

Global Map of Physical Interactions among Differentially Expressed Genes in Multiple Sclerosis Relapses and Remissions

Tamir Tuller^{1,2*}, Shimshi Atar², Eytan Ruppin^{1,3}, Michael Gurevich^{1,2&}, Anat Achiron^{1,2&}

¹School of Medicine, Tel Aviv University, Ramat Aviv, Israel; ²Multiple Sclerosis Center, Sheba Medical Center, Tel Hashomer, Israel; ³School of Computer Science, Tel Aviv University, Ramat Aviv, Israel

& Equal contribution

*Contact: tamirtul@post.tau.ac.il (TT).

Abstract

Multiple Sclerosis (*MS*) is a central nervous system autoimmune inflammatory T-cell mediated disease with a relapsing-remitting course in the majority of patients. In this study we performed a high resolution systems biology analysis of gene expression and physical interactions in *MS* relapse and remission. To this end, we integrated 164 large scale measurements of gene expression in Peripheral Blood Mononuclear Cells (*PBMC*) of *MS* patients in relapse or remission and healthy subjects, with large scale information about the physical interactions between these genes obtained from public databases. These data were analyzed with a variety of computational methods.

We find that there is a clear and significant global network-level signal that is related to the changes in gene expression of *MS* patients in comparison to healthy subjects. However, despite the clear differences in the clinical symptoms of *MS* patients in relapse vs. remission, the network level signal is weaker when comparing patients in these two stages of the disease. This result suggests that most of the genes have *relatively* similar expression levels in the two stages of the disease.

In accordance with previous studies we found that the pathways related to *Regulation of Cell Death*, *Chemotaxis*, and *Inflammatory response* are differentially expressed in the disease in comparison to healthy subjects, whilst pathways related to *Cell adhesion*, *Cell migration*, and *Cell-cell signaling* are activated in relapse in comparison to remission. However, the current study includes a detailed report of the exact set of genes involved in these pathways and the interactions between them. For example, we found that the genes *TP53* and *IL1* are 'network-hub' that interacts with many of the differentially expressed genes in *MS* patients vs. healthy subjects, and the

epidermal growth factor receptor *EGFR* is a 'network-hub' in the case of *MS* patients with relapse *vs.* remission.

The statistical approaches employed in this study enabled us to report new sets of genes that according to their gene expression and physical interactions are predicted to be differentially expressed in *MS vs.* healthy subjects, and in *MS* patients in relapse *vs.* remission. Some of these genes may be useful biomarkers for diagnosing *MS* and predicting relapses in *MS* patients.

Introduction

Multiple Sclerosis (*MS*) is an autoimmune inflammatory T-cell mediated disease, believed to result from a misdirected immune attack against Central Nervous System (*CNS*) myelin antigens. The damage of myelin in *MS* leads to neurological dysfunction (1-3). *MS* attacks mainly young adults (usually between ages 20 and 40), and is predominant in women. In northern Europe and the U.S.A. the disease has prevalence of around 100 per 100,000 (2). Thus, it has a significant socioeconomic impact.

The trigger of the autoimmune process in *MS* is unknown. It is believed that *MS* occurs as a result of some combination of genetic, environmental and infectious factors (4), and possibly other factors like vascular problems (5). For example, studies of identical twins have demonstrated a concordance of 30% to develop *MS* (3) suggesting that the genetic background has a relatively limited but significant role in triggering *MS*.

Symptoms of *MS* are unpredictable and include among others, motor weakness, loss of balance and muscle coordination, slurred speech, cognitive decline, sensory impairment, and bladder dysfunction.

In 85% of the patients, the disease has a relapsing-remitting (*RR*) course, which is characterized by the onset or deterioration of the neurological symptoms (relapses), which are followed by partial or complete recovery (remissions). Relapses are the consequence of complex immunological and neurodegenerative processes. They result in the development of new acute inflammatory lesions or the activation of old lesions within the *CNS*, and are associated with myelin and axonal loss (1, 2, 6).

Accordingly, relapses are the visible clinical expression of the complicated immunopathological mechanisms that occur in the *CNS*.

As was demonstrated in numerous previous studies, transcriptional profiling of *PBMC* is a useful tool for identifying gene expression signatures that are related to *MS* and other autoimmune diseases (for example (7-16); see a more general review in (17)). A possible explanation for the advantageousness of *PBMC* in this context is the fact that autoreactive immune cells initiate the autoimmune inflammatory process against the corresponding target organs (18-24).

The aforementioned studies concerning gene expression in *MS* were based solely on measurements of *mRNA* levels. The first stage of gene expression regulation is transcription (*i.e.* changes in *mRNA* levels). However, several steps in the gene expression process may be regulated, including transcription, post-transcriptional modifications such as *RNA* splicing, translation, and post-translational modifications of a protein (25). Thus, studying changes in gene expression should give a *very* limited and coarse depiction of the regulatory changes that occur in the disease. Indeed, it was shown that in human and other organisms, the gene *mRNA* levels explain only about 30% of their protein abundance variance (26, 27). In addition, *mRNA* measurements usually suffer from various elements of noise and bias (28).

The goal of the current study is to improve the understanding of the regulatory changes in *PBMC* in *MS*, by incorporating in addition to gene expression, prior knowledge about protein-protein (*PP*) interactions in human. By integrating this additional data we were able to identify genes that undergo regulatory changes in *MS*, which are not necessarily at the transcription level. It has been shown that such an

approach is useful for identifying markers related to metastasis (29) and yields better results compared to analyses based solely on gene expression; the current paper is the first time that such an approach was employed for the analysis of an autoimmune disease. In autoimmune diseases, as in cancer metastasis, many of the regulatory changes are posttranscriptional; thus, such an approach should be beneficial in both cases.

We identified proteins that have many *PP* interactions with differentially expressed genes (in terms of *mRNA* levels) in *MS* compared to healthy subjects, and in *MS* relapse as compared to *MS* remission. Such genes have not previously been reported in large scale studies of *mRNA* levels in *MS*.

With high probability such genes undergo transcriptional and/or post transcriptional regulation. Thus, our approach enables uncovering suspected transcriptional and post-transcriptional regulatory changes in the disease.

Indeed, as demonstrated in the following sections, we supply a more accurate and comprehensive report of the genes and pathways involved in *MS* than previously reported.

Results

We analyzed a *DNA*-microarray gene expression dataset that included a total of 164 subjects. To detect genes that are differentially expressed (in terms of *mRNA* levels) in the disease, we employed a statistical method which takes into account their demographical parameters and batch effects (Methods). Next, to better understand the molecular mechanisms related to *MS*, the data of the differentially expressed genes was

integrated with large scale measurements of protein-protein interactions (*PPI*) and protein-*DNA* interactions (*PDI*, Methods, Figure 1).

At all stages, we analyzed two datasets. Each dataset was related to a comparison of two groups of patients or healthy subjects:

1) The first dataset included patients with *MS* (relapse or remission, 123 patients) that were compared to healthy subjects (41 subjects). This dataset was utilized to understand regulatory processes that appear in *both* relapse and remission stages of *MS*.

2) The second dataset included *MS* patients with relapse (34 patients) that were compared to *MS* patients with remission (75 patients with active disease, $EDSS > 0$, Methods). This dataset was used to better understand the regulatory processes that are *specific* to each stage of the disease.

We named the first dataset *MS/healthy* and the second dataset *Rel/Rem*. The properties of these datasets (*e.g.* age, gender, disease duration of the analyzed patients) appear in Table 1 (see also Supplementary Table 1).

Differentially expressed genes in *MS* and their projection on the protein-protein and protein-*DNA* interaction networks – a global (network-level) view.

Gene expression data was normalized, statistically over/under expressed genes were found in each dataset based on Analysis of Variance (*ANOVA* (30)), and the *False Discovery Rate* (*FDR*; (31) see the Methods section for more technical details regarding the normalizations). We found that 1268 and 938 genes were significant according to *ANOVA* ($p\text{-value} \leq 0.05$) in the *MS/healthy* and the *Rel/Rem* databases respectively.

In both datasets, no gene passed the *FDR* criterion based on the *ANOVA* p-values (see Table 2 for the list of genes with the most significant *ANOVA* p-value; see also Supplementary Table 2-3 for the entire list of p-values) demonstrating the essentiality

of integrating additional information sources in-order to augment the differentially expressed genes signature .

At the next stage, we computed for each gene a p-value that was based on the number and expression levels of genes that have protein interaction with it (but not including *mRNA* levels of the gene itself). We named this p-value *PPI* p-value. Genes that have a relatively high number of protein-protein interactions with differentially expressed genes have a higher probability to undergo regulatory changes, possibly *post transcriptional*, themselves and thus have more significant *PPI* p-values.

We found that 398 and 257 genes were *PPI* significant (p-value < 0.05) in the *MS/healthy* and *Rel/Rem* databases respectively (see Supplementary Table 2-3 for the entire list of p-values).

Figure 2 includes a global view of the projection of the differentially expressed genes on the protein-protein interaction network. In the remainder of this section, we report several *global* p-values that are related to the expression levels of the *entire* set of genes and the physical interactions between them. The p-values were based on comparisons to random networks with similar properties to the original ones (Methods).

The first global p-value tests if differentially expressed genes tend to be close to each other in the protein-protein interaction network, as expected for the real biological signal. We found that this global p-value was significant in the case of the *MS/Healthy* dataset (p-value = 0.05) but was not significant in the case of the *Rel/Rem* dataset (p-value = 0.2). The result suggests that the two stages of the disease are *relatively* similar in terms of the gene expression signature; thus, in general, the differentially expressed genes in the *Rel/Rem* database do not seem to be clustered in the protein interaction network.

In the second global p-value we checked if the number of genes with significant *PPI* p-values is higher than in random networks with similar properties (Methods). This number was significantly higher than expected in randomized networks for the *MS/healthy* data set (275 in the real data versus 88 in the randomized data; p-value 0.01) but non-significant in the *Rel/Rem* dataset (Methods).

The third global p-value was related to the number of significantly expressed genes (according to the *ANOVA* p-value) that *also* have *PPI* with other significantly expressed genes (see exact details in the Methods section). We found that in both datasets this number was higher than expected from random permutations with similar properties (69 genes in the real dataset versus a mean of 34 in the randomized gene networks for the *MS/healthy* dataset, p-value <0.01; and 40 genes in the real dataset versus a mean of 22 in the random networks, p-value <0.05 for the *Rel/Rem* dataset; Methods). Thus we deduce that p-values based on more refined measures do detect significant global changes also in the *Rel/Rem* dataset. In the following sections we report additional, more specific and highly significant p-values, related to functional groups of genes and single genes, which are differentially expressed in the *Rel/Rem* dataset.

Protein–DNA interaction analysis of 111 transcription factors (*TFs*) and their targets (a total of 5787 interactions, Methods) revealed that 8 and 16 *TFs* changed their expression according to the *ANOVA* p-value in the *Rel/Rem* dataset and the *MS/Healthy* dataset respectively. We found 31 and 119 pairs of *TFs* and their targets that are both *ANOVA* significant in the *Rel/Rem* dataset and in the *MS/Healthy* dataset respectively. The number of such pairs was significantly higher compared to randomized networks

(Methods) in the case of the *MS/Healthy* dataset (mean number of genes in the real data: 119 vs. mean number of genes in the randomized data: 53; p-value = 0.03) but not in the *Rel/Rem* dataset. The resultant lists of genes that are differentially expressed and have differentially expressed *TFs* appear in Supplementary Tables 2-3.

Enriched *GO* groups for genes with significant *ANOVA* or *PPI* p-values

To gain a picture of the cellular processes that are specific to *MS* and the cellular processes that are differentially expressed in relapse vs. remission, we performed a *GO* (Gene Ontology) enrichment analysis of the genes with significant *ANOVA* p-values and *PPI* p-value (Methods). Some cellular functions with the highest enrichment p-values, which pass *FDR*, are depicted in Table 3 (all the enriched cellular functions can be found in Supplementary Tables 4-7).

In the *MS/healthy* dataset, the list of enriched cellular processes that were based on the genes with *ANOVA* significant p-values includes the following functional groups: *Positive regulation of cell death*, *Chemotaxis*, *Inflammatory response* (all p-values < 0.0014; see Table 3). These cellular processes are known to be key mechanisms in *MS*.

Almost all the cellular functions that were enriched according to genes with *ANOVA* significant p-values were also enriched based on genes with *PPI* significant p-values. However, the enrichment results according to the *PPI* significant p-values included additional relevant cellular functions that were not discovered by the *ANOVA* based enrichment. For example, the cellular functions *positive regulation of NF-kappaB transcription factor activity*, *immune system development*, and *regulation of caspase activity* (all the p-values < 0.0021; see Table 3) were uncovered only by the *PPI* based enrichment.

In the case of the *Rel/Rem* dataset, the following cellular functions were enriched based on the genes with *ANOVA* significant p-values: *Cell adhesion*, *Cell migration*, *Cell-cell signaling*, *Regulation of cell motion* (all the p-values < 0.0007; see Table 3).

A similar analysis based on genes with significant *PPI* p-values again revealed additional relevant cellular functions:

Positive regulation of cell proliferation, *T cell receptor signaling pathway*, *T cell activation*, *Regulation of T cell proliferation*, *T cell differentiation*, and *positive regulation of natural killer cell mediated cytotoxicity*, and *Regulation of apoptosis* (all the p-values < 0.00245; see Table 3).

These results demonstrate that many of the regulatory changes in *MS* cannot be detected based solely on gene *mRNA* levels; however, incorporating protein-protein interactions can further the discovery of *additional* relevant regulatory changes in *MS*.

Genes with significant *PPI* p-values

The genes with the highest *PPI* p-values appear in Table 4 (see the entire list in Supplementary Tables 2-3). In this case many of the p-values passed the *FDR* threshold (16 genes for the *Rel/Rem* dataset and 11 for the *MS/healthy* dataset). Several of the genes that passed the *FDR* test are known to be key-apoptotic genes such as *TP53*, *TRAF1* that mediate the anti-apoptotic signals from *TNF* receptors, and the anti-apoptotic gene *JUND*. Other genes that passed the *FDR* are related to growth and proliferation (for example, *FOS* and *BRCA1*), and transcription (*e.g.* *TBP*, *SMAD3*, *JUN*).

The list of genes that passed the *FDR* test according to the *PPI* p-value in the *Rel/Rem* dataset included many genes that related to phosphorylation (for example, *PTPRD*, *PTPRF*, *PPFIA1*, *PTPRS*, *PTPRA*, *PPFIA2*).

Interestingly, when we focused on genes with known *SNPs* that are associated with *MS* (32) (*HLA-DRB1*, *IL2RA*, *CD58*, *IL7R*, *PLP1*, *MAG*, *MOG*), we discovered that many of them have significant p-values (*ANOVA* or *PPI*) in at least one of the analyzed databases (Table 4); specifically, we found enrichment in the number of genes with *SNPs* associated with *MS* among the *PPI* significant genes in the *Rel/Rem* dataset (hyper geometric p-value = 0.001).

One interesting gene is *IL7R* which is a receptor for interleukine 7 (*IL7*). *IL7* has been shown to play a critical role in the *V(D)J* recombination during lymphocyte development (see (33)). This protein also controls the accessibility of the *TCR* gamma locus by *STAT5* and histone acetylation (see (34)). Knockout studies in mice have suggested that blocking apoptosis is an essential function of this protein during differentiation and activation of T lymphocytes (see (35)). We found that *IL7R*, which does not have a significant *ANOVA* p-value, has a significant *PPI* p-value in the *Rel/Rem* dataset (p-value = 0.00003), but not in the *MS/healthy* dataset. Thus, this result may suggest that protein levels of *IL7R* tend to change between relapse and remission periods of *MS*, even though no change is observed at the *mRNA* level of this gene.

Sub-networks of genes with both *ANOVA* and *PPI* significant p-values

Next we analyzed the specific set of genes that have *both ANOVA* and *PPI* significant p-values. We also considered the protein-protein interactions between these genes (Methods). The resultant protein interaction networks corresponding to these genes appear in Figure 3 (*MS/healthy*) and Figure 4 (*Rel/Rem*; see also Supplementary Tables 8-9). As can be seen in Figure 3, the *MS/healthy* network includes genes related to

transcription, regulation of proliferation and apoptosis, inflammatory response, response to cytokine stimulus, and T cell activation (all p-values < 0.004). The *Rel/Rem* network (Figure 4), on the other hand, includes genes related to *protein amino acid dephosphorylation, and cell adhesion* (all p-values < 0.003).

Clusters of genes with both ANOVA and PPI significant p-values

We clustered the sub-networks mentioned in the previous sub-section to find modules (clusters) of genes with relatively more protein-protein interactions between them ((36 Methods). The resultant networks and the clusters appear in Figure 3 (for the *MS/healthy* dataset) and Figure 4 (for the *Rel/Rem* dataset). We verified that the modularity of the resultant clustering is higher than in random networks with similar properties (p-value < 0.01; see details in the Methods section) suggesting that these clusters indeed have biological significance.

The *MS/healthy* network includes many apoptotic genes and modules that will be discussed in details in the following section. One striking module includes the cytokine *IL1B* and its receptor, which are over-expressed in *MS*. *IL1* is known to have an important role in activation of lymphocyte proliferation. Specifically, it induces *TNF-alpha* release by endothelial cell, proliferation of *CD4+* cells, *IL-2* production, co-stimulates *CD8+/IL-1R+* cells, induces proliferation of mature *B-cells* and immunoglobulin secretion, induces *IL-6* and *GCSF* secretion, and stimulates expression of *Fibroblast Growth Factor (FGF)* and *Epidermal Growth Factor (EGF)*. In addition, it was found that MS patients tend to have polymorphisms in this gene and its receptor (37-39).

The *Rel/Rem* network includes modules related to phosphorylation, the growth factor receptors (*EGFR*), and cell adhesion.

The epidermal growth factor receptor (*EGFR*), a receptor tyrosine kinase, is a hub in one of the modules. *EGFR* (and some of the proteins interacting with it) is over expressed (in terms of *mRNA* levels) in relapse relative to remission. *EGFR* signaling is initiated by ligand binding to the extra-cellular domain of *EGFR*, and following activation, phosphorylation of cytoplasmic substrates occurs, and a signaling cascade is launched. As a result many cellular responses are derived. Specifically, there are changes in gene expression, cytoskeletal rearrangement, anti-apoptosis, increased cell proliferation, and adhesion. The *EGFR* cluster is enriched with genes related to adhesion (p-value 0.03; Figure 4); thus, the genes that appear in the *EGFR* cluster suggest that in our context (relapse vs. remission in MS), *EGFR* mainly plays an important role in adhesion.

Pathways analysis of genes with significant p-values (ANOVA and PPI)

In order to better understand the molecular mechanisms that are involved in *MS/healthy* signatures, we combined the genes with significant *ANOVA p-values* with those with significant *PPI p-values*, and analyzed them with Ingenuity software (<http://www.ingenuity.com>). We considered the canonical pathways of Ingenuity and found that the most striking biological pathways that were enriched with these genes were related to *IL1-IL8 induced inflammation* ($p=4.7*10^{-19}$; Figure 5A) and *suppression of FASL and P53 dependent apoptosis* ($p=1.1*10^{-12}$; Figure 5B). *IL-1* is produced by activated macrophages, B cells and fibroblasts: it is a pivotal pro-inflammatory cytokine that is centrally involved in local and systemic responses in the immune system, which lead to typical effects of inflammation. Its deregulation, prolonged synthesis and release in chronic inflammation contribute to diseases such as multiple sclerosis. There are two forms of *IL-1* encoded by

distinct genes, *IL-1 α* and *IL-1 β* , both of them were found to be over-expressed in the MS/healthy signature. The *IL 1* receptor has two subunits, *IL-1RI* and *IL-1RAP*, that upon ligand binding form a complex; in accordance with the previous results, the genes encoding *IL 1* receptor were also up-regulated.

Binding of *IL-1* to its cell surface receptor activates G-proteins, which in turn stimulate adenylate cyclase activity, which leads to an increase in the intracellular level of *cAMP*. *Cyclic AMP* activates *PKA* which then activates the pro-inflammatory *NF- κ B* pathway as demonstrated by the significant p-value of the two subunits of *NFKB1*. Activated *NF- κ B*, *JNK1* and *p38MAPK* are translocated to the nucleus where, either directly or via the transcriptional regulators *c-Jun* and *c-Fos* (both found to be over-expressed), they induce genes that regulate inflammation.

Another mechanism of *NF- κ B*, *c-Jun*, and *c-Fos* activation involves the recruitment of *IRAK* and *TRAF6* (Figure 5A).

Interleukin 8 (*IL-8*) is a member of the *C-X-C* family of chemokines, which plays a central role in inflammation. Activation by *IL-8* can trigger inflammation in cells like neutrophils, which leads to chemotaxis, granule release, and increased cell adhesion. Interestingly, over-expressed *IL-8* similarly to *IL1*, can also induce nuclear transcription factor-kappa B (*NF- κ B*) through a *TRAF6*-dependent pathway, and activate cyclooxygenase (*PTGS2*) and prostaglandin type of inflammation, that we found in the MS/healthy signature. This last mechanism is also supported by the significant over-expression of *IL1A* and *Apl* genes. In our context, the most imperative result of *IL8* activation is the up-regulation of matrix metalloproteinase (*MMP2* and *MMP9*), which is involved in the blood brain barrier damage mechanism in multiple sclerosis. The activation of inflammatory cellular response in the MS/healthy signature was also demonstrated by the over-expression of the *CXCR4* signaling mechanism. As can be seen in Figure 5A; the *CXCR4* receptor and its specific *CXCL12* chemokine were over-expressed. This over-

expression should lead to an increased phosphorylation of multiple focal adhesion components.

The down-regulation of apoptotic mechanisms in multiple sclerosis have been demonstrated in previous studies (see, for example, (40)). Here by combining gene expression and protein-protein interactions we were able to reveal more precisely the mechanisms of apoptosis deficiency in the *MS/healthy* signature (Figure 5B). We found that *FASL*, which mediates apoptosis, and its effector *CASP8* proteolytic protein, were suppressed in *MS*. The downstream apoptotic cascade includes the genes *Daxx*, *JNK*, *BcL-XL*, *CIAP*, *CASP3* and *CAD*, which had significant *PPI* p-values. Another apoptotic pathway which appears in the *MS/healthy* signature is initiated by the binding of *TNF* family ligands to their receptors inefficiently due to under-expression of *TRAF1/2* adaptor protein, which facilitate binding to pro-caspase 8 and pro-caspase (Figure 5B). In addition, we found clues that suggest that the *p53* signaling pathway is also suppressed in *MS*: first, as we previously mentioned, we found that *p53* is under-expressed; second, we found that *p53* masters regulatory proteins such as *MDM2*, *WTN1*, *p300*, *PCAF*, which have highly significant *PPI* p-values; third, we found that *cJUN*, that is known as an antagonist of *p53*, is over-expressed in *MS*. Downstream *negative* regulators of cell cycle, like retinoblastoma 1 (*RB*), were also over-expressed. Thus, the suppression of immune cell's apoptotic mechanisms could lead to expansion of autoimmune cell clones and maintenance of autoimmunity.

The *Rel/Rem* signature (Figure 5C) characterized by activation of lymphocyte migration mechanisms ($p < 5.7 \cdot 10^{-4}$) includes a large group of significantly over-expressed cytokines and chemokines (according to the *ANOVA* test). Specifically, the set of over-expressed genes in the *Rel/Rem* dataset included the protein *CCL14*, a

cytokine which induces alterations in intracellular calcium concentrations, enzymes released in monocytes and the chemokine *CCBP2*. These proteins physically interact with each other and are critical for the recruitment of effector immune cells to the inflammation site (Figure 5C, see, for example, (41, 42)).

The activation of chemotaxis of lymphocytes and monocytes is demonstrated by the over-expression of members of the *MIP1* family such as *CCL15*, which induces changes in intracellular calcium concentrations, and acts via the *CCR1* receptor (Figure 5C; see, for example, (41, 42)).

An important chemokine which is over-expressed in *Rel/Rem* is *CCL22* (Figure 5C). It plays a role in trafficking dendritic cells and natural killer cells to the inflammatory sites of monocytes, and chronically activated T lymphocytes. It also displays a mild activity in primary activated T lymphocytes, though has no chemoattractant activity for neutrophils, eosinophils and resting T lymphocytes (see, for example, (41, 42)).

Another piece of evidence that supports lymphocyte activation in *Rel/Rem* is the over-expression of *CCL12* (Figure 5C), which is known to play a role in the pathogenesis of diseases that are characterized by monocytic infiltrates, such as psoriasis, rheumatoid arthritis and multiple sclerosis (see, for example, (43-45)).

In addition we found that the *VNT* molecule, which is found in serum and tissues, and promotes cell adhesion and spreading, is over-expressed in acute relapse (Figure 5C).

We also uncovered that acute *MS* relapse is characterized by stimulation of pro-inflammatory molecules such as *MMP1* and *ALOX15* (Figure 5C; Supplementary Table 3).

To complete the picture of lymphocyte activation and monocyte trafficking in inflammatory sites, we report a few additional relevant proteins with significant *PPI*

p-values (Figure 5C). This set of proteins includes (Figure 5C, Supplementary Table 3) *IL2R*, *CXCR4* – a receptor for *CCL12*, and the focal adhesion kinase (*P38MAPK*). In addition, it incorporates *CD44* (Figure 5C), a cell-surface glycoprotein that is involved in cell adhesion and migration, as well as interaction with other ligands, such as osteopontin, collagens, and *MMPs*. *CD44* also binds to chemokines in the intracellular matrix. Interestingly, the receptor of *TGFB* (*PDGFRA* (46)), which is an upstream regulator that increases the expression of *CCL12* and *CCBP2*, was over-expressed in the *Rel/Rem* database as well (Figure 5C).

Discussion

In this study, we reported the first systems biology study of gene expression in *MS* patients under remission and relapse, encompassing 164 large scale measurements of gene expression in *PBMC*, including clinical and demographical characteristics of the patients, batch effects, and physical interactions between proteins. Using this approach we were able to report the regulatory changes that occur in the disease with a much higher resolution than before (see, for example, (8, 12)). The results that are reported in this study can be classified to three major levels of resolution: global, pathways and gene groups, and single genes.

At the global level, when considering the expression levels of all the genes and the physical interactions between them, we show that there is a strong global signal related to the disease in the *MS/Healthy* dataset. Even though there are clear differences between the two disease stages (relapse and remission), and despite the fact that there are specific genes and pathways that are differentially expressed between the two stages of the disease, we discerned a considerably weaker global signal in the *Rel/Rem* dataset.

Thus, roughly speaking *most* of the *MS* gene expression signature remains *relatively* constant in relapses and remissions.

At the pathway and gene modules level, the current study includes a detailed report of modules and pathways that are involved in the disease (*MS/Healthy* dataset) and its two stages (*Rel/Rem* dataset), including the exact set of relevant proteins, and the interaction mechanisms amongst them (see Figures 3-5 that includes only genes with significant *PPI* and/or *ANOVA* p-values). Previous papers (*e.g.* (8, 12)) included only a high level report of the pathways that appear to be related to the disease (*e.g.* T-cell activation, apoptosis, and inflammation). Thus, the pathways outlined in this paper improve our understanding of the disease phenotypes at the signaling pathways level. For example, we found that epidermal growth factor receptor (*EGFR*) is a 'hub' which interacts with many *ANOVA* significant genes in the *Rel/Rem* database, possibly to promote improved adhesion. Thus, it is plausible that the gene expression of the *EGFR* sub-network is a good candidate for relapse prediction.

At the single gene level, we employ novel statistical analyses that overcome some of the biases related to gene expression analysis, and the fact that many of the regulatory mechanisms are not transcriptional. Thus, we detail a robust list of potential biomarkers of the disease in *PBMC* (see Tables 2,4 and supplementary tables 1-3,8-9). To the best of our knowledge, many of these genes are novel in this context, and have not been reported in previous studies (*e.g.* (8, 12)). For example, to the best of our knowledge, the genes *SMAD3*, *RBI*, *FOS*, *TP53*, *BRCA1*, *TBP* with significant p-values in the *MS/healthy* dataset (see Table 2), and the genes *PTPRD*, *PDGFRB*, *STAT5A*, *PTPRF*,

PPFIA1, *PTPRS*, *PPFIA2*, *SCNN1B* with significant p-values in the *Rel/Rem* dataset, (see Table 2) are novel potential biomarkers in the *MS* context.

Finally, since gene expression in *PBMC* has been utilized in the study of other autoimmune diseases (see, for example, (7, 15, 16)), the approach employed in this study can be useful in the case of these diseases as well.

Material and Methods

Population Study

The clinical and demographical characteristics of each of the analyzed dataset (*MS/Healthy* and *Rel/Rem*) appear in Table 1.

In the *Rel/Rem* dataset we did not include patients whose Expanded Disability Status Scale (*EDSS*) (47) was zero.

All the *MS* patients were diagnosed with definite *MS* according to the McDonald criteria (48). *MS* relapse was defined as the onset of new objective neurological symptoms/signs or the aggravation of existing neurological disability, not accompanied by metabolic changes, fever or other signs of infection, and lasting for a period of at least 48 hours accompanied by objective change of at least 0.5 in the *EDSS* score. Confirmed relapses and *EDSS* scores were recorded consecutively.

RNA isolation and microarray expression profiling in MS

PBMC were separated on Ficol-Hypaque gradient, total *RNA* was purified, labelled, hybridized to a Genechip array (HU133A-2) and scanned (Hewlett Packard,

GeneArray-TM scanner G2500A) according to the manufacturer's protocol (Affymetrix Inc, Santa Clara, CA, USA).

The normalization and analysis of the gene expression datasets

We used the Sheba *MS* center recorded computerized clinical follow up and blood gene expression measurements dataset. This dataset includes information regarding clinical variables such as: age, gender, and *EDSS* at time of blood sampling.

The following steps of the data analysis were performed by *MATLAB*:

1. Expression values were computed from raw CEL files by applying the Robust Multi-Chip Average (*RMA*) background correction algorithm. We averaged the expression levels of all the probes of each gene, and each of the datasets underwent quantile normalization.
2. In the next step, in each of the datasets (*MS/Healthy* and *Rel/Rem*) we computed significantly over/under expressed genes based on *ANOVA* (30), considering batch effects such as the scan date of each chip, and the age and gender of each patient.

The drug regimen of each patient

Supplementary Table 1 includes the drug regimen of each patient during the month of blood withdrawal (last column). We found that differing treatment is not a variable that has a significant effect on the gene expression pattern: The Spearman ranked correlation between the *ANOVA* p-values of all the genes when considering only the untreated patients vs. the *ANOVA* p-values when considering *all* the patients is close to perfect ($r = 0.943$); i.e. genes that are relatively more significant based on the

untreated group are also more significant based on the entire group of patients (roughly speaking the ranking of the genes' p-values does not change).

However, when we considered all the patients there we uncovered 16% more significant genes (the statistical power increases). Thus, we decided to employ all the patients.

The Protein-Protein interaction network of Human

The human protein-protein interaction network was gathered from public databases (49, 50) and from recently published papers (51, 52). The final reconstructed network includes 7915 proteins and 28972 protein-protein interactions (a subset of 6850 proteins and 25931 protein-protein interactions appears in the analyzed chips).

Protein-DNA interaction network of Human

This network was downloaded from the *TRED* database (53). It includes 112 *TFs*, 2964 proteins that are regulated by these *TFs*, and a total of 7298 protein-DNA interactions. When considering only the genes that appear on the analyzed Genechip array, there are 111 *TFs*, 6069 proteins that are regulated by the *TFs* and a total of 5787 protein-DNA interactions.

Gene Specific P-values based on the Protein-Protein Interaction

Network

For each protein in the protein-protein interaction network, we computed a hypergeometric p-value that was based on the number of *ANOVA* significant proteins that

interact with it, the total number of *ANOVA* significant, and the topology of the protein-protein interaction network. Roughly speaking, a protein that interacts with a larger number of interacting genes that are *ANOVA* significant proteins will have a more significant p-value. We named this p-value *PPI* (protein-protein interaction network p-value) and calculated it as follows:

Let n_i denote the number of proteins interacting with gene i , let m_i be the number of protein interactions with gene i that have significant *ANOVA* p-values, let N be the total number of proteins in the network, and let M be the total number of proteins with significant *ANOVA* p-values in the network. At the first stage, we computed for each gene an initial p-value; let *PPI0* denote this p-value. The initial p-value of gene i is:

$$\sum_{j=m_i}^{n_i} \binom{M}{j} * \binom{N-M}{n_i-j} / \binom{N}{n_i}$$

To get a more refined p-value we repeated iteratively (till convergence) on the procedure described above in the following way:

In iteration k compute for each gene a *PPIk* p-value that is based on the *PPIk-1* p-values computed in the previous iteration using the formula above, with the only change being we replace *ANOVA* significant genes with genes that are *PPIk-1* significant (*i.e.* M is the total number of proteins with significant *PPIk-1* p-values and m_i is the number of protein interactions with gene i that have significant *PPIk-1* p-value).

In this paper we considered three sets of genes:

A) Genes that satisfy two conditions: 1) significant *ANOVA* p-value and 2) significant *PPI*-network p-value. Such genes appear in Figures 3-4.

B) Genes that satisfy only condition 2). Such genes appear in the left sides of Tables 2 and 3. C) Genes that satisfy only condition 1). Such genes appear the right sides of Tables 2 and 3.

A global p-value related to the total number of protein-DNA interactions

In this subsection, we describe a global p-value which is related to the number of genes with *ANOVA* significant p-values, such that each of them is regulated by at least one *ANOVA* significant Transcription Factor (*TF*).

The aforementioned global p-value related to the number of genes was computed as follows:

Repeat 100 times:

1. Randomly choose a subset of genes of size M .
2. Assume that the subset of genes that was chosen in 1 includes the *ANOVA* significant genes, and compute the number of genes that are *ANOVA* significant and are regulated by at least one *ANOVA* significant *TF*.
3. Compute the empirical probability (frequency) that the number of genes with *ANOVA* significant p-values that are regulated by at least one *TF* which is *ANOVA* significant, obtained in a random network, is larger (or equal) than the number of such genes in the original network.

A global P-value for the number of genes with significant *PPI*-network p-values

A global p-value which is related to the number of genes with significant *PPI*-network p-values was computed as follows:

Repeat 100 times:

1. Randomly choose a subset of genes of size M .
2. Assume that the subset of genes that was chosen in 1 includes the *ANOVA* significant genes, and compute for all genes the *PPI* p-values mentioned above.
3. Compute the empirical probability (frequency) that the number of genes with significant *PPI* p-values obtained in the random network is larger (or equal) than the number of genes with significant *PPI* p-values obtained in the original network.

A global P-value for the distance between *ANOVA* significant genes in the *PPI*- network

The aim of the global p-value described in this sub-section was to demonstrate that genes with *ANOVA* significant p-values tend to be close to other genes with *ANOVA* significant p-values in the protein-protein interaction network. Therefore the observed changes in the gene expression are not random.

This global p-value was computed as follows:

1. Find for each gene that is *ANOVA* significant, the distance to the closest gene in the protein-protein interaction network that is also *ANOVA* significant.
Compute the mean distance.

Repeat 100 times:

2. Randomly select a subset of M genes and assign them to the nodes of the *PPI*-network such that the degree distribution of these nodes will be identical to the degree distribution of *ANOVA* significant nodes in the original graph.
3. Find the mean distance between each of the M random *ANOVA* significant nodes and its closest neighbor, and compute the mean distance
4. Compute the empirical probability (frequency) that the random network has a smaller (or equal) mean distance than the mean distance in the original one.

Clustering the Differentially Expressed Networks

The clustering of differentially expressed networks was performed by the Newman algorithm (36) for finding communities in biological networks. The algorithm was implemented in *MATLAB*. This algorithm detects sub-sets (clusters) of proteins that include relatively many protein interactions between them (compared to other parts of the network). By definition, a network is modular if it can be divided to modules/sub-networks with a large number of protein-protein interactions between proteins within the same module, and less protein-protein interactions between proteins that reside in different modules.

We analyzed the protein-protein interaction network consisting *only* of genes with both significant *PPI* and *ANOVA* p-values, with the aforementioned Newman algorithm

We used the modularity score that is described in the work of Girvan and Newman,(36). To evaluate the significance of the modularity score of the resultant clustering, we compare it to the modularity score of randomized inputs as follows: Let n denote the number of nodes (proteins) in the original network. Let $u1, v1, u2,$ and $v2$ denote four proteins (nodes) in the network and let $(u1, v1)$ denote protein-protein interaction between proteins $u1$ and $v1$.

Each random network was generated in the following way:

1. Start with the original protein-protein interaction sub-network.
2. Repeat $10*n$ times on the following steps:
 - a. Choose a random pair of edges $(u1, v1)$ and $(u2, v2)$ in the network.
 - b. Replace them with a new pair of edges $(u1, v2)$ and $(u2, v1)$.

GO enrichment:

GO enrichment analysis of genes with significant *ANOVA* p-values and/or of genes with significant *PPI* p-values ($p < 0.05$ in both cases) was performed by David (54) (<http://david.abcc.ncifcrf.gov/>). In all cases, we considered *GO* groups with enrichment p-values that passed the False Discovery Rate (*FDR*). To this end we utilized the *FDRs* that were reported by David, and we reported only cases with *FDR* of less than 5%.

FDR test for the p-values of single genes

In this case, we considered the nonparametric approach of (31) with a threshold of $q = 0.05$.

Fold Change

To estimate the fold change in one condition versus the second (*e.g.* relapse *vs.* remission) while taking into account the batch effects and additional variables, we performed a multivariate linear regression (see, for example, (55)) in which the *Rel/Rem* (or *MS/control*) is the dependent variable and all the other (see table 1) previously mentioned variables (batch effects, clinical and demographical variables) are the independent variables. The scan date was represented by a set of dummy binary variables, and other variables were either continuous or binary. The sign of the coefficient related to the expression levels determined the fold change in one condition *vs.* the other. For example, in the case of the *Rel/Rem* database we set the dependent variable to be '1' in the case of relapse (and '0' otherwise). Thus, a positive coefficient of the expression levels variable corresponds to increased expression level of the gene in relapse in comparison to remission.

Similar results were obtained when we used partial correlations (Spearman or Pearson) between the *Rel/Rem* (or *MS/control*) variable and the expression levels, given all the other variables.

Acknowledgements

We thank Ms. Hadas Zur for helpful comments.

Conflict of Interest Statement

The authors declare that that they have no conflict of interests.

References

- .1 Waxman, S.G. (1998) Demyelinating Diseases--New Pathological Insights, New Therapeutic Targets. *N Engl J Med*, **338**, 323-5.
- .2 Noseworthy, J.H., Lucchinetti, C., Rodriguez, M. and Weinshenker, B.G. (2000) Multiple Sclerosis. *N Engl J Med*, **343**, 938-52.
- .3 Compston, A. and Coles, A. (2002) Multiple Sclerosis. *Lancet*, **359**, 1221-31.
- .4 Compston, A. and Coles, A. (2008) Multiple Sclerosis. *Lancet*, **372**, 1502-17.
- .5 Minagar, A., Jy, W., Jimenez, J.J. and Alexander, J.S. (2006) Multiple Sclerosis as a Vascular Disease. *Neurol Res*, **28**, 230-5.
- .6 Salaman, M.R. (2003) A Two-Step Hypothesis for the Appearance of Autoimmune Disease. *Autoimmunity*, **36**, 57-61.
- .7 Centola, M., Frank, M.B., Bolstad, A.I., Alex, P., Szanto, A., Zeher, M., Hjelmervik, T.O., Jonsson, R., Nakken, B., Szegedi, G. *et al.* (2006) Genome-Scale Assessment of Molecular Pathology in Systemic Autoimmune Diseases Using Microarray Technology: A Potential Breakthrough Diagnostic and Individualized Therapy-Design Tool. *Scand J Immunol*, **64**, 236-42.
- .8 Achiron, A., Gurevich, M., Friedman, N., Kaminski, N. and Mandel, M. (2004) Blood Transcriptional Signatures of Multiple Sclerosis: Unique Gene Expression of Disease Activity. *Ann Neurol*, **55**, 410-7.
- .9 Achiron, A., Gurevich, M., Snir, Y., Segal, E. and Mandel, M. (2007) Zinc-Ion Binding and Cytokine Activity Regulation Pathways Predicts Outcome in Relapsing-Remitting Multiple Sclerosis. *Clin Exp Immunol*, **149**, 235-42.
- .10 Baranzini, S.E., Mousavi, P., Rio, J., Caillier, S.J., Stillman, A., Villoslada, P., Wyatt, M.M., Comabella, M., Greller, L.D., Somogyi, R. *et al.* (2005) Transcription-Based Prediction of Response to Ifnbeta Using Supervised Computational Methods. *PLoS Biol*, **3**, e2.
- .11 Gurevich, M., Tuller, T., Rubinstein, U., Or-Bach, R. and Achiron, A. (2009) Prediction of Acute Multiple Sclerosis Relapses by Transcription Levels of Peripheral Blood Cells. *BMC Med Genomics*, **2**, 46.
- .12 Bomprezzi, R., Ringner, M., Kim, S., Bittner, M.L., Khan, J., Chen, Y., Elkahoul, A., Yu, A., Bielekova, B., Meltzer, P.S. *et al.* (2003) Gene Expression Profile in Multiple Sclerosis Patients and Healthy Controls: Identifying Pathways Relevant to Disease. *Hum Mol Genet*, **12**, 2191-9.
- .13 Whitney, A.R., Diehn, M., Popper, S.J., Alizadeh, A.A., Boldrick, J.C., Relman, D.A. and Brown, P.O. (2003) Individuality and Variation in Gene Expression Patterns in Human Blood. *Proc Natl Acad Sci U S A*, **100**, 1896-901.
- .14 Barnes, M.G., Aronow, B.J., Luyrink, L.K., Moroldo, M.B., Pavlidis, P., Passo, M.H., Grom, A.A., Hirsch, R., Giannini, E.H., Colbert, R.A. *et al.* (2004) Gene Expression in Juvenile Arthritis and Spondyloarthritis: Pro-Angiogenic Eln+ Chemokine Genes Relate to Course of Arthritis. *Rheumatology (Oxford)*, **43**, 973-9.
- .15 Mandel, M., Gurevich, M., Pauzner, R., Kaminski, N. and Achiron, A. (2004) Autoimmunity Gene Expression Portrait: Specific Signature That Intersects or Differentiates between Multiple Sclerosis and Systemic Lupus Erythematosus. *Clin Exp Immunol*, **138**, 164-70.
- .16 Kaizer, E.C., Glaser, C.L., Chaussabel, D., Banchereau, J., Pascual, V. and White, P.C. (2007) Gene Expression in Peripheral Blood Mononuclear Cells from Children with Diabetes. *J Clin Endocrinol Metab*, **92**, 3705-11.

- .17 Chaussabel, D., Pascual, V. and Banchereau, J. (2010) Assessing the Human Immune System through Blood Transcriptomics. *BMC Biol*, **8**, 84.
- .18 Compston, A., Ebers, G., Lassmann, H., McDonald, I., Matthews B., Wekerle, H. (1998.) *Mcalpine's Multiple Sclerosis*
- .19 Cosimo Prantera, B.I.K. (1996) *Crohn's Disease*. Informa Health Care
- .20 Earl J. Brewer, E.H.G., Donald A. Person (1982) *Juvenile Rheumatoid Arthritis*. Saunders.
- .21 Eisenbarth, G.S. (2004) *Type 1 Diabetes: Molecular, Cellular and Clinical Immunology*. Springer.
- .22 Lahita, R.G. (2003) *Lupus: Systemic Erythematosus*. Academic Press.
- .23 Noel R. Rose, I.R.M. (2006) *The Autoimmune Diseases*. Academic Press.
- .24 Ó'Moráin, C.A. (1991) *Ulcerative Colitis*. CRC Press.
- .25 Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) *Molecular Biology of the Cell*, New York.
- .26 Vogel, C., Abreu Rde, S., Ko, D., Le, S.Y., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M. and Penalva, L.O. (2010) Sequence Signatures and Mrna Concentration Can Explain Two-Thirds of Protein Abundance Variation in a Human Cell Line. *Mol Syst Biol*, **6**, 400.
- .27 Tuller, T., Kupiec, M. and Ruppín, E. (2007) Determinants of Protein Abundance and Translation Efficiency in *S. Cerevisiae*. *PLoS Comput Biol*, **3**, e248.
- .28 Scherer, A. (2009) *Batch Effects and Noise in Microarray Experiments: Sources and Solutions*. John Wiley & Sons.
- .29 Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D. and Ideker, T. (2007) Network-Based Classification of Breast Cancer Metastasis. *Mol Syst Biol*, **3**, 140.
- .30 Rutherford, A. (2001) *Introducing to Anova and Ancova a Glm Approach*. SAGE Publication Inc.
- .31 Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy. Statistical Society*, **57**, 289–300.
- .32 Oksenberg, J.R., Baranzini, S.E., Sawcer, S. and Hauser, S.L. (2008) The Genetics of Multiple Sclerosis: Snps to Pathways to Pathogenesis. *Nat Rev Genet*, **9**, 516-26.
- .33 Chowdhury, D. and Sen, R. (2003) Transient Il-7/Il-7r Signaling Provides a Mechanism for Feedback Inhibition of Immunoglobulin Heavy Chain Gene Rearrangements. *Immunity*, **18**, 229-41.
- .34 Tani-ichi, S., Lee, H.C., Ye, S.K. and Ikuta, K. Accessibility Control of Tcr Vgamma Region by Stat5. *Int Immunol*, **22**, 693-703.
- .35 Kim, K., Lee, C.K., Sayers, T.J., Muegge, K. and Durum, S.K (1998) .The Trophic Action of Il-7 on Pro-T Cells: Inhibition of Apoptosis of Pro-T1, -T2, and -T3 Cells Correlates with Bcl-2 and Bax Levels and Is Independent of Fas and P53 Pathways. *J Immunol*, **160**, 5735-41.
- .36 Girvan, M. and Newman, M.E. (2002) Community Structure in Social and Biological Networks. *Proc Natl Acad Sci U S A*, **99**, 7821-6.
- .37 Sarial, S., Shokrgozar, M.A., Amirzargar, A., Shokri, F., Radfar, J., Zohrevand, P., Arjang, Z., Sahraian, M.A. and Lotfi, J. (2008) Il-1, Il-1r and Tnfalpha Gene Polymorphisms in Iranian Patients with Multiple Sclerosis. *Iran J Allergy Asthma Immunol*, **7**, 37-40.
- .38 Schrijver, H.M., van As, J., Crusius, J.B., Dijkstra, C.D. and Uitdehaag, B.M. (2003) Interleukin (Il)-1 Gene Polymorphisms: Relevance of Disease Severity

- Associated Alleles with Il-1beta and Il-1ra Production in Multiple Sclerosis. *Mediators Inflamm*, **12**, 89-94.
- .39 Niino, M., Kikuchi, S., Fukazawa, T., Yabe, I., Sasaki, H. and Tashiro, K. (2001) Genetic Polymorphisms of Il-1beta and Il-1 Receptor Antagonist in Association with Multiple Sclerosis in Japanese Patients. *J Neuroimmunol*, **118**, 295-9.
- .40 Achiron, A., Feldman, A., Mandel, M. and Gurevich, M. (2007) Impaired Expression of Peripheral Blood Apoptotic-Related Gene Transcripts in Acute Multiple Sclerosis Relapse. *Ann N Y Acad Sci*, **1107**, 155-67.
- .41 Serhan, C.N., Ward, P.A. and Gilroy, D.A. (2010) *Fundamentals of Inflammation*. Cambridge University Press, New York.
- .42 Schwiebert, L.M. (2005) *Chemokines, Chemokine Receptors, and Disease*. Elsevier Academic Press, California.
- .43 Vyshkina, T., Sylvester, A., Sadiq, S., Bonilla, E., Perl, A. and Kalman, B. (2008) Ccl Genes in Multiple Sclerosis and Systemic Lupus Erythematosus. *J Neuroimmunol*, **200**, 145-52.
- .44 Sun, D., Tani, M., Newman, T.A., Krivacic, K., Phillips, M., Chernosky, A., Gill, P., Wei, T., Griswold, K.J., Ransohoff, R.M. *et al.* (2000) Role of Chemokines, Neuronal Projections, and the Blood-Brain Barrier in the Enhancement of Cerebral Edema Following Focal Brain Damage. *J Neuropathol Exp Neurol*, **59**, 1031-43.
- .45 Okamoto, H., Katsumata, Y., Nishimura, K. and Kamatani, N. (2004) Interferon-Inducible Protein 10/Cxcl10 Is Increased in the Cerebrospinal Fluid of Patients with Central Nervous System Lupus. *Arthritis Rheum*, **50**, 3731-2.
- .46 Bonner, J.C., Badgett, A., Lindroos, P.M. and Osornio-Vargas, A.R. (1995) Transforming Growth Factor Beta 1 Downregulates the Platelet-Derived Growth Factor Alpha-Receptor Subtype on Human Lung Fibroblasts in Vitro. *Am J Respir Cell Mol Biol*, **13**, 496-505.
- .47 Kurtzke, J.F. (1983) Rating Neurologic Impairment in Multiple Sclerosis: An Expanded Disability Status Scale (Edss). *Neurology*, **33**, 1444-52.
- .48 McDonald, W.I., Compston, A., Edan, G., Goodkin, D., Hartung, H.P., Lublin, F.D., McFarland, H.F., Paty, D.W., Polman, C.H., Reingold, S.C. *et al.* (2001) Recommended Diagnostic Criteria for Multiple Sclerosis: Guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. *Ann Neurol*, **50**, 121-7.
- .49 Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N. *et al.* (2005) Towards a Proteome-Scale Map of the Human Protein-Protein Interaction Network. *Nature*, **437**, 1173-8.
- .50 Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S. *et al.* (2005) A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell*, **122**, 957-68.
- .51 Peri, S., Navarro, J.D., Amanchy, R., Kristiansen, T.Z., Jonnalagadda, C.K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T.K., Gronborg, M. *et al.* (2003) Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans. *Genome Res*, **13**, 2363-71.
- .52 Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M. and Eisenberg, D. (2002) Dip, the Database of Interacting Proteins: A Research Tool for Studying Cellular Networks of Protein Interactions. *Nucleic Acids Res*, **30**, 303-5.

- .53 Zhao, F., Xuan, Z., Liu, L. and Zhang, M.Q. (2005) Tred: A Transcriptional Regulatory Element Database and a Platform for in Silico Gene Regulation Studies. *Nucleic Acids Res*, **33**, D103-7.
- .54 Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003) David: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*, **4**, P3.
- .55 Pedhazur, E.J. (1997) *Multiple Regression in Behavioral Research*. Wadsworth Publishing.

Legends to Figures

Figure 1. Flow diagram of the analysis performed in this study.

Figure 2. Global view on genes that are differentially expressed in MS and the protein-protein interactions between them. The graph includes edges of two types: continuous edges correspond to protein-protein interaction; dashed edges corresponding to paths of length two (via a gene that is not differentially expressed). The graph also includes three types of nodes: differentially expressed in *MS/healthy* and *Rel/Rem* (red), differentially expressed in *MS/healthy* (blue), differentially expressed in *Rel/Rem* (yellow).

Differentially expressed genes in MS tend to be close to each other in the protein-protein interaction network (p-value 0.05).

Figure 3. Differentially expressed interactions in the *MS/healthy* dataset. The graph includes only genes that were differentially expressed and also have significant *PPI p-values*. The graph was clustered (Methods (36); each cluster has a different color), and for each cluster we performed a functional enrichment analysis (Methods). B. The functional enrichments for the entire graph.

Figure 4. Differentially expressed interactions in the *Rel/Rem* dataset. A. The graph includes only genes that were differentially expressed with significant *PPI p-values*. The

graph was clustered (Methods (36); each cluster has a different color), and for each cluster we performed a functional enrichment (Methods). B. The functional enrichments for the entire graph.

Figure 5. Pathway analysis by Ingenuity based on the significant (*PPI* and *ANOVA*) genes in *MS/healthy* (A.-B.) and *Rel/Rem* (C.). Two signaling pathways found to be enriched by Ingenuity in the *MS/healthy* database: the Inflammation pathway (A.) and the Apoptosis pathway (B.); one pathway was significant in the *Rel/Rem* database: the Lymphocyte and monocyte adhesion pathway (C.). Red/green nodes denote *ANOVA* significant over/under expression respectively; black nodes denote genes with significant *PPI* p-values (but with non-significant *ANOVA* p-value).

Tables

A.

Group	Age	Disease Duration	Annual Relapse Rate	EDSS	F/M
All (N = 164)	35.5 +- 10.31	N/A	N/A	N/A	99/65
MS patients (N = 123)	35.6 +- 10.8	6.16 +- 6.15	1.58 +- 1.66	2.25 +- 1.45	78/45
Healthy subjects (N = 41)	35.1 +- 8.7	N/A	N/A	N/A	21/20

B.

Group	Age	Disease Duration	Annual Relapse Rate	EDSS	F/M
All (N = 109)	36.6 +- 10.47	6.6 +- 6.31	1.46 +- 1.59	2.54 +- 1.28	72/37
Patients in Relapse	34.0 +- 8.4	6.14 +- 4.13	1.0 +- 0.65	3.29 +- 1.2	25/9

(N = 34)					
Patients in Remission (N = 75)	37.7 +- 11.14	6.8 +- 7.1	1.7 +- 1.8	2.21 +- 1.18	47/28

Table 1. Clinical and demographical characteristics of the patients/subjects in the analyzed datasets. A. The entire dataset (*MS/Healthy* - MS patients vs. Healthy subjects). B. The subset of patients in relapse and in remission (*Rel/Rem* dataset). See also supplementary table 1.

A.

ANOVA p-values				Protein-interaction p-values			
Gene	Fold change	P-value	Description	Gene	Fold change	P-value	Description
GNAI1	-1	9.80E-06	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1	JUN	1	1.52E-14	jun oncogene
PDE3A	1	5.93E-05	phosphodiesterase 3A, cGMP-inhibited	SMAD3	-1	1.43E-11	SMAD family member 3
PANK4	-1	8.99E-05	pantothenate kinase 4	RB1	1	1.26E-10	retinoblastoma 1
STK38	-1	9.99E-05	serine/threonine kinase 38	JUNB	1	6.48E-10	jun B proto-oncogene
PARP4	-1	0.0001179	poly (ADP-ribose) polymerase family, member 4	FOS	1	1.13E-09	FBJ murine osteosarcoma viral oncogene homolog
PDCD10	-1	0.0001555	programmed cell death 10	JUND	1	1.12E-08	jun D proto-oncogene
MMP9	1	0.0001914	matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	TP53	-1	7.71E-08	tumor protein p53
OR2W1	-1	0.0002247	olfactory receptor,	TRAF1	-1	1.75E-	TNF receptor-

			family 2, subfamily W, member 1			07	associated factor 1
PLEKHF1	-1	0.0002551	pleckstrin homology domain containing, family F (with FYVE domain) member 1	BRCA1	-1	1.15E- 06	breast cancer 1, early onset
SOCS1	1	0.0003106	suppressor of cytokine signaling 1	TBP	-1	1.21E- 06	TATA box binding protein

B.

ANOVA p-values				Protein-interaction p-values			
Gene	Fold change	P-value	Descriptio	Gene	Fold change	P-value	Descriptio
GNG4	1	7.12E-06	guanine nucleotide binding protein (G protein), gamma 4	PTPRD	1	2.21E- 10	protein tyrosine phosphatase, receptor type, D
C6orf103	1	2.53E-05	chromosome 6 open reading frame 103	PDGFRB	0	6.55E- 09	platelet- derived growth factor receptor, beta polypeptide
GPR116	-1	2.66E-05	G protein- coupled receptor 116	STAT5A	0	6.55E- 09	signal transducer and activator of transcription 5A

KIAA1654	1	2.68E-05	KIAA1654 protein	PTPRF	1	1.67E-08	protein tyrosine phosphatase, receptor type, F
CCL14- CCL15	1	4.00E-05	chemokine (C-C motif) ligand 14 // chemokine (C-C motif) ligand 15	PPFIA1	0	1.67E-08	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 1
ST7L	1	4.26E-05	suppression of tumorigenicity 7 like	PTPRS	1	4.40E-08	protein tyrosine phosphatase, receptor type, S
ACCN3	1	0.000108689	amiloride- sensitive cation channel 3	STAT5B	-1	4.68E-08	signal transducer and activator of transcription 5B
ALOX15	1	0.000117418	arachidonate 15- lipoxygenase	PTPRA	-1	1.65E-07	protein tyrosine phosphatase, receptor type, A
CARD14	1	0.000120208	caspase recruitment domain family, member 14	PPFIA2	1	1.65E-07	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting

							protein (liprin), alpha 2
LGR4	1	0.000132248	leucine-rich repeat- containing G protein- coupled receptor 4	SCNN1B	0	1.65E- 07	sodium channel, nonvoltage- gated 1, beta

Table 2. Genes with the most significant p-values (ANOVA and PPI) for the MS/Healthy dataset (A.) and for the Rel/Rem dataset (B.). All the PPI p-values pass the FDR.

A.

Term	Gene expression p-value	PP-net p-value
Positive regulation of cell death	4.16*E-06	1.03*E-17
Chemotaxis	9.9*E-06	---
regulation of transcription from RNA polymerase II promoter	0.00076	3.41*E-55
Inflammatory response	0.0014	1.73E-06
positive regulation of NF-kappaB transcription factor activity	----	0.00094
immune system development	---	7.22*E-08
B cell differentiation	---	0.00217
Regulation of caspase activity	---	0.000367

B.

Cellular process	Gene expression p-value	PP-net p-value

Cell adhesion	0.00012	4.99*E-07
Cell migration	0.000034	8.15*E-10
Cell-cell signaling	2.2*E-06	1.77*E-07
Regulation of cell motion	0.0007	1.48*E-10
Positive regulation of cell proliferation	---	5.95*E-08
Synaptic transmission	---	4.61*E-06
T cell receptor signaling pathway	---	0.0000395
T cell activation	---	0.000349
Regulation of T cell proliferation	---	0.0007
T cell differentiation	---	0.0009
positive regulation of natural killer cell mediated cytotoxicity	---	0.00245
Regulation of apoptosis	---	1.26*E-06

Table 3. Some of the cellular processes that were significantly enriched in A.

***MS/Healthy* dataset and B. *Rel/Rem* dataset. Full tables appear in the**

supplementary material based on the *ANOVA* p-values and based on the pp-net p-values (*Methods*).

Symbol	Entrez	P-value PPnet	Rank PPnet	P-value ANOVA	Rank ANOVA
HLA- DRB1	3123 100133661 3127	Rel/Rem: pval=1 MS/cont: pval=1	Rel/Rem: 1884 MS/cont: 2135	Rel/Rem: pval=0.3134 MS/cont: pval=0.0491	Rel/Rem: 4556 MS/cont: 1244
IL2RA	3559	Rel/Rem: 0.0382 MS/cont: 1	Rel/Rem: 208/13523 MS/cont: 2277/13523	Rel/Rem: Pval=0.8932 MS/cont: pval=0.9509	Rel/Rem: 12130 MS/cont: 12951
CD58	965	Rel/Rem: 1 MS/cont: 1	Rel/Rem: 943/13523 MS/cont: 1244/13523	Rel/Rem: pval=0.9266 MS/cont: pval=9.3791e- 004	Rel/Rem: 12588 MS/cont: 42
IL7R	3575	Rel/Rem: 0.0000335 MS/cont:	Rel/Rem: 24 MS/cont:	Rel/Rem: pval=0.6162 MS/cont:	Rel/Rem: 8498 MS/cont:

		1	2290	0.8654	11984
PLP1	5354	Rel/Rem: 1	Rel/Rem: 2844	Rel/Rem: pval=0.062	Rel/Rem: 1126
		MS/cont: 1	MS/cont: 3050	MS/cont: pval=0.3548	MS/cont: 5735
MAG	4099	Rel/Rem: 0.0646257	Rel/Rem: 280	Rel/Rem: pval=0.927	Rel/Rem: 12597
		MS/cont: 1	MS/cont: 2564	MS/cont: pval=0.9969	MS/cont: 13488
MOG	4340	Rel/Rem: 1	Rel/Rem:	Rel/Rem: pval=0.06997	Rel/Rem: 1240
		MS/cont: 1	MS/cont: 2666	MS/cont: pval=0.2466	MS/cont: 4312

Table 4. ANOVA p-values, PPI p-values, and the p-value ranks for genes with SNPs related to MS (by (32)).

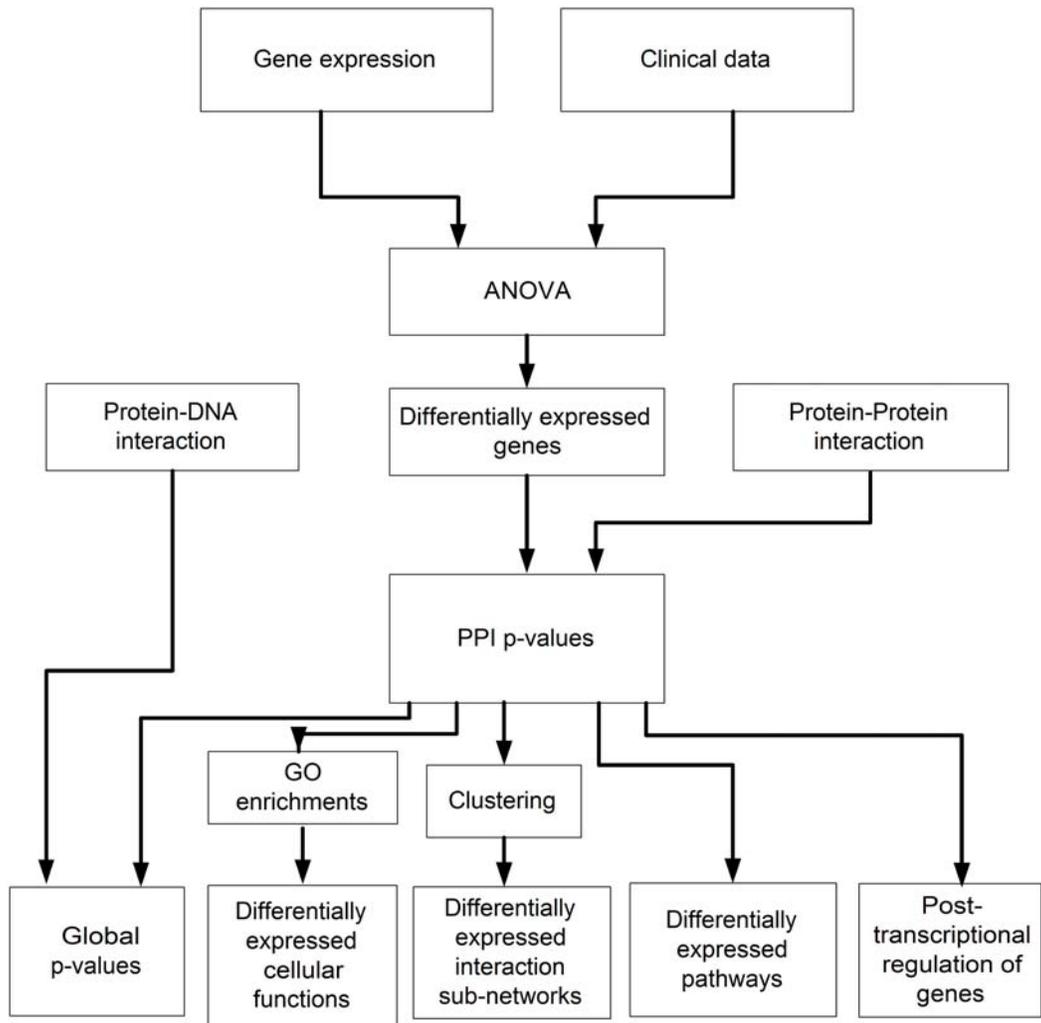


Fig. 1

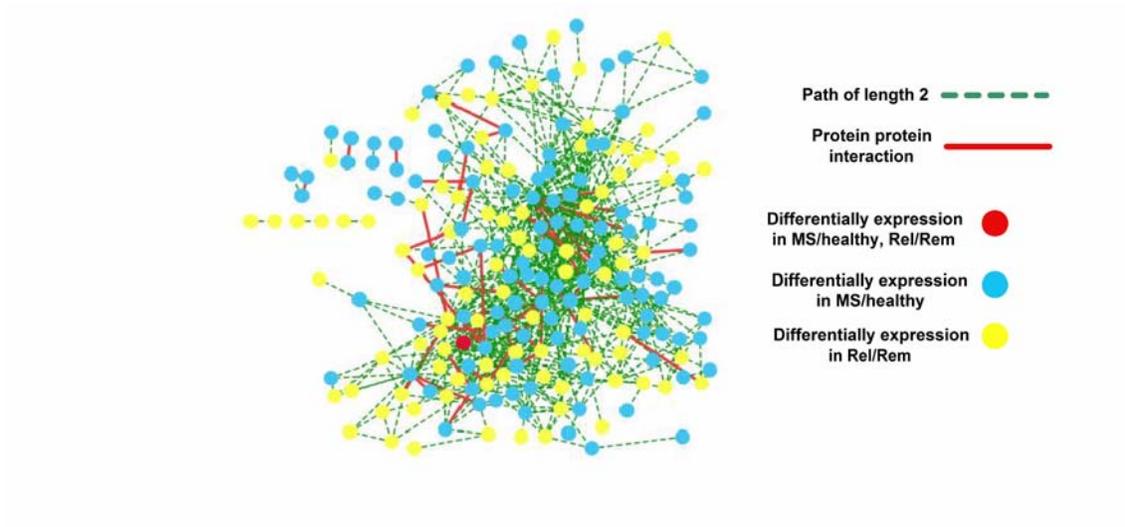


Fig. 2

A.

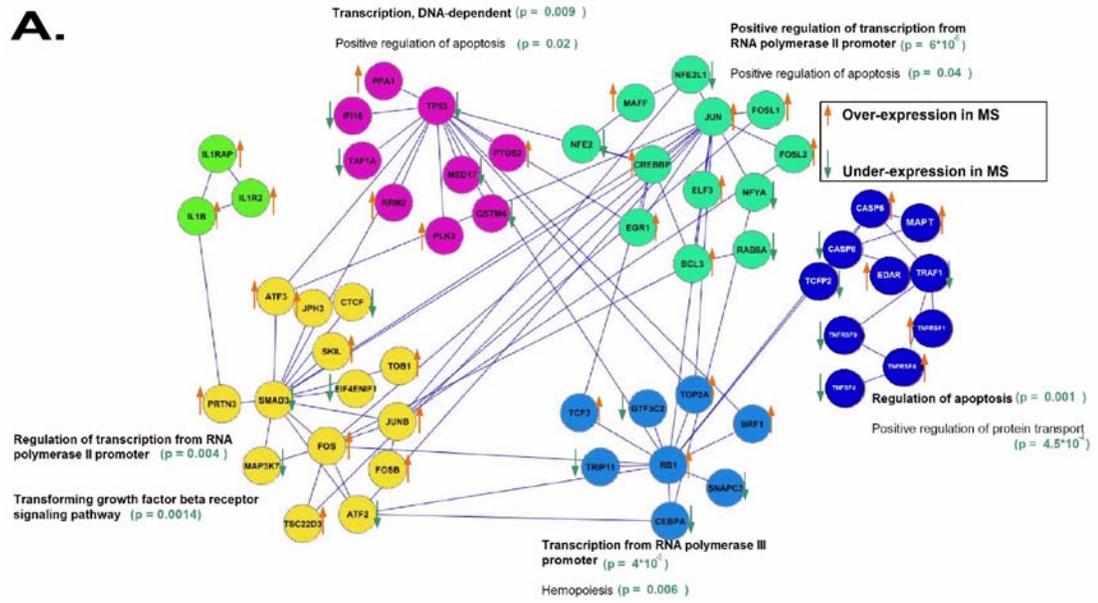


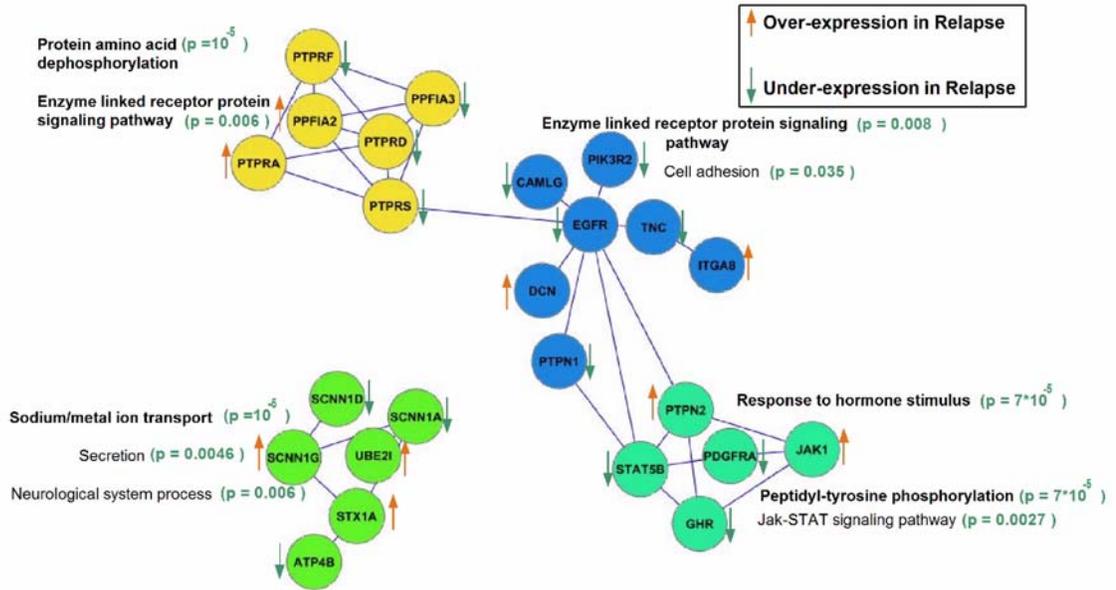
Fig. 3A

B.

Term	P-value
Transcription, DNA -dependent	6.74E -10
Regulation of cell proliferation	4.11E -07
Positive regulation of apoptosis	4.09E -06
Response to abiotic stimulus	2.91E -04
Release of cytochrome c from mitochondria	0.004234
Leukocytedifferentiation	3.18E -05
T cell activation	2.95E -04
Macromolecular complex subunit organization	3.89E -04
Inflammatory response	2.30E -06
Cytokine-cytokine receptor interaction	4.85E -06
Transmembrane receptor protein serine/threonine kinase signaling pathway	0.001338
Response to cytokine stimulus	1.71E -06
Positive regulation of developmental process	0.000297
Regulation of protein transport	1.85E -04
Regulation of cytokine production	0.001523
Protein import into nucleus, translocation	5.20E -04

Fig. 3B

A.

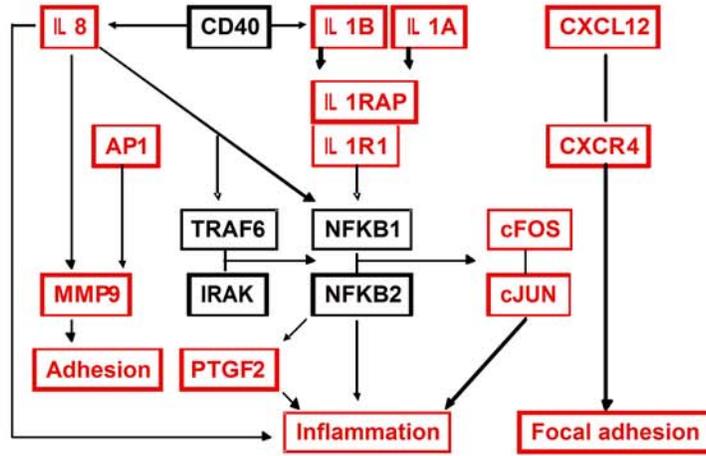


B.

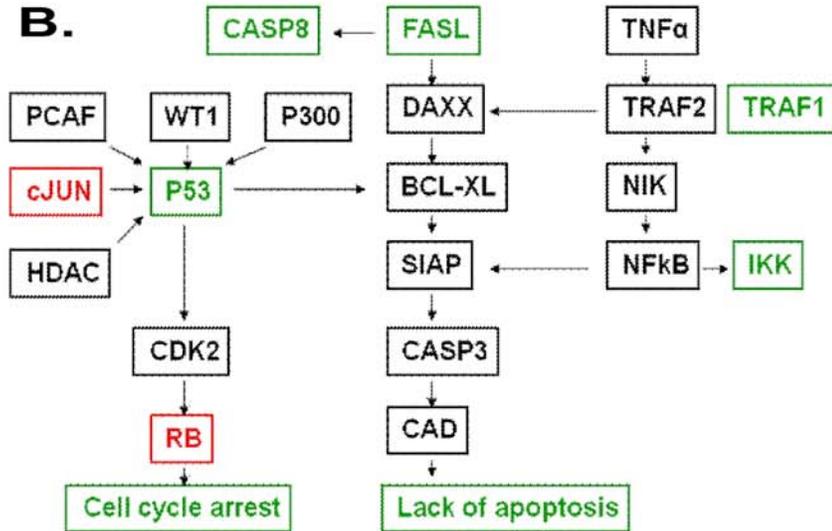
Term	P-value
Enzyme linked receptor protein signaling pathway	8.44E-11
Protein amino acid dephosphorylation	3.29E-05
Cell adhesion	0.002986
Response to endogenous stimulus	8.50E-04
Regulation of neurological system process	6.44E-05
Regulation of synaptic transmission	5.53E-04

Fig. 4

A.



B.



C.

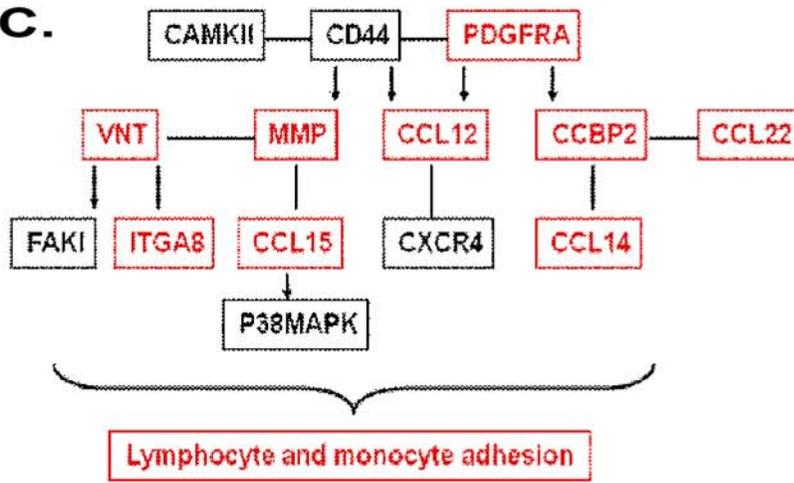


Fig. 5