

RESEARCH

Open Access

# Integrated genomic and prospective clinical studies show the importance of modular pleiotropy for disease susceptibility, diagnosis and treatment

Mika Gustafsson<sup>1\*</sup>, Måns Edström<sup>2</sup>, Danuta Gawel<sup>1</sup>, Colm E Nestor<sup>1</sup>, Hui Wang<sup>1</sup>, Huan Zhang<sup>1</sup>, Fredrik Barrenäs<sup>1</sup>, James Tojo<sup>3</sup>, Ingrid Kockum<sup>3</sup>, Tomas Olsson<sup>3</sup>, Jordi Serra-Musach<sup>4</sup>, Núria Bonifaci<sup>4</sup>, Miguel Angel Pujana<sup>4</sup>, Jan Ernerudh<sup>2</sup> and Mikael Benson<sup>1</sup>

## Abstract

**Background:** Translational research typically aims to identify and functionally validate individual, disease-specific genes. However, reaching this aim is complicated by the involvement of thousands of genes in common diseases, and that many of those genes are pleiotropic, that is, shared by several diseases.

**Methods:** We integrated genomic meta-analyses with prospective clinical studies to systematically investigate the pathogenic, diagnostic and therapeutic roles of pleiotropic genes. In a novel approach, we first used pathway analysis of all published genome-wide association studies (GWAS) to find a cell type common to many diseases.

**Results:** The analysis showed over-representation of the T helper cell differentiation pathway, which is expressed in T cells. This led us to focus on expression profiling of CD4<sup>+</sup> T cells from highly diverse inflammatory and malignant diseases. We found that pleiotropic genes were highly interconnected and formed a pleiotropic module, which was enriched for inflammatory, metabolic and proliferative pathways. The general relevance of this module was supported by highly significant enrichment of genetic variants identified by all GWAS and cancer studies, as well as known diagnostic and therapeutic targets. Prospective clinical studies of multiple sclerosis and allergy showed the importance of both pleiotropic and disease specific modules for clinical stratification.

**Conclusions:** In summary, this translational genomics study identified a pleiotropic module, which has key pathogenic, diagnostic and therapeutic roles.

## Background

Medical research typically focuses on individual diseases and individual genes. This focus is complicated by many patients having more than one disease, the heterogeneity of diseases as well as the involvement of thousands of genes, many of which are shared by more than one disease. Several observations point to the need to systematically investigate the pathogenic roles of shared, or pleiotropic genes: their importance is well established in model organisms and monogenic diseases [1-3]. For example, mutations

in the same gene, *ARX*, cause two highly diverse diseases, namely ambiguous genitalia and lissencephaly, a disease of the brain (Online Mendelian Inheritance in Man no. 300215). Genome-wide association studies (GWAS) have also shown the importance of pleiotropy in complex diseases [4-8]. Identification of pleiotropic genes by GWAS of complex diseases is limited by the relatively modest effects of genetic variants [5] and because epigenetic causes also play important roles [9,10]. Since mRNA expression can be affected by both genetic and epigenetic variants, expression profiling may be an optimal way to identify pleiotropic genes. In order to get an overview of the large number of genes involved in complex diseases, a module-based strategy can be applied. Briefly, this strategy is based

\* Correspondence: [mika.gustafsson@liu.se](mailto:mika.gustafsson@liu.se)

<sup>1</sup>The Centre for Individualised Medicine, Department of Clinical and Experimental Medicine, Linköping University, 58185 Linköping, Sweden  
Full list of author information is available at the end of the article

on the concept that disease-associated genes are not randomly distributed when mapped on the human protein-protein interaction (PPI) network. Instead, they tend to co-localize and form modules in the PPI network [11-22]. The advantages of modules are that they are thought to contain a limited number of the most disease-relevant genes, and that a functional overview of disease mechanisms can be obtained by pathway analysis.

In this study, we used a module-based approach to systematically investigate the pathogenic, diagnostic and therapeutic roles of pleiotropic genes. The approach was applied to public expression profiling data from different diseases. Because gene expression varies in different cell types, such data should ideally be derived from the same cell type. To reach this ideal, we first performed pathway analysis of all published GWAS to find a cell type common to many diseases. This resulted in the identification of the T helper (Th) cell differentiation pathway, which is expressed in T cells. We therefore focused on expression profiling of CD4<sup>+</sup> T cells from highly diverse inflammatory and malignant diseases. Analysis of the expression profiling data, which were independent from the GWAS data, resulted in the identification of pleiotropic genes. These genes were highly interconnected and formed a pleiotropic module. Despite being derived from the same cell type, this module was associated with a general increase in disease susceptibility: it was enriched for highly diverse pathways, as well as genetic variants described by all GWAS and cancer studies. The module was also enriched for known diagnostic and therapeutic targets. Moreover, the pleiotropic or disease-specific modules could be used to stratify patients in prospective clinical studies of multiple sclerosis and allergy. Our results show the clinical importance of pleiotropic genes as well as of translational genomics.

## Materials and methods

### Ethics statement

The studies were approved by the ethics board of University of Gothenburg and the ethics board of University of Linköping. All participants provided written consent for participation.

### Study subjects

The study comprised two different sets of experimental data: 1) response to cortisone treatment in 48 seasonal allergic rhinitis (SAR) patients, of which 8 were classified as high responders (HRs; see below for definitions) and 8 as low responders (LRs; 5 men and 11 women, mean age  $35.6 \pm 2.3$  years); and 2) response to natalizumab treatment in 50 multiple sclerosis (MS) patients, of which 8 were classified as HRs and 8 as LR (14 men and 2 women, mean age  $35.6 \pm 1.5$  years). The SAR patients and healthy donors were of Swedish origin and recruited at The Queen Silvia Children's Hospital, Gothenburg. SAR was defined by

a positive seasonal history and a positive skin prick test or by a positive ImmunoCap Rapid (Phadia) to birch and/or grass pollen. Patients with perennial symptoms or asthma were not included. The healthy subjects did not have any history for SAR and had negative ImmunoCap Rapid tests. The MS patients suffering from relapsing-remitting disease were of Swedish origin and recruited at the Department of Neurology at Linköping University Hospital. Patients had definite MS according to the McDonald criteria. All the microarray data from this study have been deposited under the Gene Expression Omnibus superseries GSE44966.

### Definition of low and high responders to treatment

An initial cohort of 48 allergic patients was asked to mark their symptoms (rhinorrhea, congestion, and itching) on a visual analogue scale (VAS) before and after two weeks of nasal treatment with fluticasone. We defined LR/HRs from the decrease in symptom scores when treated for two weeks in the pollen season, using previously described methods [23]. Fifty patients were included in the MS group and treated with natalizumab (NZB; Tysabri®); patients were classified as LR (n = 8) if they had experienced at least one relapse during this follow-up period and as HR if this had not occurred (n = 42, of which 8 were selected because of age, gender and disease matching to the LR group). All patients were followed for three years, except two LR, who were followed for two and one year, respectively. The annual relapse rate was  $0.65 \pm 0.26$  for LR and 0 for HRs. Samples analyzed for gene expression were taken prior to clinical natalizumab treatment. Both groups had comparable age distribution (mean age  $37.3 \pm 5.8$  (standard deviation) years for LR and  $33.9 \pm 6.2$  for HRs), Expanded Disability Status Scale score (LR, median 2.5, range 0.0 to 7.0; HRs, 2.0, range 1.0 to 5.0), disease duration (LR, mean  $9.1 \pm 6.6$  (standard deviation) years; HR,  $9.9 \pm 6.3$  years) and gender distribution (seven males in both groups) at the time of inclusion.

### Bioinformatics analyses

Public data were downloaded from the Gene Expression Omnibus that met the following criteria on 31 December 2012 (Table S6 in Additional file 1): i) expression profiling of CD4<sup>+</sup> T cells from healthy controls and patients with T-cell-related non-virus diseases; ii) at least five samples per disease and controls; and iii) the patients were not drug treated. In this study, data were quantile-normalized and log-transformed. The GCs and natalizumab materials were corrected for potential batch effects using COMBAT [24], with phenotype and stimulation as covariates. We mapped all probes to the corresponding genes and in cases where multiple probes for the same gene were present we used the median probe levels. Differential expression was computed using the LIMMA package in R. Classification by the LASSO was performed by the MatLab function `lassoglm` in

Statistics Toolbox choosing  $\lambda$  from the minimum deviance of leave-one-out cross-validation starting from all the measured genes of the platform. To determine if a set of genes was differentially expressed we used the mean of the squares of the student  $t$ -test statistic. This particular choice was made since its theoretical distribution is  $\chi^2(n - 1)$  and thereby the size effect was easy to interpret [25]. However, for calculating the  $P$ -values we still used permutation test of the mean value of the squared student  $t$ -values. We also performed complementary analysis using the log  $P$ -values with similar results. Permutation tests were performed using  $10^6$  permutations if nothing else is stated, and for the estimation of small  $P$ -values we approximated the distribution using [26]. The enrichments of GWAS genes, cancer genes, mouse knockout phenotypes, therapeutic targets, and biomarkers was also controlled for connectivity biases. This was performed by randomly sampling 158 genes from the STRING network with the same median (182) and minimal degree (39) as the pleiotropic module and repeating this procedure  $10^6$  times. The false discovery rate (FDR) was determined using the Benjamini Hochberg [27] correction method. The size effect was represented by fold enrichments (FEs), which is defined here for a specific feature and test set by the frequency of genes in the tested set with a certain feature divided by the frequency of the feature among all annotated genes. For example, we found the GWAS frequency 33.5% (53/158) among the pleiotropic genes, whereas this in the background of all genes is about 10.2% (2,298/22,500), and therefore  $FE = 33.5/10.2 = 3.3$ .

#### Databases

In order to test the robustness of the pleiotropic module we downloaded the latest versions (15 November 2013) of five different databases covering different aspects of the human interactome, namely HPRD [28], Reactome [29], Intact [30], HI2 (updated union of HI-2011, HI-2009 and HI-2005 from [31]) [30], and a high confidence database [32]. For each of these databases we then first computed the shortest distances between all pairs of proteins in the pleiotropic module that were also present in the largest connected component. Then we randomly selected the same number of genes from the largest connected component and similarly computed their shortest paths. We repeated the random selection procedure 1,000 times and calculated  $P$ -values using a Wilcoxon test of the mean values of the shortest paths. The enrichment of pathways within GWAS genes was tested using all pathways (as of 15 November 2013) in three different annotation databases, namely Kyoto Encyclopedia of Genes and Genomes (KEGG) [33], Ingenuity Pathway Analysis (IPA; Ingenuity® Systems, [34]), and Gene Ontology (GO) [35]. These enrichment analyses were performed using all human genes in NCBI (National Center for Biotechnology Information) as background. The analysis

of mouse knockout phenotypes was performed by downloading phenotypic information from [36] as of 31 January 2013. The analysis of therapeutic targets and biomarkers was also performed by first downloading all therapeutic target drugs and all genes annotated as markers for disease or prognosis in the IPA database.  $P$ -values for enrichment were calculated using Fisher's exact test (FET) using all genes as background ( $n = 22,500$  for human and  $n = 6,964$  for mice).

For the correlation analysis between the basal expression of the 158 genes and the drug responses across cancer cell lines, data were downloaded from the Genomics of Drug Sensitivity project [37]. This dataset included IC50 values for a total of 131 drugs that were assessed in a panel of 638 human cancer cell lines. In combination with this information, gene expression data from the same study (corresponding to the basal transcriptional profiles for 595 cell lines and publicly available on the ArrayExpress reference E-MTAB-783) were analyzed. Normalized expression values were obtained using the RMA algorithm [38]. Non-annotated probes were removed from the analysis and correlations were computed using the Pearson's correlation coefficient (PCC). For the null distribution, 1,000 sets of 158 genes randomly selected from the same microarray dataset were analyzed.

#### Disease-associated network modules

The network and modules were constructed using a modified version of a previously described method [39]. First, all maximal cliques [29] were extracted from the human PPI database STRING [40] (version using 9.03 interactions with a confidence score  $\geq 0.7$ ). Then, each clique was assigned a weight, defined as the sum of the  $-\log(P\text{-values})$  for all the genes in the clique, based on the differential expression analysis (henceforth referred to as the real weight). To assess how the weight for each clique differed from the null distribution, the  $P$ -values for all the genes in the differential expression analysis were randomized, and the weight for each clique was again determined by the same method. This process was repeated 10,000 times, to give a null distribution of weights for each clique. The significance of each clique's weight was defined as the fraction of random permutations that resulted in a higher weight than the real weight. Disease modules were then constructed by mapping the cliques that had a significantly higher weight than expected by the null distribution ( $P < 0.01$ ). This analysis resulted in a module associated to each disease. We then identified the intersection between the disease modules. To determine how the size of the intersection differed from random, we repeated the above analysis 100 times using randomized  $P$ -values. This resulted in 100 disease modules for every disease. We then determined the intersection between these 100 sets of disease modules and compared this to the real intersection. Furthermore, we used these 100 sets of disease modules to identify genes that were represented in more

disease modules than expected by chance. Like before, the  $P$ -value of each gene was equivalent to the ratio of random permutations that included the gene in as many or more disease modules as the disease modules that resulted from real (non-randomized) data. The R code for the identification of disease modules from interactome and  $P$ -values can be found in Additional file 2, and the output disease modules are provided in Table S22 in Additional file 1.

#### **Culturing of CD4<sup>+</sup> T cells from patients with SAR**

Challenges with allergen or allergen plus glucocorticoids (GCs) were performed as previously described [41-43]. Briefly, peripheral blood mononuclear cells were prepared from fresh blood by means of centrifugation in Ficoll-Hypaque, washed, and stimulated with allergen extract (100 µg/ml; ALK-Abelló A/S Hørsholm, Denmark), or allergen plus hydrocortisone ( $10^{-7}$  M; Sigma-Aldrich, St. Louis, Missouri, USA). After 17 hours and 7 days, respectively, of incubation at 37°C and 5% CO<sub>2</sub>, the supernatant was removed and CD4<sup>+</sup> T cells enriched. The RNA samples were then analyzed using Agilent Sureprint G3 Human Gene Expression.

#### **Culturing of CD4<sup>+</sup> T cells from patients with MS**

Briefly, mononuclear cells were thawed from liquid N<sub>2</sub>. Next, CD4<sup>+</sup> T cells were separated using the Isolation Kit II (Miltenyi Biotech, Bergisch Gladbach, Germany). Purity of samples was typically >97%. The CD4<sup>+</sup> T cells were split into three cultures and an unstimulated culture served as baseline control. Two cultures were stimulated with pre-coated anti-CD3/-CD28 monoclonal antibodies (0.1 µg/ml). Cells were cultured for 48 hours in media consisting of Iscove's modified Dulbecco's medium (IMDM) supplemented with 5% fetal calf serum (Sigma-Aldrich), L-glutamine (292 mg/ml; Sigma-Aldrich), sodium bicarbonate (3.024 mg/ml; Sigma-Aldrich), penicillin (50 IE/ml; Cambrex, East Rutherford, New Jersey, USA), streptomycin (50 µg/ml; Cambrex), and 100× non-essential amino acids (Gibco BRL, New York, USA). One of these cultures was supplemented with Tysabri® at a final concentration of 25 µg/ml. After culturing, cells were lysed using TRI Reagent (MRC, London UK). Total RNA was extracted according to the manufacturer's instructions. RNA quality and quantity were assessed with the NanoDrop ND-1000 (NanoDrop Technologies Wilmington, New Jersey, USA). The Agilent Sureprint G3 Human Gene Expression 8x60k was used for gene expression analysis (Agilent Technologies, Santa Clara, California, USA).

## **Results and discussion**

### **Pathway analysis of GWAS leads to the selection of T cells to study the expression of pleiotropic genes**

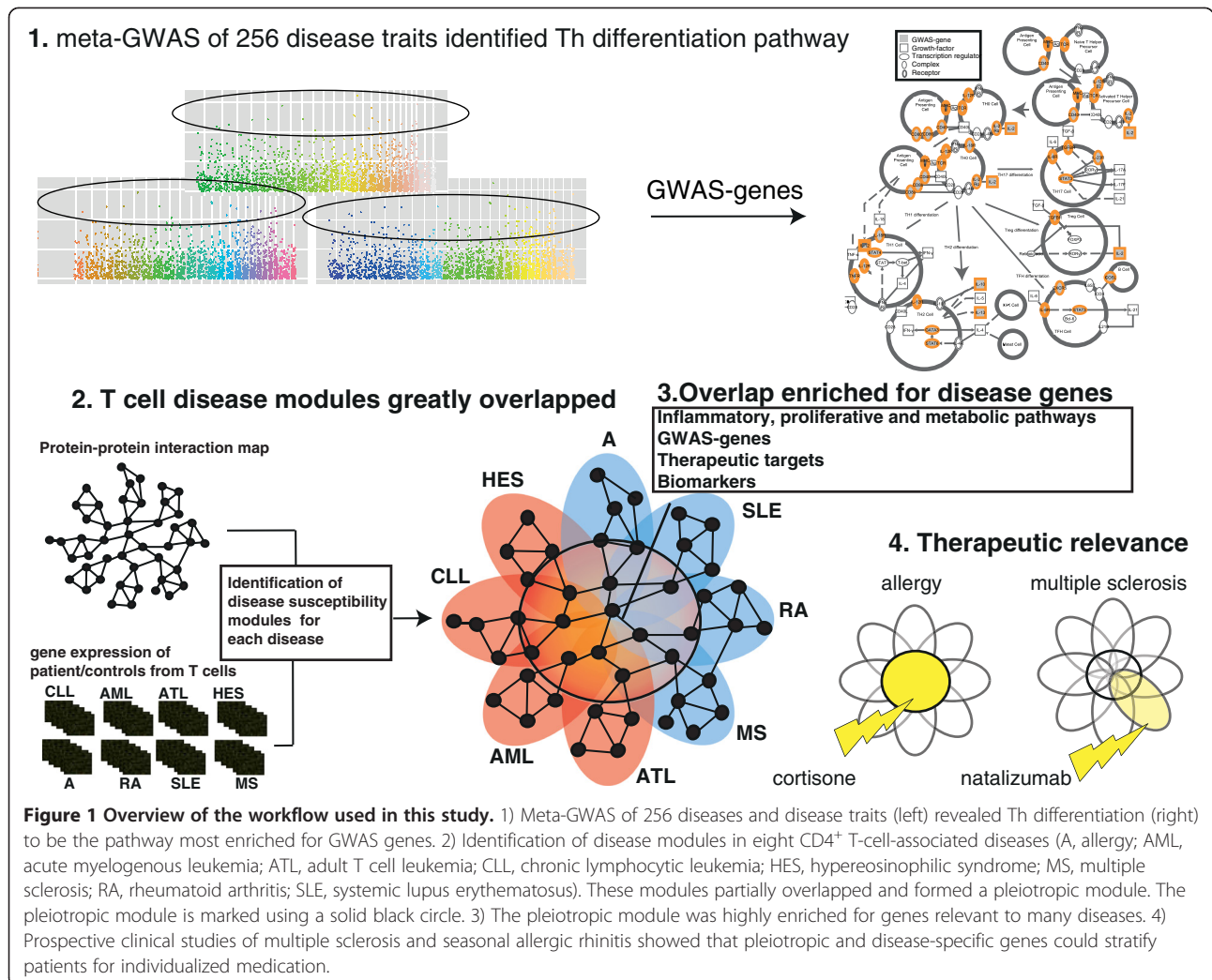
In order to identify pleiotropic genes at the transcriptomics level, the expression of those genes should ideally be

measured in the same cell type. Because expression varies in different cell types, we performed pathway analysis of GWAS to find a pleiotropic pathway. The aim was to use that pathway to find a pleiotropic cell type that was accessible for independent expression profiling studies. This led to the identification of Th differentiation, which is expressed in T cells, as described below.

First, meta-analysis of GWAS combined with pathway enrichment analysis was performed (Figure 1, analysis 1). We downloaded GWAS data compiled by the National Human Genome Research Institute [44]: this included 256 diseases and traits, and 2,298 gene loci harboring potentially associated SNPs (mapping inter-genic SNPs to the nearest upstream and downstream genes at a significance threshold  $P < 10^{-5}$ ; see Table S1 in Additional file 1 and Figure S2 in Additional file 3 for disease/trait distribution). These genes, which represent potential associations with diseases or disease traits, are henceforth referred to as 'GWAS genes' (Table S2 in Additional file 1). Next, using a large curated database (Ingenuity® Systems [34]), the Th cell differentiation pathway (Figure S2 in Additional file 3) was identified as the most enriched in the GWAS genes measured by the  $P$ -values from a FET ( $n = 38$ ,  $FE = 5.3$ , Bonferroni-corrected  $P < 10^{-15}$ ; Table S3 in Additional file 1 shows all genes and pathways). Th differentiation results in different T-cell subsets, such as Th1, Th2 and Th17, which are thought to have key roles in immune-related diseases, but are also associated with many other diseases [45-48].

In order to test the importance of Th differentiation in diseases that were not *a priori* immune related, and to minimize knowledge biases related to immune diseases, we performed six complementary analyses. These analyses supported the importance of Th differentiation.

First, we queried the GO [35] database for enrichments of the GO process T-cell differentiation and found that this was the case ( $P < 2 \times 10^{-6}$ , FET). Second, we tested that the corresponding pathway (T-cell receptor pathway,  $P < 4 \times 10^{-3}$ ) was enriched for the GWAS genes using KEGG (Table S4 in Additional file 1; see Figure S3 in Additional file 3 for a comparison of the overlap of IPA, KEGG, and GO). Third, we manually classified the disease traits into immune and non-immune (Tables S1 and S2 in Additional file 1) and again found the Th differentiation to be highly enriched also in the non-immune disease category ( $P < 1 \times 10^{-7}$ ; Table S5 in Additional file 1). Fourth, we repeated the analyses for all genes within each block, where the corresponding gene is in linkage disequilibrium [49] (Th differentiation,  $P < 3 \times 10^{-7}$ ; Tables S6 and S7 in Additional file 1). Fifth, we repeated the analysis for 446 genes with known somatic cancer mutations [50], henceforth referred to as 'cancer genes' (Th differentiation,  $P < 5 \times 10^{-4}$ ; Tables S8 and S9 in Additional file 1). Sixth, 4,613 genes annotated to disease from the Online Mendelian Inheritance in Man database [51] (visited on 12



November 2013; Th differentiation,  $P < 2 \times 10^{-11}$ ; Tables S10 and S11 in Additional file 1).

Since the Th differentiation pathway is expressed in CD4<sup>+</sup> T cells, we proceeded to analyze expression profiling data from this cell type. Note that the analyses below are based on all genes, and not only genes associated with Th differentiation.

#### A pleiotropic module in CD4<sup>+</sup> T-cell-associated diseases was associated with generally increased disease susceptibility

In order to identify pleiotropic genes, publicly available gene expression microarray data for eight diseases were analyzed: four inflammatory diseases (allergy, multiple sclerosis, rheumatoid arthritis, and systemic lupus erythematosus), and four malignant or proliferative diseases (acute myelogenous leukemia, adult T cell leukemia, chronic lymphocytic leukemia, and hypereosinophilic syndrome; Figure 1, analysis step 2; Table S12 in Additional file 1; the

diseases were chosen because they met the criteria described in Materials and methods).

Despite the diversity of these diseases, we found that the genes that were differentially expressed between patients and healthy controls were not generally dispersed when mapped on the human PPI network. Instead, the genes from each disease formed modules, which partially overlapped. The overlapping genes were highly interconnected and formed a pleiotropic module. This module was associated with a wide variety of diseases, as shown by pathway analysis, phenotypes resulting from mouse knockout studies, as well as an enrichment of GWAS and cancer genes. The analyses are detailed below.

We first confirmed that patients and controls could be correctly classified using the LASSO approach ( $10^{-16} < P < 10^{-3}$ , permutation test; Figure S4 in Additional file 3; Table S12 in Additional file 1; see Materials and methods for details).

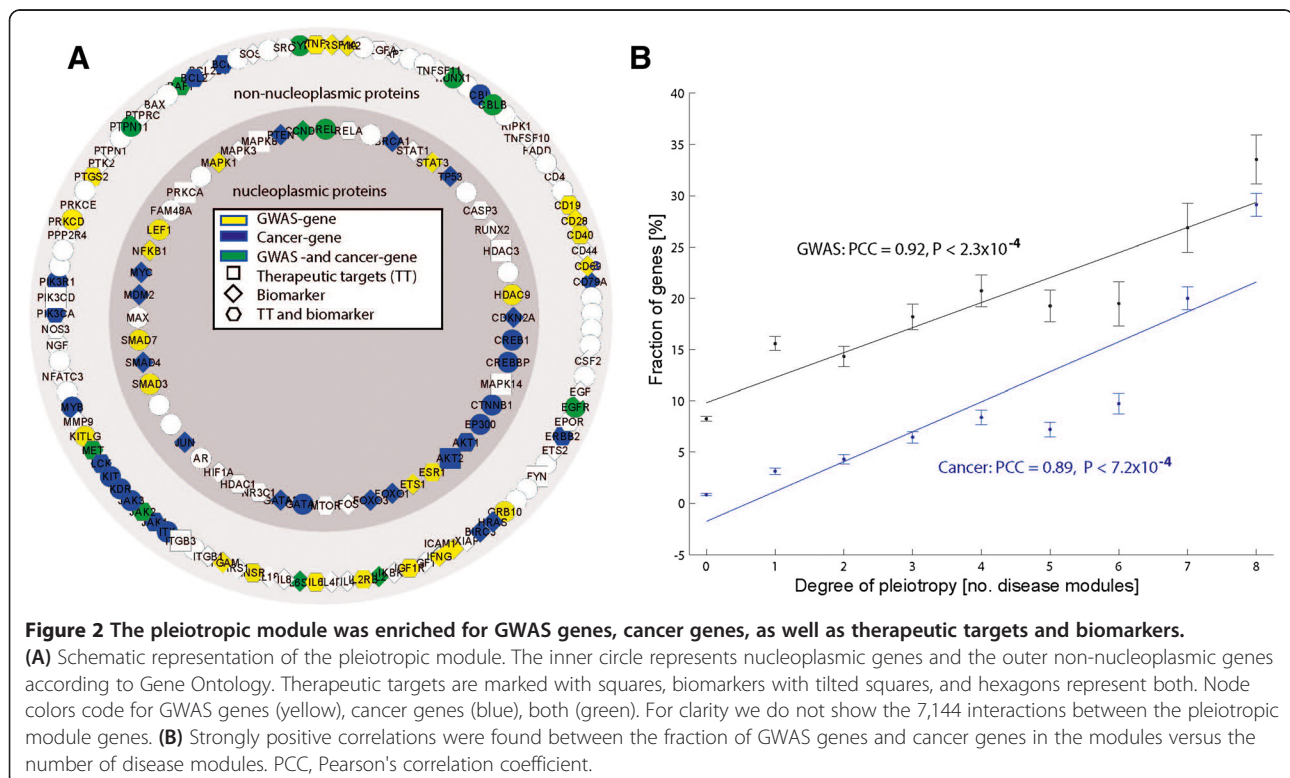
Differentially expressed genes for each of the eight diseases were identified and subsequently mapped on a

network of known and predicted human PPIs from the STRING database [40] (Materials and methods). For each disease, we identified subnetworks of highly interconnected and differentially expressed genes, using our previously published method [39]. Such interconnected and differentially expressed genes are henceforth referred to as disease modules. In our study, this led to the identification of eight different sets of partially overlapping and interconnected disease modules, which each consisted of 1,215 to 1,933 genes.

We examined if the disease modules for each of the eight different diseases overlapped significantly (Materials and methods). We found a highly significant overlap of 158 genes (Table S13 in Additional file 1) in all eight disease modules ( $FE = 7.0$ ,  $P < 10^{-61}$  permutation test; Materials and methods; Figure 2). The 158 genes were highly interconnected and had 7,144 interactions between each other. By contrast, the mean number of connections between the 158 genes in the disease module and 158 random genes was only 376 ( $P < 10^{-300}$ , permutation test). As different PPI databases have different inclusion criteria, we tested if the overlapping 158 genes had a lower mean shortest path among each other than 158 random genes, using five different databases covering different aspects of the human interactome (Materials and methods). This analysis showed that, for all five analyzed data sets, we had significantly lower mean shortest paths among the

158 genes in the overlap of all the disease modules than expected ( $P < 10^{-3}$  in four databases and  $P < 0.011$  for one database using one-sided Wilcoxon rank test; see Materials and methods and Figure S5 in Additional file 3 for details).

Henceforth, the 158 genes are referred to as the 'pleiotropic module'. Despite the expression profiling data being derived from T cells, pathway analysis of the pleiotropic module showed enrichment not only of inflammatory pathways, but also of multiple other disease-associated pathways (Table S14 in Additional file 1 shows results for all pathways enriched in the pleiotropic module). Because of the pleiotropy and interconnectivity of the pathways, we hypothesized that the genes in the pleiotropic module would generally increase disease susceptibility. This was supported by analysis of the mouse knockout database, which showed that a wide range of phenotypes and diseases resulted from knock out of genes in the pleiotropic module (Bonferroni corrected  $P$ -values from  $P < 10^{-50}$ , FET; Table S15 in Additional file 1). The pleiotropic module was also significantly enriched for both GWAS and cancer genes (GWAS,  $n = 53$ ,  $FE = 3.3$ ,  $P < 10^{-15}$ ; cancer,  $n = 47$ ,  $FE = 15.4$ ,  $P < 10^{-41}$ ; Figure 1). Moreover, we tested the general disease importance of the pleiotropic module by only studying GWAS for the 158 diseases and disease traits that were not *a priori* associated with immune-related diseases and cancer (see Table S1 in Additional file 1 for disease categorizations), which then consisted of



1,437 genes. e found 19 of the pleiotropic genes were significantly associated with these diseases and disease traits ( $FE = 1.9$ ,  $P < 4 \times 10^{-3}$ ).

We also confirmed that the enrichment was significantly more than expected by the degree of the corresponding genes (Materials and methods; Table S16 in Additional file 1). Then, in order to test if the enrichment was dependent on the definition of the pleiotropic module, we performed a correlation analysis: we correlated the fraction of disease susceptibility genes and the number of disease modules for each gene and found a strong significant positive PCC for both GWAS ( $PCC = 0.91$ ,  $P < 4 \times 10^{-4}$ ) and cancer ( $PCC = 0.94$ ,  $P < 2 \times 10^{-4}$ ) genes (Figure 1B). Thus, the enrichment was not due to the definition of the pleiotropic module.

A limitation in the above analyses of SNPs from public databases was that those SNPs had been defined based on genome-wide significance levels, although we had only examined 158 genes. In order to increase statistical power, we also analyzed the 158 genes for nominally significant associations in the original GWAS data from MS [52], comprising approximately 25,000 individuals. We defined a gene to be nominally significant if it harbored at least one SNP with a nominally low  $P$ -value. We found that the pleiotropic genes were enriched for MS SNPs using  $P$ -values in the range  $10^{-5}$  to 0.05 (Figure S6 in Additional file 3). For example, we found that 60% (89/148) of the assayed probes had at least one SNP with  $P$ -value  $< 0.05$  ( $P < 2 \times 10^{-6}$ , FET) and 40% had SNPs with a  $P$ -value  $< 0.01$  ( $P < 10^{-12}$ ). Furthermore, by examining all assayed SNPs ( $n = 4,990$ ) of the pleiotropic genes we found that 150 SNPs for 51 unique pleiotropic genes (Table S17 in Additional file 1) were disease associated using a 20% FDR, which represents a huge increase from the five genes originally reported using genome-wide significance criterion.

#### **The pleiotropic module was highly enriched with known diagnostic markers and therapeutic targets**

Having identified a link with disease development, we next assessed if the pleiotropic module was enriched for known diagnostic markers ( $n = 1,177$ ) and/or therapeutic targets ( $n = 404$ ) compiled from all diseases [53]. The results from these analyses revealed a significant enrichment in both markers and targets ( $n = 88$ ,  $FE = 5.5$ ,  $P < 10^{-46}$ ;  $n = 36$ ,  $FE = 12.8$ ,  $P < 10^{-28}$ , respectively, using FET; Figure 1A). These enrichments were not confounded by the high connectivity of the pleiotropic module or by the number of genes/proteins and diseases examined (Figure S7 in Additional file 3; Table S11 in Additional file 1). Since the interaction partners of known drug targets are likely to be functionally related, we hypothesized that those partners may be novel therapeutic candidates across diseases. Indeed, examination of predicted druggable PPIs [54] in the

pleiotropic module identified 173 novel therapeutic candidates (Table S18 in Additional file 1).

Since the pleiotropic module was associated with generally increased disease susceptibility, the identification of approved drugs that may significantly alter its activity could be clinically relevant. To address this hypothesis, data from the determination of the half maximal inhibitory concentration (IC<sub>50</sub>) of more than 100 drugs across 100 cancer cell lines [37] was analyzed. Although some correlations may be specific to cancer status, this dataset represents a large collection of different tissue types. Next, relative to 1,000 random sets of 158 genes compiled from the same expression dataset, several drugs were found to be frequently correlated with genes from the pleiotropic module (Figure S8 in Additional file 3; Table S19 in Additional file 1). This analysis showed 37 of the drugs significantly changed the expression of the genes in the pleiotropic module more than other genes at a FDR of 5%. Thus, those drugs may be therapeutic candidates in multiple diseases. The top significant drugs were two allosteric AKT inhibitors (AKT inhibitor VIII and MK-2206), which suggests that signaling through this kinase would be critical in different pathologies. Importantly, the pleiotropic influence of AKT function has been previously discussed in different scenarios, which include immune-related processes and pathologies.

#### **Disease stratification based on pleiotropic or disease-specific genes**

Because the pleiotropic module was enriched for known therapeutic targets, we hypothesized that changes in the expression of pleiotropic or disease-specific genes could be used to stratify patients for response to drugs targeting these two gene categories. Briefly, we carried out prospective clinical studies of 50 patients with MS and 48 patients with SAR to define HRs and LR to treatment. In order to identify genes that responded to treatment, CD4<sup>+</sup> T cells from the patients were analyzed with gene expression microarrays before and after *in vitro* exposure of the drug. Next, we examined if genes that were shared or specific for the two diseases could separate HRs and LR. We found that HRs and LR could be separated by shared genes in SAR, and by MS module genes in MS.

The 48 SAR patients were treated with GCs for two weeks during the pollen season. GCs generally reverse the expression levels of genes involved in the inflammatory response [41] and are used in the treatment of several immune diseases. The 50 MS patients were treated with natalizumab, and followed clinically during three years. Natalizumab is a drug that is mainly used in MS and specifically targets a membrane protein responsible for lymphocyte passage through the blood-brain barrier, and also influences gene expression in lymphocytes [55,56]. In both diseases, clinical specialists classified subsets of

patients that were HRs and LRs (Materials and methods). The CD4<sup>+</sup> T cells were obtained from untreated patients during symptom-free periods. Gene expression microarray analyses of the SAR patients showed increased likelihood of GC response for genes within many disease modules (PCC = 0.79,  $P = 0.0011$ ; Figure 3A). By contrast, in MS, the opposite non-significant trend was found for natalizumab (PCC = -0.61,  $P = 0.079$ ; Figure 3A).

Next, we analyzed if the disease module knowledge could be translated to prediction of treatment response for the two drugs. Therefore, we first selected only genes differentially expressed *in vitro* by the drug for further analysis ( $P < 0.05$ ). Since genes in multiple disease modules more frequently responded to GCs, we further limited the GC classifier into genes within at least two disease modules ( $n = 311$ ). Then, using the LASSO and cross-validation for model parameter fits, most patients in both diseases were correctly classified as LRs or HRs (permutation test  $P < 3 \times 10^{-4}$ ; Figure 3B). Analysis of the classifiers showed that they exploited different features. The GC classifier used GC pathway ( $n = 29$ ,  $P < 10^{-13}$ , FET; see Table S20 in Additional file 1 for all pathways). By contrast, since natalizumab did not affect genes in multiple modules more frequently, we limited the MS classifier to genes in the MS module ( $n = 28$ ). Conversely, the natalizumab classifier utilized genes dispersed in different pathways, and instead the main difference between HRs and LRs was due to a higher perturbation of the LRs than HRs

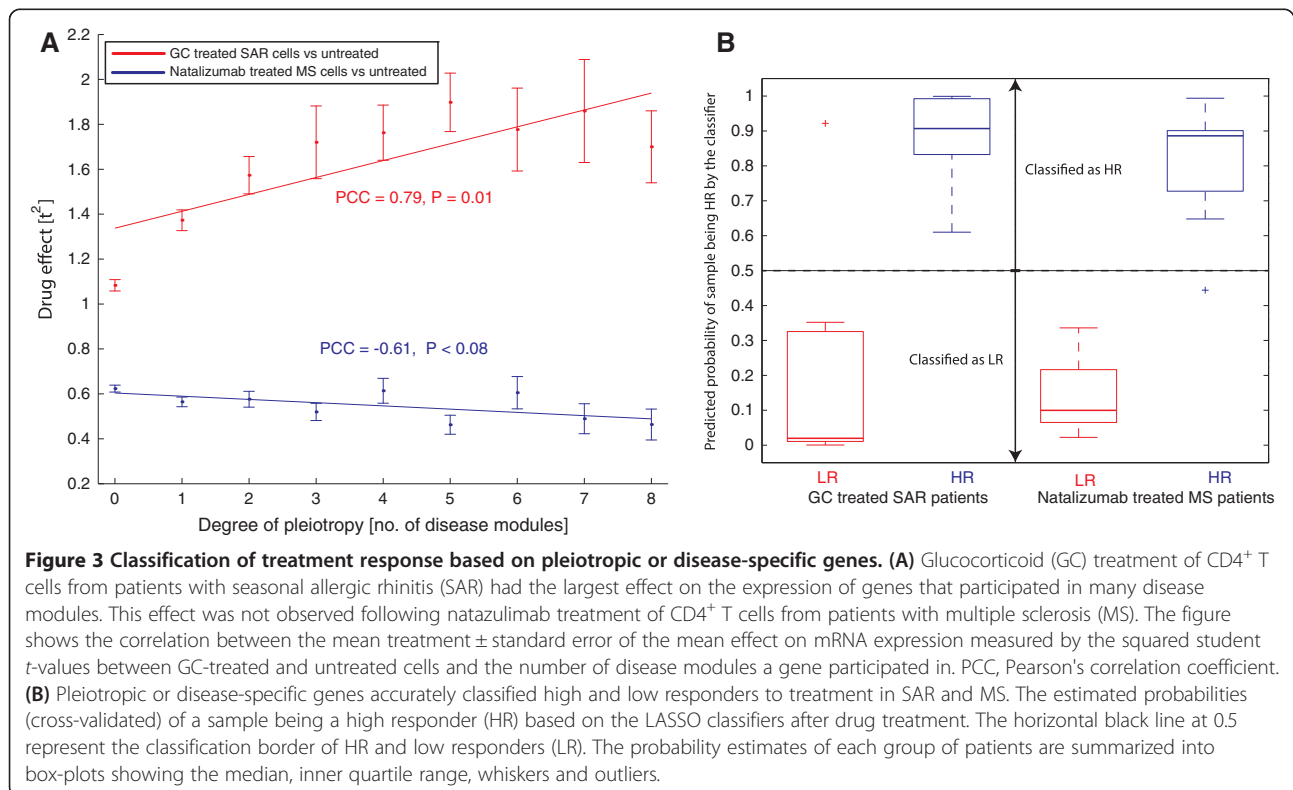
by the drug ( $\langle t_{LR}^2 \rangle = 1.20$ ,  $\langle t_{HR}^2 \rangle = 0.66$ , permutation test  $P < 10^{-300}$ ; Figure S9 in Additional file 3; see Table S21 in Additional file 1 for enriched pathways in the differentially expressed genes for LRs).

Thus, pleiotropic or disease-specific genes have the potential to stratify patients for drug response depending on which genes are targeted by the drug.

## Conclusions

The main findings of this study are that pleiotropic genes in T-cell-associated diseases formed a module of highly interconnected genes. This pleiotropic module generally increases disease risk and is an important source for diagnostic markers and therapeutic targets. To our knowledge, this is the first study to systematically address the functional and clinical implications of pleiotropic genes in complex diseases. Our study may have important implications for the general understanding of complex disease mechanisms and translational genomics.

The background to our study was that translational research mainly focuses on investigating disease-specific genes/proteins. Nonetheless, recent observations from GWAS have suggested an important etiological role for pleiotropic genes/proteins [4,6-8,39]. However, an important limitation in the studies of pleiotropic genes based on GWAS is that epigenetic causes may play important roles for complex disease susceptibility [9].





In this study, we reasoned that mRNA expression would be an optimal level to identify pleiotropic genes because it can be affected by both genetic and epigenetic variants. To address the problem that gene expression varies in different diseases and cell types, we searched for a common cell with pleiotropic genes based on pathway analysis of GWAS. This novel approach resulted in the identification of the Th differentiation pathway, which is expressed in CD4<sup>+</sup> T cells. The Th differentiation pathway has a key role in orchestrating immune responses in auto-immune and allergic diseases [45], atherosclerosis [46], cancer [47], and obesity [48]. Our finding is in line with previous GWAS showing the importance of immune-related genes in complex diseases [5]. On the other hand, the identification of the Th pathway could be confounded by the selection of diseases analyzed with GWAS. Some definitions of pleiotropy require that the studied diseases or disease traits are seemingly unrelated, while others argue that such definitions are too subjective [57]. In this study, we addressed this potential confounder by complementary analyses. For example, to test if the identification of the Th pathway was due to the selection of diseases examined by GWAS, we repeated the analyses on somatic mutations in cancer and found that the Th pathway was significantly enriched. Next, we analyzed expression profiling data from CD4<sup>+</sup> T cells from eight inflammatory and malignant diseases. In each of the diseases we could define modules of differentially expressed genes. Despite the diversity of the diseases, the modules overlapped. The overlapping part was highly interconnected and formed a pleiotropic module, which was enriched not only for the Th differentiation pathway, but also for several other inflammatory, metabolic and proliferative pathways. These pathways contained individual genes of known or recently recognized importance for multiple diseases. For example, *GATA3* is known to have a key role in Th2 differentiation in allergy, but is also a potential diagnostic marker in epithelial cancers [58]. Conversely, the tumor suppressor gene *BRCA1*, which has an important role in breast cancer, was also differentially expressed in allergen-challenged T cells. This finding is in line with *BRCA1* potentially having a role in regulating inflammation [59].

We speculated that because of the pleiotropy and interconnectivity of the pathways, the pleiotropic module would generally increase disease susceptibility. Remarkably, despite the pleiotropic module being derived from eight inflammatory and hematological diseases, it was significantly enriched with GWAS genes from all published analyses of diseases and disease traits. Thus, rather than being dispersed in the interactome, a limited number of highly interconnected genes that regulated key pathways generally increased disease susceptibility.

A confounding factor when using collections of experimental studies might be potential knowledge-related biases,

that is, the experimental studies are not equally distributed among the genes. Therefore, to limit this problem we confirmed that the identified pleiotropic module genes were highly associated using databases with several other inclusion criteria, including systematic and high confidence databases. It is also important to note that the genes in the pleiotropic module were more enriched for both GWAS genes and disease phenotypes than the disease-specific genes. The findings for the GWAS genes were also replicated for cancer genes. This indicates that pleiotropic genes may have larger impact on disease phenotypes than specific genes, which has important diagnostic and therapeutic implications that are discussed below.

We then tested the potential of therapy of the pleiotropic module, and found that it was highly enriched for known diagnostic markers and therapeutic targets. Interestingly, the module also contained a large number of druggable genes that were not known drug targets. Those genes represent new therapeutic candidates across diseases. Thus, while disease-specificity is often desired for both types of molecules, our observations point to pleiotropic genes/proteins being relevant for '4P medicine' (predictive, preventative, personalized and participatory medicine [60]). Knowledge from the pleiotropic module also provides insight into disease stratification for individualized medication. Prospective clinical studies in MS and SAR combined with gene expression profiling revealed differences relative to the modular impact of the respective disease treatments. While the GCs (used in SAR) affected multiple disease-associated modules, natalizumab (used in MS) was linked to a single module. Subsequently, the expression of the pleiotropic genes was found to classify GC responders with high accuracy, while the disease-specific genes classified natalizumab responders. This observation leads us to propose that the definition of pleiotropic versus specific genes/proteins may help to identify markers for stratification, based on network molecular consequences of the therapies. Prospective studies in other diseases are required to assess this hypothesis.

Another important future research direction is to examine how the same pleiotropic genes can be associated with different diseases. Possible explanations include altered interactions between those genes, as well as between disease-specific genes. Furthermore, analyses of pleiotropic genes in other pleiotropic cells are warranted. Our pathway analysis of GWAS data indicates that there may be several different pleiotropic cells, for example, B lymphocytes, macrophages and epithelial cells. Since these cell types interact, it would be very interesting to simultaneously profile their gene expression profiles. This could help to understand disease-associated interactions on a multi-cellular scale.

In summary, given the results of this study, we propose that the pleiotropic genes/proteins warrant extensive functional and clinical studies across diverse diseases.

## Additional files

**Additional file 1: Table S1.** Manual classification of all traits and diseases in the GWAS catalogue. **Table S2.** Gene symbols of all mapped genes that harbor at least one disease-associated SNP corresponding to a disease trait. **Table S3.** All pathways for all the GWAS genes reported by IPA. **Table S4.** All KEGG pathways for all the GWAS genes. **Table S5.** Ingenuity pathways for the GWAS genes with any non-immune trait. **Table S6.** Gene symbols of all genes harboring at least one disease-associated SNP in the corresponding linkage disequilibrium block. **Table S7.** Ingenuity pathways for the genes in Table S6. **Table S8.** Gene symbols of the mapped genes with a somatic cancer mutation in the COSMIC database. **Table S9.** Ingenuity pathways for the genes in Table S8. **Table S10.** Gene symbols of the mapped genes within the OMIM database. **Table S11.** Ingenuity pathways for the genes in Table S10. **Table S12.** Statistics of all eight T-cell diseases. **Table S13.** Descriptions of the pleiotropic module genes. **Table S14.** Ingenuity pathways for the genes in Table S13. **Table S15.** Knockout mice phenotypes of the pleiotropic module genes. **Table S16.** Statistics of median degree preserving randomization of the enrichment of GWAS genes, biomarkers, therapeutic targets, cancer genes, immune mice knockout genes and pleiotropic module genes. **Table S17.** Nominal *P*-values for disease association in original GWAS of MS for all measured SNPs of the pleiotropic module genes. **Table S18.** Analysis of all druggable interactions between the pleiotropic module genes. **Table S19.** Drugs significantly associated with the expression of the pleiotropic module genes. **Table S20.** Ingenuity pathways for the GC-perturbed genes. **Table S21.** Ingenuity pathways for the genes differentially expressed in low responders of natalizumab treatment. **Table S22.** Identified disease modules for each of the studied diseases.

**Additional file 2: R code for the derivation of the disease modules, which integrate binding and expression data.**

**Additional file 3: Figure S1.** Pie chart of the distribution of manual classification (Table S2) of disease traits from all GWAS. **Figure S2.** The Th differentiation pathway (Ingenuity® Systems [34]), where GWAS genes (grey), growth factors (squares), transcription factors (broad ovals), complexes (standing ovals), and receptors (double standing ovals) are marked. **Figure S3.** Comparison of the overlap of genes annotated to T helper (Th) differentiation according to different data sources. **Figure S4.** The estimated probabilities (cross-validated) of a sample being a patient based on the LASSO classifiers of each of the eight T-cell diseases. **Figure S5.** The genes in the pleiotropic module had relative mean shortest paths among each other. **Figure S6.** Enrichment of nominally associated SNPs of MS for the genes in the pleiotropic module (black bars) using different cutoffs for the *P*-value for determining disease association. **Figure S7.** Mean (dots) ± standard error of the mean (bars) of the fraction of genes that are biomarkers (red) and therapeutic targets (blue) versus number of disease modules. **Figure S8.** Correlation analysis between the expression of the pleiotropic module genes and the responses of 131 drugs across cancer cell lines measured using IC50 (see Materials and methods). **Figure S9.** Differential expression of genes between MS patients before versus after *in vitro* natalizumab treatment, measured by the squared student *t*-values.

## Abbreviations

FE: fold enrichment; FET: Fisher's exact test; GC: glucocorticoid; GO: Gene Ontology; GWAS: genome-wide association study; HR: high responder; IPA: Ingenuity Pathway Analyses; KEGG: Kyoto Encyclopedia of Genes and Genomes; LR: low responder; MS: multiple sclerosis; PCC: Pearson's correlation coefficient; PPI: protein-protein interaction; SAR: seasonal allergic rhinitis; SNP: single-nucleotide polymorphism; Th: T helper.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

MG and MB conceived the study. ME, HW, and HZ performed wet lab experiments. MG, DG, FB, CEN, JS, NB, and MAP performed the data analysis. CEN, MAP, JE, MB, JT, IK, and TO was involved in preparing figures and writing the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

This study was supported by grants from the European Community's Seventh Framework Programme (FP7/2007-2012) under grant agreement 223367 (MultiMod), AFA Foundation, the Knut and Alice Wallenbergs Foundation, the Swedish Brain Foundation, the Swedish Research Council, and the Spanish Ministry of Health ISCIII FIS 12/01528 and RTICC RD12/0036/0008.

## Author details

<sup>1</sup>The Centre for Individualised Medicine, Department of Clinical and Experimental Medicine, Linköping University, 58185 Linköping, Sweden. <sup>2</sup>Clinical and Experimental Medicine, Faculty of Health Sciences, Division of Clinical Immunology, Unit of Autoimmunity and Immune Regulation, Linköping University, 58185 Linköping, Sweden. <sup>3</sup>Department of Clinical Neurosciences, Karolinska Institutet and Centrum for Molecular Medicine, 17177 Stockholm, Sweden. <sup>4</sup>Cancer and Systems Biology Unit, Catalan Institute of Oncology, IDIBELL, L'Hospitalet del Llobregat, 08908 Barcelona, Spain.

Received: 29 November 2013 Accepted: 21 February 2014

Published: 26 February 2014

## References

1. Wang Z, Liao BY, Zhang J: **Genomic patterns of pleiotropy and the evolution of complexity.** *Proc Natl Acad Sci USA* 2010, **107**:18034–18039.
2. Albin RL: **Antagonistic pleiotropy, mutation accumulation, and human genetic disease.** *Genetica* 1993, **91**:279–286.
3. Brunner HG, van Driel MA: **From syndrome families to functional genomics.** *Nat Rev Genet* 2004, **5**:545–551.
4. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, Abecasis GR, Barrett JC, Behrens T, Cho J, De Jager PL, Elder JT, Graham RR, Gregersen P, Klareskog L, Siminovitsh KA, van Heel DA, Wijmenga C, Worthington J, Todd JA, Hafler DA, Rich SS, Daly MJ, FOCIS Network of Consortia: **Pervasive sharing of genetic effects in autoimmune disease.** *PLoS Genet* 2011, **7**:e1002254.
5. Wellcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447**:661–678.
6. Ramos PS, Criswell LA, Moser KL, Comeau ME, Williams AH, Pajewski NM, Chung SA, Graham RR, Zidovetzki R, Kelly JA, Kaufman KM, Jacob CO, Vyse TJ, Tsao BP, Kimberly RP, Gaffney PM, Alarcon-Riquelme ME, Harley JB, Langefeld CD, International Consortium on the Genetics of Systemic Erythematosus: **A comprehensive analysis of shared loci between systemic lupus erythematosus (SLE) and sixteen autoimmune diseases reveals limited genetic overlap.** *PLoS Genet* 2011, **7**:e1002406.
7. Gutierrez-Achury J, Coutinho-de-Almeida R, Wijmenga C: **Shared genetics in coeliac disease and other immune-mediated diseases.** *J Intern Med* 2011, **269**:591–603.
8. Lees CW, Barrett JC, Parkes M, Satsangi J: **New IBD genetics: common pathways with other diseases.** *Gut* 2011, **60**:1739–1753.
9. Feil R, Fraga MF: **Epigenetics and the environment: emerging patterns and implications.** *Nat Rev Genet* 2012, **13**:97–109.
10. Nestor CE, Barrenäs F, Wang H, Lentini A, Zhang H, Bruhn S, Jörnsten R, Langston MA, Rogers G, Gustafsson M, Benson M: **DNA methylation changes separate allergic patients from healthy controls and may reflect altered CD4+ T-cell population structure.** *PLoS Genet* 2014, **10**:e1004059.
11. Lamb J, Ramaswamy S, Ford HL, Contreras B, Martinez RV, Kittrell FS, Zahnaw CA, Patterson N, Golub TR, Ewen ME: **A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer.** *Cell* 2003, **114**:323–334.
12. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC: **PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.** *Nat Genet* 2003, **34**:267–273.

13. Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B, Assmann V, Elshamy WM, Rual JF, Levine D, Rozek LS, Gelman RS, Gunsalus KC, Greenberg RA, Sobhian B, Bertin N, Venkatesan K, Ayivi-Guedehoussou N, Sole X, Hernandez P, Lazaro C, Nathanson KL, Weber BL, Cusick ME, Hill DE, Offit K, et al: **Network modeling links breast cancer susceptibility and centrosome dysfunction.** *Nat Genet* 2007, **39**:1338–1349.
14. Jonsson PF, Cavanna T, Zicha D, Bates PA: **Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis.** *BMC Bioinforma* 2006, **7**:2.
15. Segal E, Friedman N, Kaminski N, Regev A, Koller D: **From signatures to models: understanding cancer using microarrays.** *Nat Genet* 2005, **37**:S38–45.
16. Tuller T, Atar S, Ruppin E, Gurevich M, Achiron A: **Global map of physical interactions among differentially expressed genes in multiple sclerosis relapses and remissions.** *Hum Mol Genet* 2011, **20**:3606–3619.
17. Wang IM, Zhang B, Yang X, Zhu J, Stepaniants S, Zhang C, Meng Q, Peters M, He Y, Ni C, Slipetz D, Crackower MA, Houshyar H, Tan CM, Asante-Appiah E, O'Neill G, Luo MJ, Thieringer R, Yuan J, Chiu CS, Lum PY, Lamb J, Boie Y, Wilkinson HA, Schadt EE, Dai H, Roberts C: **Systems analysis of eleven rodent disease models reveals an inflammotome signature and key drivers.** *Mol Syst Biol* 2012, **8**:594.
18. Huang E, Ishida S, Pittman J, Dressman H, Bild A, Kloos M, D'Amico M, Pestell RG, West M, Nevins JR: **Gene expression phenotypic models that predict the activity of oncogenic pathways.** *Nat Genet* 2003, **34**:226–230.
19. Zhou X, Wu W, Hu H, Milosevic J, Konishi K, Kaminski N, Wenzel SE: **Genomic differences distinguish the myofibroblast phenotype of distal lung fibroblasts from airway fibroblasts.** *Am J Respir Cell Mol Biol* 2011, **45**:1256–1262.
20. Bruhn S, Fang Y, Barrenas F, Gustafsson M, Zhang H, Konstantinell A, Kronke A, Sonnichsen B, Bresnick A, Dulyaninova N, Wang H, Zhao Y, Klingelhofer J, Ambartsumian N, Beck MK, Nestor C, Bona E, Xiang Z, Benson M: **A generally applicable translational strategy identifies S100A4 as a candidate gene in allergy.** *Sci Transl Med* 2014, **6**:218ra4.
21. Pe'er D, Hachohen N: **Principles and strategies for developing network models in cancer.** *Cell* 2011, **144**:864–873.
22. McKinney EF, Lyons PA, Carr EJ, Hollis JL, Jayne DR, Willcocks LC, Koukoulaki M, Brazma A, Jovanovic V, Kemeny DM, Pollard AJ, Macary PA, Chaudhry AN, Smith KG: **A CD8+ T cell transcription signature predicts prognosis in autoimmune disease.** *Nat Med* 2010, **16**:586–591. 1p following 591.
23. Wang H, Gottfries J, Barrenas F, Benson M: **Identification of novel biomarkers in seasonal allergic rhinitis by combining proteomic, multivariate and pathway analysis.** *PLoS One* 2011, **6**:e23563.
24. Johnson WE, Li C, Rabinovic A: **Adjusting batch effects in microarray expression data using empirical Bayes methods.** *Biostatistics* 2007, **8**:118–127.
25. Irizarry RA, Wang C, Zhou Y, Speed TP: **Gene set enrichment analysis made simple.** *Stat Methods Med Res* 2009, **18**:565–575.
26. Knijnenburg TA, Wessels LF, Reinders MJ, Shmulevich I: **Fewer permutations, more accurate P-values.** *Bioinformatics* 2009, **25**:161–8.
27. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc* 1995, **57**:289–300.
28. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, Deshpande N, Suresh S, Rashmi BP, Shanker K, Padma N, Niranjana V, Harsha HC, Talreja N, Vrushabendra BM, Ramya MA, Yatish AJ, Joy M, Shivashankar HN, Kavitha MP, Menezes M, Choudhury DR, Ghosh N, Saravana R, Chandran S, Mohan S, Jonnalagadda CK, Prasad CK, et al: **Human protein reference database as a discovery resource for proteomics.** *Nucleic Acids Res* 2004, **32**:D497–501.
29. Milacic M, Haw R, Rothfels K, Wu G, Croft D, Hermjakob H, D'Eustachio P, Stein L: **Annotating cancer variants and anti-cancer therapeutics in reactome.** *Cancers (Basel)* 2012, **4**:1180–1211.
30. Yu H, Tardivo L, Tam S, Weiner E, Gebreab F, Fan C, Svrzikapa N, Hirozane-Kishikawa T, Rietman E, Yang X, Sahalie J, Salehi-Ashtiani K, Hao T, Cusick ME, Hill DE, Roth FP, Braun P, Vidal M: **Next-generation sequencing to generate interactome datasets.** *Nat Methods* 2011, **8**:478–480.
31. **CCSB Interactome Database.** [<http://interactome.dfci.harvard.edu/>].
32. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H: **Three-dimensional reconstruction of protein networks provides insight into human genetic disease.** *Nat Biotechnol* 2012, **30**:159–164.
33. Kanehisa M: **Molecular network analysis of diseases and drugs in KEGG.** *Methods Mol Biol* 2013, **939**:263–275.
34. **Ingenuity Pathway Analyses (IPA).** [<http://www.ingenuity.com/>].
35. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology.** *The Gene Ontology Consortium. Nat Genet* 2000, **25**:25–29.
36. **Mammalian Phenotype database.** [<http://informatics.jax.org/pub/reports/index.html>].
37. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson JA, Barthorpe S, Lutz SR, Kogera F, Lawrence K, McLaren-Douglas A, Mitropoulos X, Mironenko T, Thi H, Richardson L, Zhou W, Jewitt F, et al: **Systematic identification of genomic markers of drug sensitivity in cancer cells.** *Nature* 2012, **483**:570–575.
38. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data.** *Nucleic Acids Res* 2003, **31**:e15.
39. Barrenas F, Chavali S, Alves AC, Coin L, Jarvelin MR, Jornsten R, Langston MA, Ramasamy A, Rogers G, Wang H, Benson M: **Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms.** *Genome Biol* 2012, **13**:R46–2012. 13–6–46.
40. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C: **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.** *Nucleic Acids Res* 2011, **39**:D561–568.
41. Zhao Y, Wang H, Gustafsson M, Muraro A, Bruhn S, Benson M: **Combined multivariate and pathway analyses show that allergen-induced gene expression changes in CD4+ T cells are reversed by glucocorticoids.** *PLoS One* 2012, **7**:e39016.
42. Wang H, Barrenas F, Bruhn S, Mobini R, Benson M: **Increased IFN-gamma activity in seasonal allergic rhinitis is decreased by corticosteroid treatment.** *J Allergy Clin Immunol* 2009, **124**:1360–1362.
43. Benson M, Langston MA, Adner M, Andersson B, Torinsson-Nalua A, Cardell LO: **A network-based analysis of the late-phase reaction of the skin.** *J Allergy Clin Immunol* 2006, **118**:220–225.
44. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.** *Proc Natl Acad Sci USA* 2009, **106**:9362–9367.
45. Zhu J, Paul WE: **Heterogeneity and plasticity of T helper cells.** *Cell Res* 2010, **20**:4–12.
46. Engelbertsen D, Andersson L, Ljungcrantz I, Wigren M, Hedblad B, Nilsson J, Bjorkbacka H: **T-helper 2 immunity is associated with reduced risk of myocardial infarction and stroke.** *Arterioscler Thromb Vasc Biol* 2013, **33**:637–644.
47. Hung K, Hayashi R, Lafond-Walker A, Lowenstein C, Pardoll D, Levitsky H: **The central role of CD4(+) T cells in the antitumor immune response.** *J Exp Med* 1998, **188**:2357–2368.
48. Cheng X, Wang J, Xia N, Yan XX, Tang TT, Chen H, Zhang HJ, Liu J, Kong W, Sjoberg S, Folco E, Libby P, Liao YH, Shi GP: **A guanidine-rich regulatory oligodeoxynucleotide improves type-2 diabetes in obese mice by blocking T-cell differentiation.** *EMBO Mol Med* 2012, **4**:1112–1125.
49. Palleja A, Horn H, Eliasson S, Jensen LJ: **Distill Database: diseases and traits in linkage disequilibrium blocks.** *Nucleic Acids Res* 2012, **40**:D1036–1040.
50. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA: **COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer.** *Nucleic Acids Res* 2011, **39**:D945–950.
51. **Online Mendelian Inheritance in Man.** [<http://omim.org/>].
52. International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S, Gray E, Booth DR, Potter SC, Goris A, Band G, Otrai AB, Strange A, Saarela J, Bellenguez C, Fontaine B, Gillman M, Hemmer B, Gwilliam R, Zipp F, Jayakumar A, Martin R, Leslie S, Hawkins S: **Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis.** *Nature* 2011, **476**:214–219.
53. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS: **DrugBank 3.0: a comprehensive resource for 'omics' research on drugs.** *Nucleic Acids Res* 2011, **39**:D1035–1041.
54. Sugaya N, Kanai S, Furuya T: **Dr. PIAS 2.0: an update of a database of predicted druggable protein-protein interactions.** *Database (Oxford)* 2012, **2012**:bas034.
55. Mittelbrunn M, Molina A, Escribese MM, Yanez-Mo M, Escudero E, Ursa A, Tejedor R, Mampaso F, Sanchez-Madrid F: **VLA-4 integrin concentrates at the peripheral supramolecular activation complex of the immune**

- synapse and drives T helper 1 responses. *Proc Natl Acad Sci USA* 2004, **101**:11058–11063.
56. Lindberg RL, Achtrichts L, Hoffmann F, Kuhle J, Kappos L: **Natalizumab alters transcriptional expression profiles of blood cell subpopulations of multiple sclerosis patients.** *J Neuroimmunol* 2008, **194**:153–164.
57. Wagner GP, Zhang J: **The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms.** *Nat Rev Genet* 2011, **12**:204–213.
58. Miettinen M, McCue PA, Sarlomo-Rikala M, Rys J, Czapiewski P, Wazny K, Langfort R, Waloszczyk P, Biernat W, Lasota J, Wang Z: **GATA3: a multispecific but potentially useful marker in surgical pathology: a systematic analysis of 2500 epithelial and nonepithelial tumors.** *Am J Surg Pathol* 2013, **13**:22.
59. Teoh H, Quan A, Creighton AK, Annie Bang KW, Singh KK, Shukla PC, Gupta N, Pan Y, Lovren F, Leong-Poi H, Al-Omran M, Verma S: **BRCA1 gene therapy reduces systemic inflammatory response and multiple organ failure and improves survival in experimental sepsis.** *Gene Ther* 2013, **20**:51–61.
60. Bousquet J, Anto JM, Sterk PJ, Adcock IM, Chung KF, Roca J, Agusti A, Brightling C, Cambon-Thomsen A, Cesario A, Abdelhak S, Antonarakis SE, Avignon A, Ballabio A, Baraldi E, Baranov A, Bieber T, Bockaert J, Brahmachari S, Brambilla C, Bringer J, Dauzat M, Ernberg I, Fabbri L, Froguel P, Galas D, Gojobori T, Hunter P, Jorgensen C, Kauffmann F, *et al*: **Systems medicine and integrated care to combat chronic noncommunicable diseases.** *Genome Med* 2011, **3**:43.

doi:10.1186/gm534

**Cite this article as:** Gustafsson *et al.*: Integrated genomic and prospective clinical studies show the importance of modular pleiotropy for disease susceptibility, diagnosis and treatment. *Genome Medicine* 2014 **6**:17.

**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](http://www.biomedcentral.com/submit)

