented a strong susceptibility gene, which was shared by many autoimmune diseases such as T1D, psoriasis, and SLE. This suggested the presence of common genetic factors that predisposed to multiple autoimmunity diseases. Besides, pleiotropic genes *HLA-DRB1*, *PTPN22*, *AFF3* and *IL2RA* were all included in a

T cell activation pathway. In this pathway, some predisposing T cell repertoire selection, antigen presentation, or alteration in peptide affinity had a role in promoting autoreactive adaptive immune responses [1]. There were studies indicating that intracellular machinery was affected in T cells of RA patients, which might alter the behavior of T cells during activation. Different therapeutic approaches may modulate the abnormal T cell functions [30]. Protein tyrosine phosphatases (*PTPN22*) were the critical regulators of T cell signal transduction. In conjunction with protein tyrosine kinases, protein tyrosine phosphatases regulated the reversible phosphorylation of tyrosine residues and thereby played important roles in many diverse aspects of T cell physiology. Abnormalities in tyrosine phosphorylation had turned out to be involved in the pathogenesis of numerous human diseases, from autoimmunity to cancer [31].

More recently, we occasionally found that one of the 116 genes, *TNFAIP3*, should be a susceptibility factor for RA in the northern Chinese Han population [32] and its pleiotropic association with SLE and RA in the Korean population were also reported [33]. *TNFAIP3* played a major role in downregulating TNF-induced nuclear factor kappa B (NF- κ B) activation. As is well known, acute inflammation is a part of the defense response; chronic inflammation can lead to cancer, diabetes, arthritis, Alzheimer's disease, pulmonary, and neurological diseases. Regulation of NF- κ B helped in explaining the linkages of inflammation and cancer at the molecular level. Transcription factor NF- κ B mainly regulated the expression of several proinflammatory gene products TNF and its superfamily members [34]. NF- κ B was constitutively active in most tumors and was induced by carcinogens, tumor

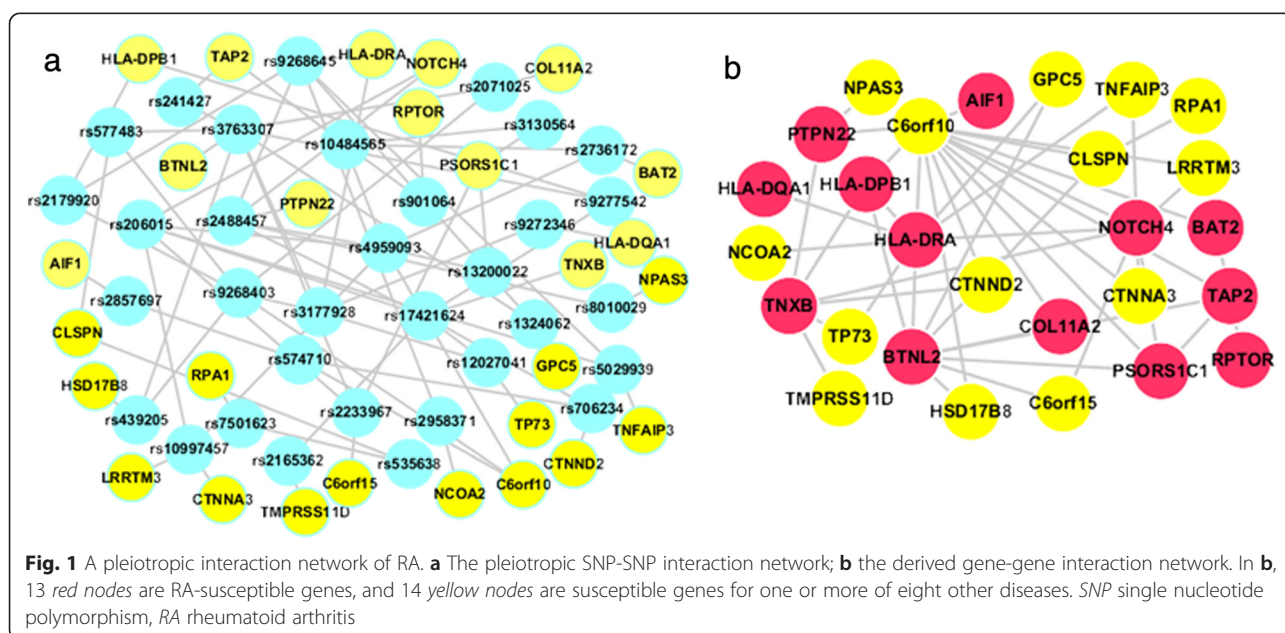
promoters, carcinogenic viral proteins, chemotherapeutic agents, and γ -irradiation. Hence, anti-inflammatory agents, which suppressed NF- κ B or NF- κ B-regulated products, should have a potential in both the prevention and treatment of cancer [35].

Identifying pleiotropic SNP-SNP interactions and hub genes for RA and related diseases

These 205 pleiotropic SNPs were subject to an epistasis analysis using the above mentioned two approaches. First, 127 pairs of SNP interactions ($P < 0.01$) were detected by using classification tree-based analysis. Then, logistic regression models were used to compute the odds ratios and their 95 % CIs for these SNP-SNP interactions, confirming that 46 SNP pairs were significant interactions ($P < 0.01$) (see Additional file 5). A SNP-SNP interaction network was constructed as shown in Fig. 1. By turning into a gene-gene network, 27 pleiotropic genes were included in the network and five hub genes were extracted by using a Poisson-based test of connectivity: *C6orf10* (degree = 14, $P = 1.1 \times 10^{-11}$), *BTNL2* and *HLA-DRA* (degree = 9, $P = 1.3 \times 10^{-5}$), *NOTCH4* (degree = 8, $P = 2.3 \times 10^{-5}$), *TNXB* (degree = 6, $P = 8.6 \times 10^{-3}$).

HLA-DRA, as one of the genes contained in major histocompatibility complex (MHC) region, had been discussed in the previous section. Involvement of the MHC region in chromosome 6p21 was in no doubt for most autoimmune diseases [36].

NOTCH4 encoded a receptor protein that was involved in multiple cellular processes, such as cell differentiation, proliferation and apoptosis. OMIM recorded that *NOTCH4* was highly related to schizophrenia. However, we also found its association with RA, breast



neoplasm, and Alzheimer disease in PubMed. AlFadhli [37] was the first one reporting a statistically significant association between *NOTCH4* and RA. Furthermore, SNP hla58, located in the intron of *NOTCH4* gene was observed to be in strong linkage disequilibrium with *HLA-DRB1* allele in RA [38]. Also, *NOTCH4* was a potential new therapeutic target for triple-negative breast neoplasm [39]. An mRNA transcript encoding the intracellular domain of *NOTCH4* was detected in two human breast neoplasm lines [40]. Furthermore, *NOTCH4* may play important roles in the pathogenesis of Alzheimer disease in the Japanese population and in the United Kingdom population [41, 42].

Published reports for *TNXB* gene were relatively limited. *TNXB* gene, located near the MHC class III region (6p21.3), had anti-adhesive effects, and functions in matrix maturation in connective tissues [43]. *TNXB* gene encoded an extracellular matrix protein, tenascin XB, which regulated collagen synthesis and deposition [44, 45]. In OMIM, *TNXB* was a pleiotropic gene of Ehlers-Danlos syndrome and Vesicoureteral reflux diseases. Furthermore, Rupert et al. [46] reported an unequal crossover between *RCCX* modules of human *MHC* leading to the presence of a *CYP21B* gene and a tenascin *TNXB/TNXA-RP2* recombinant in patients with juvenile RA. In other research, *BTNL2*, *C6orf10*, *NOTCH4*, *TAP2*, and *TNXB* were all identified as the novel RA-associated genes [47]. Not only that, a study by Kamatani and colleagues [45] identified rs3130342 in the 50 flanking region of the *TNXB* as a possible candidate gene susceptible to SLE in the Japanese population.

BTNL2 gene in OMIM was labeled as a sarcoidosis disease susceptibility gene. However, *BTNL2* (rs3817963) was reported as a new susceptibility locus of lung adenocarcinoma as well, the most common histologic type of lung neoplasm in the Japanese population [48]. As reported, patients with RA more often had lymphomas and lung tumors with the standardized incidence ratios of 2.1 and 1.6 respectively [49]. Functional variant rs2076530 of the *BTNL2* gene was identified conferring susceptibility to the autoimmune diseases T1D, RA, and SLE (G allele was linked to T1D and RA, and the A allele associated with SLE) [50]. Furthermore, there were new results implicating *BTNL2* as a novel prostate neoplasm-related gene [51].

C6orf10 gene had not yet been included in OMIM. *C6orf10* lies between *NOTCH4* and *BTNL2*. However, *C6orf10* was a completely new pleiotropic gene that had not been reported before. In our analysis, 12 SNPs located within *C6orf10* (rs9268402, rs9268403, rs574710, rs4959093, rs3129932, rs3132959, rs9368716, rs2894249, rs3129934, rs3129933, rs910050, rs9268208) were significantly related to RA. *C6orf10* was one of five hub disease genes in our constructed pleiotropic genetic network.

There were 14 genes interacting with *C6orf10*, indicating its strong risk for multiple diseases. We provided the associated diseases with 14 genes (see Additional file 6) and exhibited a corresponding gene-disease network to predict the pleiotropic function of *C6orf10* (see Fig. 2).

We manually searched for literature in PubMed for *C6orf10* to explore its function. Table 1 summarizes the related diseases of *C6orf10* reported to date.

Formerly, *C6orf10* was only known as testis-specific basic protein, and for which no transcripts had been assigned to date. It had been known that expression of this gene in keratinocytes was induced by exposure to tumor necrosis factor alpha (TNF- α). TNF- α -directed therapy had turned out to be valuable in the treatment of patients with refractory RA [52]. TNF- α inhibitors had been a cornerstone in the treatment of several chronic inflammatory diseases. Among patients with RA and a history of breast neoplasm, those who started a TNF- α -inhibitor treatment did not experience more breast neoplasm recurrences than patients with RA treated otherwise. TNF- α inhibitors might impact the risk of cancer development, or modify the risk of recurrence of previous cancers [53].

As non-HLA genes, we hypothesized that *NOTCH4*, *TNXB*, *BTNL2*, and *C6orf10* all would be good candidates for further clinical and laboratory studies. We postulated that to achieve success in the study of RA (as well as other diseases), it was necessary to take into account the multivariate nature of these diseases.

Discussion

GWAS analysis, by searching the entire human genome for association, is a promising approach to unravel the genetic basis of complex genetic diseases. Analysis of omic SNPs permitted determining relationships between genotypes and phenotypes, identification of risk SNPs and genes related to disease [54]. In this study, we presented an integrative approach to combine network analysis of GWAS data with literature reviews in order to recognize pleiotropic genes linked to RA and related diseases. In total, we found 116 genes harboring variants associated with RA and eight other diseases. In order to verify these findings, we sought supportive evidence by reviewing two additional databases for disease genes (MalaCards and HuGENavigator).

MalaCards [55] is an integrated web database of human maladies and their annotations, which included 64 data sources. In MalaCards, using GeneCards Suite and keywords of eight other diseases (breast neoplasm, Alzheimer disease, T1D, prostate neoplasm, lung neoplasm, psoriasis, SLE and leukemia), we compared 116 pleiotropic genes identified in this study with the disease genes related to the diseases. In total, 87 of the 116 pleiotropic genes were found to be related to the eight

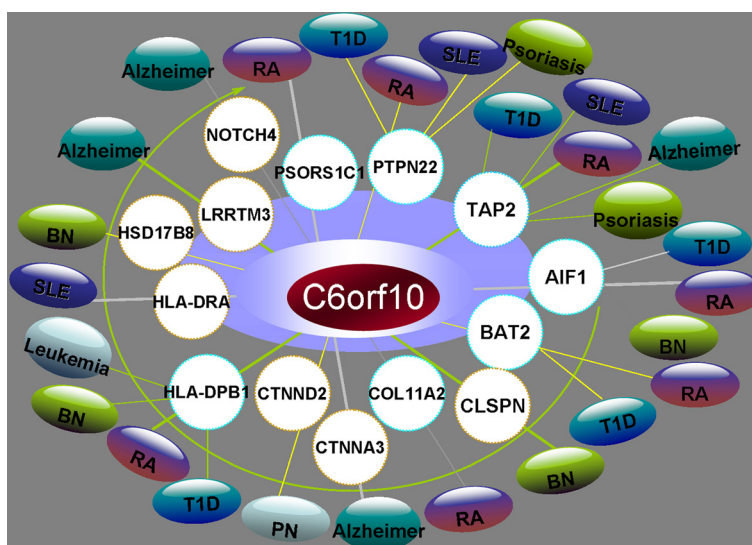


Fig. 2 The gene-disease network of *C6orf10* and its interaction genes

diseases. However, we found that some disease-related genes confirmed in the OMIM database, such as T1D-associated genes (*FOXP3, HNF1A, OAS1, ITPR3, PTPN22*) and SLE-associated genes (*BANK1*) were not included in MalaCards database. Thus, this verification using only MalaCards was incomplete and inadequate. So, we chose another database, HuGENavigator, for further comparison.

HuGENavigator [56] is an integrated, searchable knowledge base of genetic associations and human genome epidemiology. In HuGENavigator, we performed a similar analysis, and found that a total of 89 pleiotropic genes

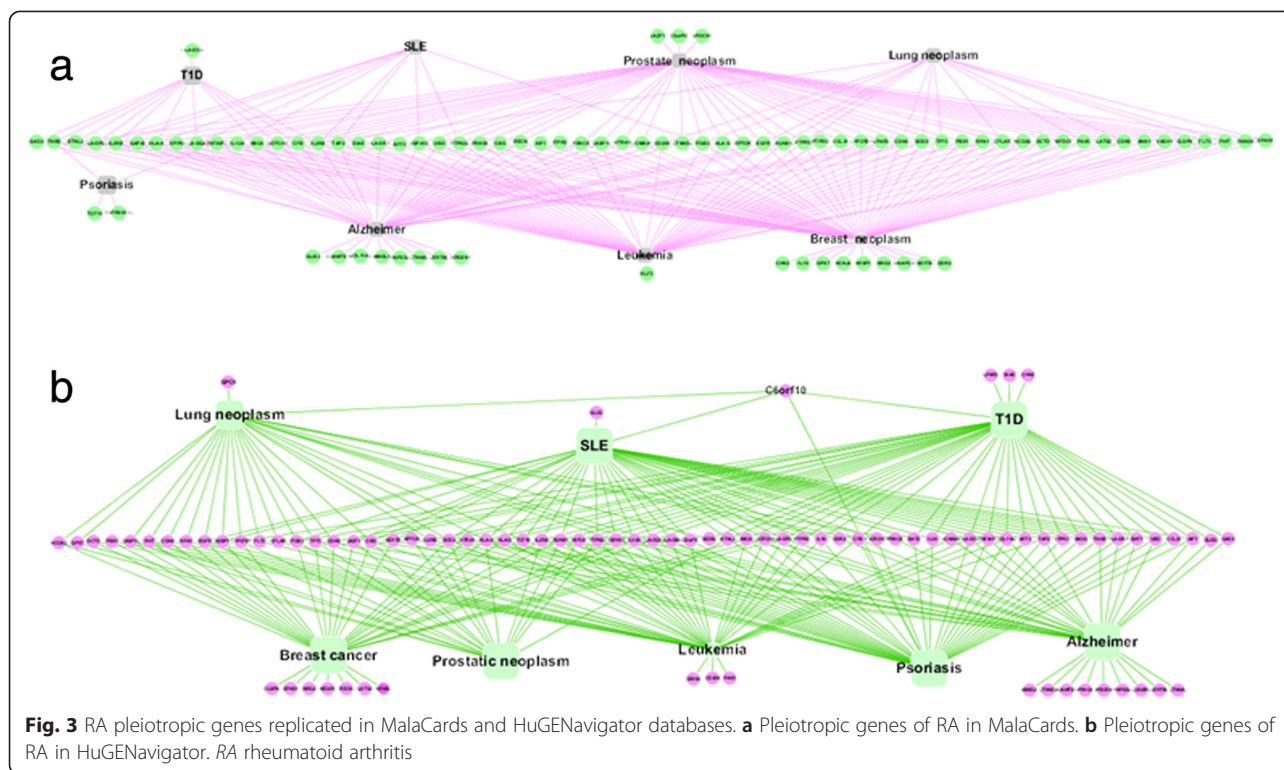
were replicated by HuGENavigator (i.e., also related with the eight diseases). The list of genes replicated in MalaCards and HuGENavigator is provided in Additional file 7. Moreover, we found that *C6orf10* was also proved a pleiotropic risk gene for SLE, T1D, psoriasis and lung neoplasm in HuGENavigator. For comprehensive results of this additional analysis, see Fig. 3.

Among these 116 pleiotropic genes, most of our extracted pleiotropic genes were found to link with multiple diseases, while others are novel, requiring further studies (either experimental or computational) to verify

Table 1 The list of gene *C6orf10* related diseases, as a novel RA pleiotropic gene identified by GWAS analysis

| SNPs | Population | Disease relevance | PubMed ID |
|---|-----------------------|-------------------|--|
| rs1265777, rs574710, rs539703, rs2894249 | British | RA | 20018019 |
| rs6910071, rs9391858, rs10484560, rs6910071 | North American | RA | 20018025, 20018077, 20018006, 21592391 |
| rs2395148 | Canadian, U.S. | PBC | 19458352 |
| rs2073048, rs28732201 | European, Chinese Han | Psoriasis | 19680446, 20692714 |
| rs926070, rs2073044 | Cleveland | Leukemia | 20460636 |
| rs3130320 | | SLE | 21408207 |
| rs3117103, rs7746019 | European | SLE | 19851445 |
| rs7775397 | | NL | 22031281 |
| rs7758128 | European | Vitiligo | 21326295 |
| rs485774 | U.S., Canada | LOAD | 22245343 |
| rs498422 | Chinese Han | NOA | 22541561 |
| rs2050190 | Utah | CAD | 22703881 |
| DG65398, D652889 | Caucasians | T1D | 19143811 |
| rs9391858 | British | T1D | 20549515 |
| rs3129934 | European | MS | 19010793 |

GWAS genome-wide association study, RA rheumatoid arthritis, PBC primary biliary cirrhosis, Leukemia T cell large granular lymphocyte leukemia, SLE systemic lupus erythematosus, NL neonatal lupus, Vitiligo generalized vitiligo, LOAD late-onset Alzheimer disease, NOA nonobstructive azoospermia, CAD coronary artery disease, T1D diabetes mellitus, type 1, MS multiple sclerosis



them prior to applying these findings to more practical settings.

We believed that examples of pleiotropy would accumulate over time, for it was already clear that pleiotropy was a common property of genes and SNPs associated with disease traits [21]. The clustering of autoimmune disorders within families and/or individuals also provided supportive evidence for these shared genetic factors [29].

Pleiotropic genes built up a bridge between RA and eight other diseases; they also would shed light on some studies of the inflammation-cancer link. Clinical studies suggested that persistent inflammation functions as a driving force in the journey to cancer. The possible mechanisms by which inflammation could contribute to carcinogenesis include induction of genomic instability, alterations in epigenetic events and subsequent inappropriate gene expression, enhanced proliferation of initiated cells and resistance to apoptosis [57]. Pleiotropic genes and their products should play an important role in this process. Also there was a study that clearly pointed to the importance of anti-inflammatory drugs in preventing the initiation and progression of both gastrointestinal and other solid organ cancers (including lung and prostate), and suggested that inflammation might be an underlying cause of cancer even in tumor types that had not been traditionally considered to originate within chronically inflamed tissues [58].

Despite that this study provided a pioneering approach to analyze large-scale SNP omic data and to fusing data from rich knowledge databases, we should recognize the limitations of this study in exploring pleiotropic mechanisms between complex diseases. First, because a single GWAS dataset was used for analyzing RA, we believe that only a small proportion of pleiotropic genes were identified, and perhaps a long list of genes with modest effects remains to be found in further independent studies. The omic SNP data, provided by WTCCC, are for Caucasians, whether these findings could be extended to other nationalities remains unclear. Furthermore, due to lack of information, this study did not model several important epidemiological covariates like gender, age and so on, which may have an impact on this pleiotropic analysis. Second, currently accumulated knowledge about genes involved in complex diseases, especially of their pleiotropic effects, was incomplete and fragmented, and part of this analysis that relied on knowledge mining suffers from this limitation. Finally, although we tried our best to control Type I errors in various steps toward identification of pleiotropic genes, whether the overall Type I error was well controlled remains unclear. In this sense, we view our analysis exploratory in nature.

In spite of the encouraging successes in finding pleiotropic genes for RA achieved by this study, further deeper studies of their detailed molecular etiology or links for these correlated diseases are certainly required. We

deeply believe that the pleiotropic genes identified by this study only represent a tip of the iceberg in the genetic architectures for complex diseases. Further studies, with more sophisticated designs and more involved multivariate analysis, are called for to directly test the hypothesis that a pleiotropic gene has a duplicated genetic effect on multiple disease phenotypes, followed by experimental or clinical validation.

Conclusions

Identification of pleiotropic genes is fundamental to reveal the underlying genetic links that connect multiple related diseases at the molecular level. In this study, we provided an integrative approach to combine network analysis of GWAS data with literature reviews in order to identify pleiotropic genes for complex diseases. Application to a GWAS dataset for RA identified a list of potential pleiotropic genes, which may be valuable clues for experimental studies to decipher the molecular mechanisms underlying these pleiotropisms.

Additional files

Additional file 1: The list of susceptible SNPs involved in RA by GWAS. (XLS 188 kb)

Additional file 2: The list of susceptible genes involved in RA by GWAS. (XLS 251 kb)

Additional file 3: The list of significant pleiotropic SNPs and genes that are involved in RA and one or more other diseases. (XLS 61 kb)

Additional file 4: Risk RA genes were proved to be susceptible to RA and eight other diseases respectively confirmed by the PubMed literature database. (XLS 35 kb)

Additional file 5: The list of pleiotropic SNP-SNP interactions that are involved in RA and one or more other diseases. (XLS 23 kb)

Additional file 6: The interaction genes with *C6orf10* in gene networks and their associated diseases. (XLS 21 kb)

Additional file 7: The list of RA pleiotropic genes replicated in MalaCards and HuGENavigator. (XLS 38 kb)

Abbreviations

BANK1: B-cell scaffold protein with ankyrin repeats 1; BTNL2: butyrophilin-like 2; C6orf10: chromosome 6 open reading frame 10; FOXP3: forkhead box P3; GWAS: genome-wide association study; HLA: human leukocyte antigen; HLA-DRA: major histocompatibility complex, class II, DR alpha; HNF1A: HNF1 homeobox A; ITPR3: inositol 1,4,5-trisphosphate receptor, type 3; MHC: major histocompatibility complex; MS: multiple sclerosis; NF- κ B: nuclear factor kappa B; NOTCH4: neurogenic locus notch homolog 4; OAS1: 2'-5'-oligoadenylate synthetase 1, 40/46 kDa; OMIM: Online Mendelian Inheritance in Man; PTPN22: tyrosine phosphatase nonreceptor 22; RA: rheumatoid arthritis; SLE: systemic lupus erythematosus; SNP: single nucleotide polymorphism; T1D: diabetes mellitus, type 1; TNFAIP3: tumor necrosis factor alpha-induced protein 3; TNF- α : tumor necrosis factor alpha; TNXB: tenascin XB; UTR: untranslated region; WTCCC: Wellcome Trust Case Control Consortium.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SR gave the initial instructions for the work. WZ analyzed the data. Both authors wrote and edited the manuscript. Both authors read and approved the final version of the manuscript.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (grant numbers 30830104, 31071166 and 81101641).

Received: 17 December 2014 Accepted: 13 July 2015

Published online: 08 August 2015

References

- McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. *N Engl J Med*. 2011;365:2205–19.
- Parikh-Patel A, White RH, Allen M, Cress R. Risk of cancer among rheumatoid arthritis patients in California. *Cancer Causes Control*. 2009;20:1001–10.
- Jessica A, Stanich JDC, Whittum-Hudson J, Hudson AP. Rheumatoid arthritis: disease or syndrome? *Open Access Rheumatol Res Rev*. 2009;1:179–92.
- Lewder JN. Diagnostic use of in the management HLA-B27 in the management of seronegative spondyloarthropathies: methods for detection. *Clin Immunol Newslett*. 1996;16:1–32.
- Eaton WW, Byrne M, Ewald H, Mors O, Chen CY, Agerbo E, et al. Association of schizophrenia and autoimmune diseases: linkage of Danish national registers. *Am J Psychiatry*. 2006;163:521–8.
- Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet*. 2008;40:310–5.
- Veeriah S, Taylor BS, Meng S, Fang F, Yilmaz E, Vivanco I, et al. Somatic mutations of the Parkinson's disease-associated gene *PARK2* in glioblastoma and other human malignancies. *Nat Genet*. 2010;42:77–82.
- Khurana R, Wolf R, Berney S, Caldito G, Hayat S, Berney SM. Risk of development of lung cancer is increased in patients with rheumatoid arthritis: a large case control study in US veterans. *J Rheumatol*. 2008;35:1704–8.
- Isomaki HA, Hakulinen T, Joutsenlahti U. Excess risk of lymphomas, leukemia and myeloma in patients with rheumatoid arthritis. *J Chronic Dis*. 1978;31:691–6.
- Michaud K, Wolfe F. Comorbidities in rheumatoid arthritis. *Best Pract Res Clin Rheumatol*. 2007;21:885–906.
- Young A, Koduri G, Batley M, Kulinskaya E, Gough A, Norton S, et al. Mortality in rheumatoid arthritis. Increased in the early course of disease, in ischaemic heart disease and in pulmonary fibrosis. *Rheumatology*. 2007;46:350–7.
- Tsuchiya Y, Takayanagi N, Sugiura H, Miyahara Y, Tokunaga D, Kawabata Y, et al. Lung diseases directly associated with rheumatoid arthritis and their relationship to outcome. *Eur Respir J*. 2011;37:1411–7.
- Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc Natl Acad Sci U S A*. 2007;104:8685–90.
- Barrenas F, Chavali S, Holme P, Mobini R, Benson M. Network properties of complex human disease genes identified through genome-wide association studies. *PLoS One*. 2009;4:e8090.
- Park J, Lee DS, Christakis NA, Barabasi AL. The impact of cellular networks on disease comorbidity. *Mol Syst Biol*. 2009;5:262.
- Chavali S, Barrenas F, Kanduri K, Benson M. Network properties of human disease genes with pleiotropic effects. *BMC Syst Biol*. 2010;4:78.
- Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, et al. Sarcoidosis is associated with a truncating splice site mutation in *BTNL2*. *Nat Genet*. 2005;37:357–64.
- Thorsby E, Lie BA. HLA associated genetic predisposition to autoimmune diseases: Genes involved and possible mechanisms. *Transpl Immunol*. 2005;14:175–82.
- Zanelli E, Breedveld FC, de Vries RR. HLA association with autoimmune disease: a failure to protect? *Rheumatology*. 2000;39:1060–6.
- Sirota M, Schaub MA, Batzoglou S, Robinson WH, Butte AJ. Autoimmune disease classification by inverse association with SNP alleles. *PLoS Genet*. 2009;5:e1000792.
- Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, et al. Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet*. 2011;89:607–18.
- Vineis P, Brennan P, Canzian F, Ioannidis JP, Matullo G, Ritchie M, et al. Expectations and challenges stemming from genome-wide association studies. *Mutagenesis*. 2008;23:439–44.
- Golden Helix GenomeBrowse® visualization tool (Version 2.x) Available from <http://goldenhelix.com/>.

24. Claude C, Arshad K, Lemoine NR, SNPnexus. A web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics*. 2009;25:655–61. Available at <http://snp-nexus.org/>.
25. Hua L, Zhou P, Liu H, Li L, Yang Z, Liu ZC. Mining susceptibility gene modules and disease risk genes from SNP data by combining network topological properties with support vector regression. *J Theor Biol*. 2011;289:225–36.
26. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 2004;5:101–13.
27. Jiang W, Li X, Rao S, Wang L, Du L, Li C, et al. Constructing disease-specific gene networks using pair-wise relevance metric: application to colon cancer identifies interleukin 8, desmin and enolase 1 as the central elements. *BMC Syst Biol*. 2008;2:72.
28. Torkamani A, Topol EJ, Schork NJ. Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics*. 2008;92:265–72.
29. Dieude P. Rheumatic diseases: environment and genetics. *Joint Bone Spine*. 2009;76:602–7.
30. Szalay B, Cseh A, Meszaros G, Kovacs L, Balog A, Vasarhelyi B. The impact of DMARD and anti-TNF therapy on functional characterization of short-term T-cell activation in patients with rheumatoid arthritis—a follow-up study. *PLoS One*. 2014;9:e104298.
31. Orozco G, Sanchez E, Gonzalez-Gay MA, Lopez-Nevot MA, Torres B, Caliz R, et al. Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Rheum*. 2005;52:219–24.
32. Zhang X, Li W, Zhang X, Zhao L, Zhang X, Jiang L, et al. Single nucleotide polymorphisms in TNFAIP3 were associated with the risks of rheumatoid arthritis in northern Chinese Han population. *BMC Med Genet*. 2014;15:56.
33. Kim SK, Choe JY, Bae J, Chae SC, Park DJ, Kwak SG, et al. TNFAIP3 gene polymorphisms associated with differential susceptibility to rheumatoid arthritis and systemic lupus erythematosus in the Korean population. *Rheumatology*. 2014;53:1009–13.
34. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet*. 2009;10:43–55.
35. Aggarwal BB, Shishodia S, Sandur SK, Pandey MK, Sethi G. Inflammation and cancer: how hot is the link? *Biochem Pharmacol*. 2006;72:1605–21.
36. Baranzini SE. The genetics of autoimmune diseases: a networked perspective. *Curr Opin Immunol*. 2009;21:596–605.
37. AlFadhli S, Nanda A. Genetic evidence for the involvement of NOTCH4 in rheumatoid arthritis and alopecia areata. *Immunol Lett*. 2013;150:130–3.
38. Kochi Y, Yamada R, Kobayashi K, Takahashi A, Suzuki A, Sekine A, et al. Analysis of single-nucleotide polymorphisms in Japanese rheumatoid arthritis patients shows additional susceptibility markers besides the classic shared epitope susceptibility sequences. *Arthritis Rheum*. 2004;50:63–71.
39. Nagamatsu I, Onishi H, Matsushita S, Kubo M, Kai M, Imaizumi A, et al. NOTCH4 is a potential therapeutic target for triple-negative breast cancer. *Anticancer Res*. 2014;34:69–80.
40. Zang S, Ji C, Qu X, Dong X, Ma D, Ye J, et al. A study on Notch signaling in human breast cancer. *Neoplasma*. 2007;54:304–10.
41. Lambert JC, Mann D, Harris J, Araria-Goumidi L, Chartier-Harlin MC, Cottel D, et al. Association study of Notch 4 polymorphisms with Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 2004;75:377–81.
42. Shibata N, Ohnuma T, Higashi S, Higashi M, Usui C, Ohkubo T, et al. Genetic association between Notch4 polymorphisms and Alzheimer's disease in the Japanese population. *J Gerontol A Biol Sci Med Sci*. 2007;62:350–1.
43. Ajayi OO, Adefenwa MA, Agaviezor BO, Ikeobi CO, Wheto M, Okpeku M, et al. A novel TaqI polymorphism in the coding region of the ovine TNXB gene in the MHC class III region: morphostructural and physiological influences. *Biochem Genet*. 2014;52:1–14.
44. Mao JR, Taylor G, Dean WB, Wagner DR, Afzal V, Lotz JC, et al. Tenascin-X deficiency mimics Ehlers-Danlos syndrome in mice through alteration of collagen deposition. *Nat Genet*. 2002;30:421–5.
45. Kamatani Y, Matsuda K, Ohishi T, Ohtsubo S, Yamazaki K, Iida A, et al. Identification of a significant association of a single nucleotide polymorphism in TNXB with systemic lupus erythematosus in a Japanese population. *J Hum Genet*. 2008;53:64–73.
46. Rupert KL, Rennebohm RM, Yu CY. An unequal crossover between the RCCX modules of the human MHC leading to the presence of a CYP21B gene and a tenascin TNXB/TNXA-RP2 recombinant between C4A and C4B genes in a patient with juvenile rheumatoid arthritis. *Exp Clin Immunogenet*. 1999;16:81–97.
47. Yang HC, Liang YJ, Chung CM, Chen JW, Pan WH. Genome-wide gene-based association study. *BMC Proc*. 2009;7:S135.
48. Shiraiishi K, Kunitoh H, Daigo Y, Takahashi A, Goto K, Sakamoto H, et al. A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. *Nat Genet*. 2012;44:900–3.
49. Love T, Solomon DH. The relationship between cancer and rheumatoid arthritis: still a large research agenda. *Arthritis Res Ther*. 2008;10:109.
50. Orozco G, Eerligh P, Sanchez E, Zhernakova S, Roep BO, Gonzalez-Gay MA, et al. Analysis of a functional BTNL2 polymorphism in type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus. *Hum Immunol*. 2005;66:1235–41.
51. Fitzgerald LM, Kumar A, Boyle EA, Zhang Y, McIntosh LM, Kolb S, et al. Germine missense variants in the BTNL2 gene are associated with prostate cancer susceptibility. *Canc Epidemiol Biomarkers Prev*. 2013;22:1520–8.
52. Morris JC, Janik JE, White JD, Fleisher TA, Brown M, Tsudo M, et al. Preclinical and phase I clinical trial of blockade of IL-15 using Mikbeta1 monoclonal antibody in T cell large granular lymphocyte leukemia. *Proc Natl Acad Sci U S A*. 2006;103:401–6.
53. Raaschou P, Frisell T, Askling J, for the ASG: TNF inhibitor therapy and risk of breast cancer recurrence in patients with rheumatoid arthritis: a nationwide cohort study. *Ann Rheum Dis*. 2014. 10.1136/annrheumdis-2014-205745.
54. Li C, Zhang G, Li X, Rao S, Gong B, Jiang W, et al. A systematic method for mapping multiple loci: an application to construct a genetic network for rheumatoid arthritis. *Gene*. 2008;408:104–11.
55. MalaCards: an integrated database of human maladies and their annotations, modeled on the architecture and richness of the popular GeneCards database of human genes. <http://www.malacards.org/>. Accessed on 1 June 2015.
56. HuGE Navigator (version 2.0): An integrated, searchable knowledge base of genetic associations and human genome epidemiology. <http://www.hugenavigator.net>. Accessed on 3 July 2015.
57. Kundu JK, Surh YJ. Inflammation: gearing the journey to cancer. *Mutat Res*. 2008;659:15–30.
58. Trinchieri G. Innate inflammation and cancer: is it time for cancer prevention? *F1000 Med Rep*. 2011;3:11.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

