UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biologiche e Molecolari

XXV Ciclo

TRANSLATIONAL BIOINFORMATICS AND SYSTEMS BIOLOGY APPROACHES TO GENETIC AND TRANSCRIPTIONAL DATA IN COMPLEX HUMAN DISORDERS

Ramesh Menon

PhD Thesis

Tutor: Dr. Cinthia Farina

Co-tutor: Dr. David Horner

Coordinator: Prof. Gianni Dehò

Anno Accademico 2011-2012

SSD: BIO/13

Thesis performed at Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Milan.

Table of Contents

PART I

PART II

PART III

PART I

1. ABSTRACT

Human complex diseases are caused by genetic and environmental factors. Genome-wide association studies (gwas) are aimed to identify common variants predisposing to those disorders. However, till date, the data generated from such studies have not been extensively explored to identify the molecular and functional framework hosting the susceptibility genes. We reconstructed the multiple sclerosis-MS genetic interactome and searched for their interactions with genes predisposing to either neurodegenerative or autoimmune diseases such as Parkinson's disease-PD, Alzheimer's disease-AD, multiple sclerosis-MS, rheumatoid arthritis-RA and Type 1 diabetes-T1D. It was observed that several genes predisposing to the other autoimmune or neurodegenerative disorders may come into contact with MS interactome, suggesting that susceptibility to distinct diseases may converge towards common molecular and biological networks. In order to test this hypothesis, we performed pathway enrichment analyses on each disease interactome independently. Several issues related to immune function and growth factor signaling pathways appeared in all autoimmune diseases. Further, the paired analyses of disease interactomes revealed significant molecular and functional relatedness among the diseases. Therefore, the shift from single genes to molecular frameworks via systems biology approach highlighted several known pathogenic processes, indicating that changes in these functions might be driven or sustained by the framework linked to genetic susceptibility. Notably, MS is a complex disease of the central

nervous system (CNS), but many of the susceptibility genes play a role in immune system. Interestingly, the most widely used therapeutic drugs in MS are either immunosuppressive or immunomodulatory agents, indicating that targeting peripheral immune system is beneficial to patients with this CNS disorder. Next, we measured the global gene expression in peripheral blood mono nuclear cells (PBMCs) from MS and healthy subjects to discover disease genes, molecular biomarkers and drug targets. Extending the bioinformatics analysis of the transcriptome data to network-biology level enabled us to identify few crucial transcriptional regulators in MS. Further, as a first step towards translational research, studies were conducted in the animal model of MS, based on the outcomes of the bioinformatics analysis. Significant amelioration of disease activity was observed in diseased animals treated with drug targeting SP1 transcription factor, compared to the untreated group. Hence, disease transcriptomics combined with network-biology analysis provided a powerful platform for the identification of functional networks and molecular targets in MS.

2.1 HUMAN COMPLEX DISORDERS

Complex diseases arise from (unfavorable) multifaceted interaction of genetic and environmental factors. A genetic predisposition means that an individual has a genetic susceptibility to developing a certain disease. However, different from monogenic disorders, where mutations in a single gene/genetic component is sufficient to produce the clinical phenotype, complex diseases do not obey the single-gene dominant or recessive Mendelian pattern of inheritance, rather they arise from the action of environmental and life-style factors (cigarette smoking, sunlight, obesity, viruses etc) in individuals carrying numerous susceptibility variants, each conferring only limited risk [1]. The diseases falling into this category are numerous; here we focus our attention on Alzheimer's disease, diabetes, multiple sclerosis, Parkinson's disease, rheumatoid arthritis.

2.1.1 Parkinson's disease

Parkinson's disease (Park) is the second most common neurodegenerative disorder with a prevalence of 0.3% prevalence in industrialized countries, which rises to more than 1% over 60 years of age and above 4% at the threshold of 80 years of age [1, 2]. The mean age of onset is approximately 60 years. However, 10% of cases are classified as young onset, occurring between 20 and 50 years of age, which may represent a distinct disease group. Park is more prevalent in men than in women by upto 3:1 male to female

ratio [3]. Approximately 20% of patients with Park report a family history of the disease and monogenic forms of Park are relatively rare. Like any other complex disease, Park is multifactorial disorder, likely to be arising from a combination of polygenic inheritance, environmental exposures, and gene-environment interactions.

2.1.1.1 Symptoms and Diagnosis

Impaired motor function is the hallmark symptom of Park, which may be featured with bradykinesia, rigidity, tremor, and postural instability with an asymmetric onset spreading to become bilateral with time. Other motor features include gait and posture changes that manifest as festination (rapid shuffling steps with a forward-flexed posture when walking), speech and swallowing difficulties, and a masklike facial expression and micrographia [4]. Further, a good response to these symptoms with dopaminergic treatment is confirmatory of the diagnosis. However, in recent years non motor symptoms (NMS) such as depression, sleep disturbance, sensory abnormalities, autonomic dysfunction, and cognitive decline are also recognized as symptoms of Park. Whereas the causes of motor dysfunction in Park are reasonably well understood, the cause of NMS in Park remains poorly researched and they may largely relate to pathology outside of the basal ganglia [5].

2.1.1.2 Risk factors for developing Parkinson's disease

Even though age represents the biggest predisposing factor for Park, it remains unknown whether it is chronological age or the aging process that is [6]. In women, loss of estrogen production with age may remove a protective effect [7]. More importantly, genetic causes may account for up to 40% of risk in the population [8]. On the other hand, environmental causes include exposure to certain solvents (eg: n-hexane, methanol), carbon monoxide poisoning and hydrogen sulfide intoxication [2]. In addition, it has been shown that pesticides such as Paraquat ($C_{12}H_{14}Cl_2N_2$) and Rotenone ($C_{23}H_{22}O_6$) are able to destroy dopaminergic cells in rodents, and they may contribute to the increased risk of developing Parkinson's disease [2].

2.1.1.3 Genetics and pathogenesis of Parkinson's disease

Investigation of familial Park patients has revealed several autosomal dominant and autosomal recessive gene mutations responsible for variants of the disease [9]. These include α-synuclein mutations and triplication, parkin, ubiquitin carboxyl-terminal hydrolase L1 (UCH-L1), DJ-1, phosphatase and tensin homolog-inducible kinase 1 (PINK1), leucine-rich repeat kinase 2 (LRRK2), and glucocerebrosidase (GBA). Among these, Parkin and LRRK2 are probably the most common genetic link to young-onset and late-onset Park respectively. The α-synuclein is identified as a major component of Lewy bodies and Lewy neurites, and the labeling of Park as a synucleinopathy, underpinning much of the consensus on the final stages of neuronal loss in Parkinson's

disease being related to altered protein aggregation [10, 11]. LRRK2 interacts with Parkin, and mutant LRRK2 induces apoptotic cell death in cultured neurons. Autosomal loss of function mutations include those in the ubiquitin E3 ligase Parkin, which in combination with the ubiquitin-conjugating enzyme causes the attachment of ubiquitin as a marker on proteins destined for destruction by the proteasome. Additionally mutations in the mitochondrial PINK1 protect cells from mitochondrial stress/dysfunction, and the rarest mutation in the redox-sensitive chaperone DJ-1, protects cells against oxidative stress [12, 13]. *The International PD Genomics Consortium*, has recently revealed 14 "risk gene" loci for PARK, including α-synuclein, LRRK2, human leukocyte antigen (HLA), and tau [14, 15].

Familial forms of Parkinson's disease and the associated gene mutations currently account for approximately 10% of cases and they have distinct clinical and pathological phenotypes. Notably, mechanisms such as mitochondrial dysfunction, oxidative stress, and altered protein handling are common to both familial and sporadic Park (Fig. 1) [9]. More recently, the role of mitochondria in familial Parkinson's disease is widely investigated and reported that the mutations in α-synuclein, parkin, PINK1, and DJ-1 and perhaps LRRK2 have been associated with altered mitochondrial function (Fig. 1). [16] These mutations can lead to altered protein localization in mitochondria in Park, abnormalities in mitochondrial structure and function, and a decrease in complex I assembly and activity. Loss of function of DJ-1, Parkin and PINK1 genes decreases

mitochondrial protection against oxidative stress, which in turn increases mitochondrial dysfunction. Another important role for Parkin and PINK1 is in the turnover of mitochondria by autophagy, specifically mitophagy; they act in tandem to regulate this process. This may be very important in Park, where autophagy seems impaired and thus reducing the ability of the cell to remove the damaged mitochondria.

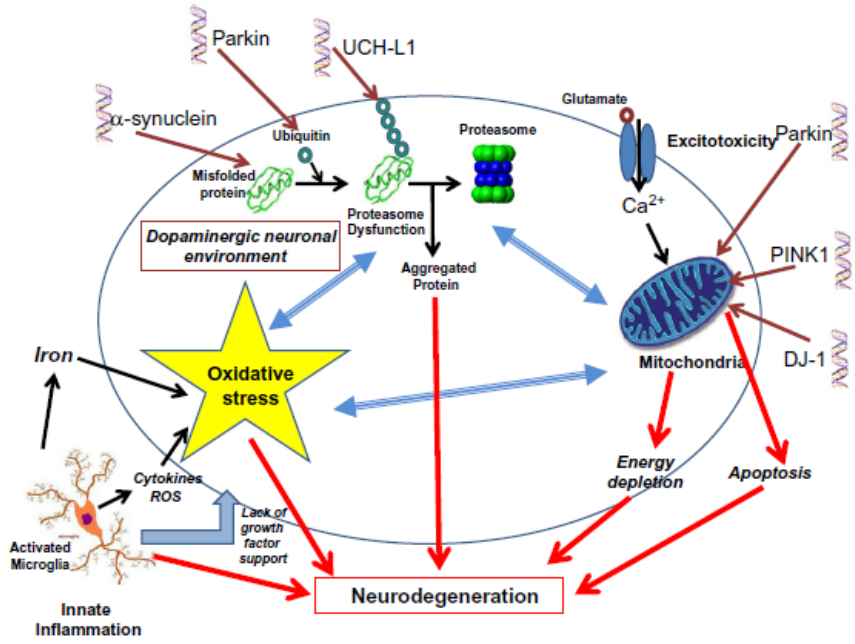


Figure 1. Molecular mechanisms in Parkinson's disease from a genetic point of view. Mutations in key genes contribute to the neurodegenerative process in dopaminergic neurons in the substantia nigra. Double-headed arrows (blue) indicate molecular mechanisms that may not only be toxic in their own right but may also influence other molecular mechanisms known to be features in Parkinson's disease. Double helix structures identify some of the

common gene mutations found in familial Parkinson's disease and brown arrows indicate where the altered protein may interfere with cell function and leads to cell death. (Derived from [1, 2]).Even though, our understanding of the nature and causes of Parkinson's disease is improving, there are several fundamental gaps in the present knowledge, which hinders the ability to develop effective neuroprotective strategies for the disease. The mechanisms currently used to explain pathogenesis in Parkinson's disease may just be the downstream consequence of a so far unknown trigger. Importantly, we still do not understand the molecular mechanisms that account for the spreading pathology of the disease. Further advances in the understanding of the pathology will enable the discovery of therapeutics for Park patients.

2.1.2 Alzheimer's disease

Alzheimer's disease (Alz) is the most common neurodegenerative disease and a common cause of dementia in the elderly. Dementia is recognized with use of the criteria of the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV), American Psychiatric Association In the UK, over 700,000 people have dementia [17] .In the United States, Alzheimer's disease is the seventh leading cause of all deaths, which has increased to 46.1 % between 2000 and 2006.

2.1.2.1 Symptoms and Diagnosis

The classic clinical features of Alzheimer's disease are memory impairment, deterioration of language, and visuospatial disability[18]. Motor and sensory abnormalities, gait disturbances, and seizures are uncommon until the late phases of the disease [19]. The diagnosis of Alzheimer's disease was developed by the National Institute of Neurologic and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS–ADRDA)[19]. This was later updated with more reliable diagnosis tools such as through structural MRI, molecular neuroimaging with PET, and cerebrospinal fluid analyses [20].

2.1.2.2 Risk factors for developing Alzheimer's disease

There is a clear genetic influence in the manifestation of Alz. Being a first-order relative of Alz patient increases the risk by two times in one's life. Alz can be thought of as 2 separate entities: a

rare early-onset form (early-onset familial Alzheimer disease) before 65 years of age and a common late-onset form (late-onset Alzheimer disease), which manifests at divergent ages. Each involves a different set of genes. Three autosomal dominant causal genes have been reliably associated with early-onset familial Alzheimer disease: presenilin 1(PSEN1), presenilin 2 (PSEN2), and amyloid precursor protein (APP). Mutations of these 3 genes make up approximately 70% cases of early-onset familial Alzheimer disease, in which a mutation in PSEN1 is most common and PSEN2 is the rarest [21]. Further, mutations apolipoprotein E gene provides a greater risk for developing early-onset or late-onset of the disease, based on the location of the mutations. . Finally, trisomy 21 may also led to Alz [22].

Diet is a major non-genetic risk factor for developing Alz. Dietary fats, and total energy intake (in terms of calories) were found to be significant risk factors for the development of Alz. Interestingly, consumption of fish reduces Alz risk [23]. Aluminum may cause neurological damage and a number of studies have linked aluminum to an increased risk for developing Alzheimer's disease.

2.1.2.3 Genetics and pathogenesis in Alzheimer's disease

Although the mechanisms involved in the pathogenesis of Alz largely remain unknown, the accumulation of Aβ (a product of APP gene) is a common observation in the disease. In the amyloidogenic pathway, APP is initially cleaved by β-secretase,

followed by γ-secretase, resulting in products that include Aβ [24]. An imbalance between Aβ production and clearance is thought to leAlz to the accumulation of amyloid in senile plaques, which results in disease progression [25, 26] (Fig.2).

The brain pathology of Alz is characterized by neuronal loss, *tau*-positive neurofibrillary tangles and amyloid plaques, consisting mainly of Aβ40/42 peptides generated by cleavage of the β-amyloid precursor protein (Fig.2). The finding that mutations in the *tau* gene are responsible for frontotemporal dementia proved that the formation of neurofibrillary tangles has neurotoxic consequences. The combination of genetic and biochemical data led to the amyloid cascade hypothesis which suggested that Aβ deposition is the primary event in disease pathogenesis.

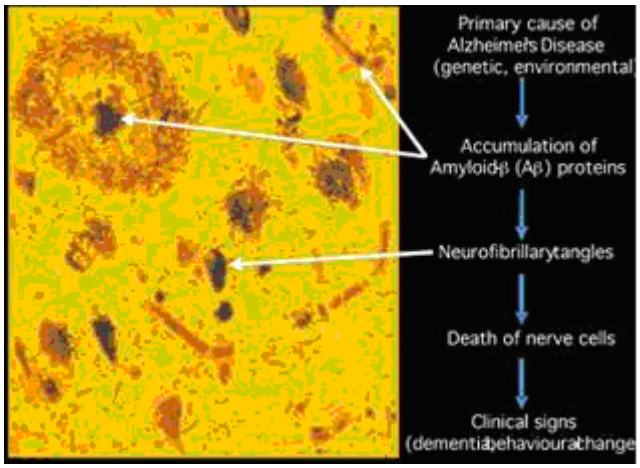


Figure 2. The path of neurodegeneration in Azheimer's disease (Derived from http://tanz.med.utoronto.ca/page/science-alzheimer)

2.1.2.4 Recently discovered GWAS genes

In the past five years, large collaborative efforts significantly contributed to the discovery of novel susceptibility genes in Alzheimer's disease, with the discovery of at least nine novel risk loci were uncovered [27]. Strikingly, the gene *CLU* was detected simultaneously in two independent studies [28, 29]. Additionally, genome-wide association and replication have been noted for single nucleotide polymorphisms in or near *CR1, PICALM,* and *BIN1*. Finally, association with single nucleotide polymorphisms in *MS4A* cluster, *CD2AP, CD33, EPHA1*, and *ABCA7* further increased the pace of genetic discovery in Alzheimer's disease [27].

2.1.3 Type 1 diabetes

Type 1 diabetes (T1D) represents one of more than 80 diseases considered to have an autoimmune aetiology, which affects millions of people worldwide. The disease occurs as a consequence of the organ-specific immune destruction of the insulin-producing β-cells in the islets of Langerhans within the pancreas [30]. Once they are destroyed, patients with type1 diabetes lose modulation of blood glucose, which can result in both acute conditions (for example, ketoacidosis and severe hypoglycaemia) and secondary complications (including heart disease, blindness and kidney failure) [31]. As a complex disorder, T1D develops as a consequence of a combination of genetic predisposition and environmental factors. Interestingly, unlike most other autoimmune diseases where risk is greatest in females, type 1 diabetes is the only major organ-specific autoimmune disorder that does not to show a strong female bias [32].

2.1.3.1 Symptoms and diagnosis

Type 1 diabetes symptoms include: increased thirst and frequent urination, extreme hunger, weight loss, fatigue, blurred vision [source: http://www.mayoclinic.com] . In diagnostics, diabetes-related autoantibodies are used to predict the appearance of T1D before any hyperglycemia arises [33]. Further, the presence of ketones (byproduct from the breakdown of fat) in urine is suggestive for type 1 rather than type 2 diabetes [source: http://www.mayoclinic.com].

2.1.3.2 Risk factors for developing Type 1 diabetes

Environmental factors have a major weight in the incidence of type 1 diabetes. Studies in identical twins showed that the co-incidence percentages range from 30-50% [33]. A study reported that there is 10-fold increased risk in occurrence of T1D among Caucasians living in different areas of Europe, and a tendency to acquire the incidence of the disease of the destination country for people who migrate [33]. Karjalainen and colleagues reported autoimmune response in T1D is influenced by antibodies against cow's milk proteins [34]. However, it has been shown that breastfeeding decreases the risk in later life [35].

2.1.3.3 Genetics and pathology of Type 1 diabetes

The hallmark of T1D is the progressive loss of β-cell function over a period of years. The precise immunologic, genetic and physiologic events that control disease initiation and progression is still to be uncovered [36]. However, studies in animal models of T1D have demonstrated that the pathology is consequence of a breakdown in immune regulation, resulting in the expansion of autoreactive $CD4^+$ and $CD8^+$ T cells, autoantibody-producing B lymphocytes, and thereby activation of the innate immune system that work together to destroy the insulin-producing β-cells [36]. Both genetic and gene-environmental interactions have an important role in T1D. As common among autoimmune disorders, the

highest risk alleles from HLA- DR2, DR3, DR4 etc. [37]. In the animal model it has been shown that the HLA class II molecules present peptide to CD4 cells and HLA class I molecules present peptides to cytotoxic CD8+ T cells, which infiltrates to pancreas [32]. The second susceptibility locus is in the variable number tandem repeat (VNTR) region consisting of a 14 to 15 bp consensus sequence upstream of the INS gene which regulates transcription rates of insulin and its precursors [38, 39]. Another gene linked to T1D is CTLA-4 [40], which is found on activated T cells that produces a negative signal by inhibiting the T cell receptor signaling complex ligand interactions (blocks binding of CD80 and CD86). It is thought that inherited changes in CTLA-4 gene expression can increase T cell self-reactivity and therefore play an important role in T1D [41]. More recently, PTPN22 susceptibility has been reported in various autoimmune disorders including T1D. The protein product lymphoid protein tyrosine phosphatase (LYP) is responsible for preventing spontaneous T cell activation and they have the ability to prevent the response to antigen by dephosphorylating and inactivating T cell receptors pancreas [32]. It has been demonstrated that PTPN22 susceptibility leads to the decrease in negative regulation of hyper-reactive T cells [42]. In addition to T cells, LYP is expressed in natural killer (NK) cells, B cells, macrophages and dendritic cells (DCs). So, alterations in PTPN22 can deregulate functions of several immune cells [32]. In 2005, the IL2RA region on chromosome 10p15 was found to be associated with T1D [43]. IL2RA encodes the α-chain of the IL-2 receptor complex (also called as CD25) which is responsible for

binding IL-2 and helps the proliferation of regulatory T cells. Two IL-2R SNPs associated with the increased risk of T1D have been reported (with rs52580101 the most closely associated [44]. A recent study measuring expression of IL-2R in individuals homozygous for susceptible and protective SNPs associated with T1D demonstrated that on stimulation, higher percentages of CD69+ CD4+ memory T cells secreted IL-2 from individuals with the protective SNP compared to individuals with the susceptible SNP [45]. There are several other susceptibility genes reported recently from the gwas studies, which include: CCR5, UBASH3A, IFIH1, TLR7, PDCA1 etc.

From the numerous susceptibility genes identified by gwas, it appears that a global problem of immune regulation may underlie disease susceptibility. For instance, mutations of genes encoded in several of the susceptibility loci including IL2A (CD25), CTLA-4, PTPN22, and PDCA1 (PD-1) in multiple animal models lead to the development of a diverse array of autoimmune diseases, including T1D. It has been shown that the number and function of the T_{reg} cells (especially the stable $CD25^+$ subset) in the pancreas is significantly reduced in the inflamed islet tissue, which may be due to lack of IL-2 production [46]. Further, it has been demonstrated that a significant increase in the number of IFNγ-producing $Foxp3^+$ cells in the new-onset T1D patients, occuring with slightly reduced Foxp3 expression in the circulating T_{reg} cell subset [47]. With these evidences, a T1D disease progression model has been proposed

which considers T1D as direct consequence of the imbalance of T_{reg} cell to effector T cells [36] as shown in Fig.3.

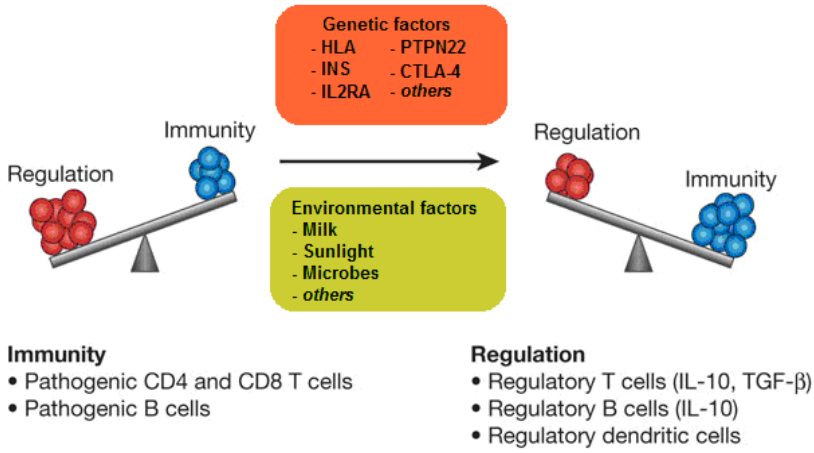


Figure 3. This schematic illustrates the fine balance of immune regulation versus pathogenesis, highlighting a number of genes that are likely to influence the balance through effects on central and peripheral tolerance and the environmental factors that control immunity. The key cell types that affect the balance locally during immune responses are listed (derived from: [36])

On the other hand, T1D researchers are trying to understand the complex gene–environment interaction: for instance, relationship between cow's milk and PTPN22 and INS or IFIH1 and enterovirus [48]. Further investigations are needed in this direction[48].

2.1.4 Rheumatoid Arthritis (RA)

Rheumatoid arthritis is an autoimmune, chronic, systemic inflammatory disorder that principally attacks synovial joints. It can be a disabling and painful condition, which can lead to substantial loss of functioning and mobility. RA afflicts up to 1% of the general population worldwide. It is clinically heterogeneous, with particular disease phenotypes defined according to a complex interplay of genes and the environment [49].

2.1.4.1 Symptoms and diagnosis

Inflammation on the affected joints is the primary sign of RA, which causes swelling, warmness, pain and stiffness. Increased stiffness early in the morning is often a prominent feature of the disease and typically lasts for more than an hour. In arthritis of non-inflammatory causes, signs of inflammation and early morning stiffness are less prominent with stiffness typically less than 1 hour, and movements induce pain caused by mechanical arthritis. In RA, the joints are often affected in a fairly symmetrical fashion, although this is not specific, and the initial presentation may be asymmetrical. The diagnosis of rheumatoid arthritis is primarily based on clinical symptoms. Typical examination findings include swelling, bogginess, tenderness and warmth of, with atrophy of muscles near, the involved joints [50].

2.1.4.2 Risk factors for developing Rheumatoid Arthritis

As a complex disorder RA risk factors include both genetic and environmental factors. It has been shown that RA clusters in families: the risk of a first-degree relative of a patient is upto 10 times more than that of the general prevalence of the disease [51]. The genetic basis of such clustering is confirmed by observations that RA concordance amongst monozygotic twins is approximately 15%, which is up to 5 times greater than in dizygotic twins. Through twin studies, RA heritability of is reported to be upto 60% [52].

Smoking is the most strongly linked environmental/life style risk factor of RA [53]. Investigations showed that smoking leads to the over production of rheumatoid factor (RF), which in turn contributes to the disease process [54]. Another factor is the influence of microflora: it has been shown that patients with early RA have different intestinal microflora than non-RA patients [55]. Notably, the intestinal microbes *P. gingivalis* has been shown to be linked with the development of immunity against citrullinated proteins due to their ability to produce citrullinated epitopes and its presence in an environment highly analogous to RA [56].

2.1.4.3 Genetics and pathogenesis in Rheumatoid Arthritis

The genetic components of RA susceptibility have been only been partially elaborated in terms of functional implications in the pathogenesis [57]. The fine-mapping of MHC polymorphisms

associates RA with HLA-B and HLA-DPB1. However, the most studied polymorphism in RA is the PTPN22, which is a non-synonymous Arg620Trp SNP rs2476601 [57]. Interestingly, PTPN22 downregulates TCR signalling by dephosphorylating Src family kinases, such as Lck or Fyn (Fig. 4). It has been shown that the PTPN22 risk allele affects the enzymatic activity of the encoded phosphatase and the reduced levels of the protein correlate with increased number, activation and thymic positive selection of T cells, and with dendritic-cell and B-cell activation [57]. Another suseptibility gene is PADI4, which mediates post-translational conversion of arginine residues to citrulline. Citrullinated peptides bind with higher affinity to HLA-DRβ1 shared epitope molecules, are naturally processed and are immunogenic [58]. Thus, it seems that increased translation of variant PADI4 mRNA boosts production of citrullinated peptides, which act as autoantigens and elicit profound adaptive immune responses activation [57]. Another gene, *CCR6* encodes a chemokine receptor expressed by $CD4^+$ type 17 T helper (T_H17) cells. A polymorphism in CCR6 is correlated with the expression level of *CCR6* mRNA and with the presence of IL-17 in the sera of patients with RA, highlighting the importance of the T_H17 pathway in RA pathogenesis [59]. IL2RA is another susceptibility gene in RA, which has been shown to be exclusively expressed in monocytes, $CD4^+$ naive T cells and memory T cells [45]. Notably, according to the *quantal theory of immunity* (proposed by Smith and colleagues in 2008), T-cell responses depend on a critical number of stimuli mediated by TCR and IL2R [60]. Another important gene is TNFAIP3, an ubiquitin-

modifying enzyme that is a key regulator of nuclear factor κB activity (Fig. 4). Within the TNFAIP3 locus, three SNPs are independently associated with RA susceptibility [61]. Like other susceptibility genes, TNFAIP3 has also been describes in other diseases. But interestingly, it has been shown that mice with conditional knockout of Tnfaip3 expression in dendritic cells develop an SLE-like phenotype, whereas mice lacking Tnfaip3 in myeloid cells develop a RA-like phenotype [62-65].

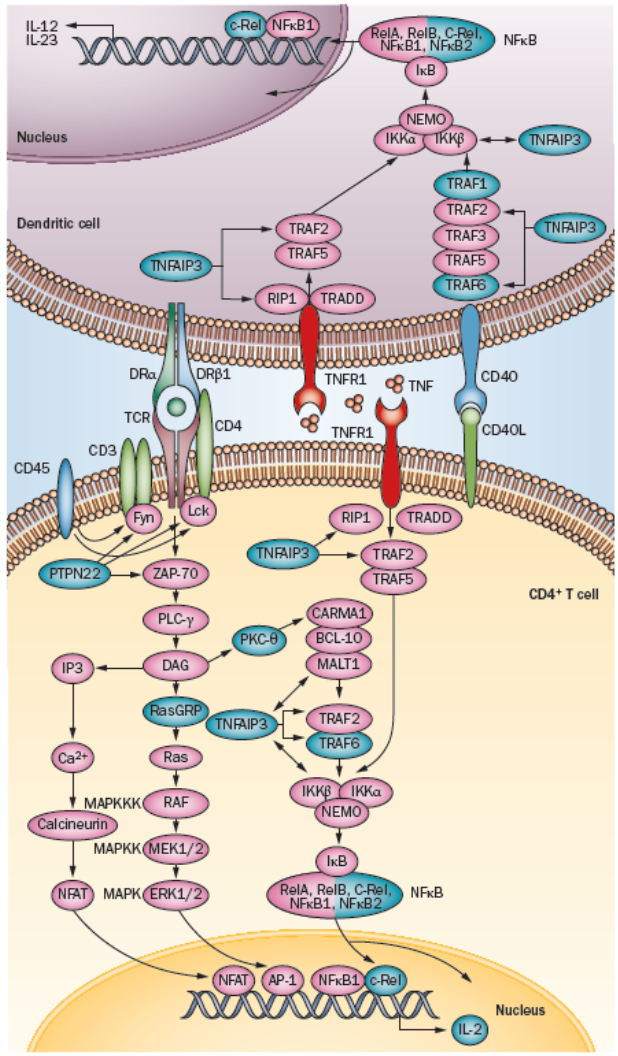


Figure 4. Mapping of RA susceptibility loci to pathways involved in the *T cell–dendritic cell* interaction. The RA susceptibility genes (blue) are implicated in TCR, TNF and CD40 signaling pathways (derived from: [57]).

2.1.5 Multiple Sclerosis (MS)

So far we have seen complex diseases which are either autoimmune or neurodegenerative. In MS, the complexity increases as it is presumably an autoimmune disease targeting the CNS. MS affects around 2 million individuals and is the leading cause of neurological disability in young adults, and there is no effective cure till date.

2.1.5.1 Historical notes on MS

The recorded history of multiple sclerosis begins with Robert Carswell (1793–1857) - a British professor of pathology, and Jean Cruveilhier (1791–1873) - a French professor of pathologic anatomy, have described and illustrated many of the disease's clinical details. However, they did not identify MS as a separate disease. It was a French neurologist named Jean-Martin Charcot (1825–1893) who recognized MS (*sclerose en plaques*) as a distinct disease in 1868, summarizing previous reports and adding his own clinical and pathological observations. Charcot also observed cognition changes, describing his patients as having a *marked enfeeblement of the memory* and *conceptions that formed slowly* [66]. He also formulated the first diagnosis criteria for MS, known as Charcot's triad: *nystagmus, intention tremor,* and *scanning speech*. The discovery of MS, along with other neurological diseases made Charcot to be remembered as the

Founder of modern neurology. [67]. The first major case reports on MS (reported as *insular sclerosis*) appeared on *The Lancet* journal on 15 February 1873 and 3-17 April / 1 May 1875.

2.1.5.2 Clinical classificaiton of MS

In the year 1996, the National Multiple Sclerosis Society (NMSS, USA) has classified MS into four clinical classes (Fig. 5). The relapsing-remitting form (RR-MS) affects approximately 85% of patients [68]. RR-MS typically starts in the early adulthood (20's or 30's) and has a female prevalence of approximately 2:1 [69]. RR-MS begins with a clinically isolated syndrome (CIS), in which a person may have experienced an attack, but does not fulfill the criteria for multiple sclerosis. However, 70% of persons experiencing CIS later will develop RR-MS. RR-MS is characterized by recurrent acute episodes of dysfunction (called relapses), followed by a partial recovery (remissions) phase [68, 69]. In approximately 70% of cases, RR-MS converts to a secondary progressive form (SP-MS) in later stages of disease [69, 69] Early symptoms in RR-MS include sensory disturbances, unilateral optic neuritis, limb weakness, gait ataxia, trunk and limb paraesthesia on neck flexion [70].

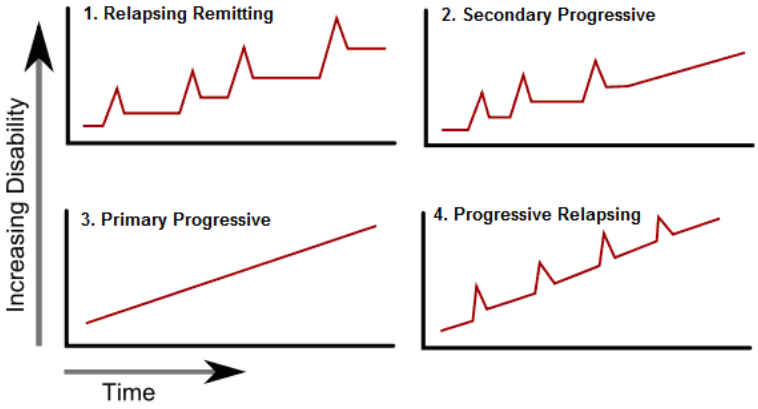


Figure 5. *Clinical classification of MS*

About 10% of total MS patients start the disease with a progressive course, which is called as primary-progressive form (PP-MS), which is characterized by progression of disability from onset, with no clear remissions and improvements. Notably, unlike other forms there is no gender bias in the incidence of PP-MS [68]. In addition, the damage is more often the spinal cord than the brain or optic nerve. PP-MS often presents with an upper-motor neuron syndrome of the legs and gradually worsens leading to cognitive impairment, quadriparesis, visual loss, bladder and sexual dysfunction [68, 69]. Finally, approximately 5% of patients experience a progressive-relapsing form of disease (PR-MS), which is characterized by a progressive onset but later associated to one or more relapses [68].

2.1.5.3 Epidemiology and diagnosis of MS

According to the prevalence of MS, world can divided into three regions. High prevalence regions comprise northern Europe, northern USA, Canada, New Zealand and southern Australia. Medium prevalence areas include southern Europe, southern USA and northern Australia. Low prevalence zones include Asia and South America [71]. Migration studies have shown that people migrating from a low prevalence region to a high prevalence region retains the low risk of their area of origin. However, the children of those immigrants acquire increased risk for MS [72]. It was also shown that the age of migration could influence the MS risk [72, 73]. As a complex disease, the prevalence and incidence of MS in different regions of the world are determined by genetic and environmental factors.

Diagnosis of MS takes advantage of both magnetic resonance imaging (MRI) and clinical observations. After the first clinical attack, dissemination of lesions in space and time can be demonstrated by the evaluation of subsequent MRI abnormalities, according to specific criteria [74]. Three types of MRI scans are commonly used to investigate MS: namely T1-weighted, T2-weighted, and proton density scans, which are obtained by manipulating the radiowave of MRI. Each provides complementary information to the neurologist about the nature of MS. For instance, T1-weighted scans along with a contrast agent (eg: gadolinium) can highlight any areas of recent inflammation that indicate active disease and areas of breakdown of blood-brain

barrier (BBB) [75]. T2-weighted scans shows both new and old lesions, and is often used when diagnosing MS. T2-weighted scans can be repeated regularly over time (snapshots over time) to understand disease progression [75]. Proton density scan can identify both old and new lesions, but they are particularly used to identify lesions near the ventricles of brain [75] .

2.1.5.4 Risk factors for developing MS

2.1.5.4.1 Genetic Factors

The genetic basis of MS has been widely accounted by the familial aggregations and incidence of disease in specific ethnics [76]. Monozygotic twins show concordance rate between 20 and 35% [77]. Some ethnic groups have the higher risk of MS such as North America, Scandinavian countries, Iceland and British Isles [68]. Like other autoimmune diseases, the major histocompatibility complex (MHC) locus is described as an important genetic factor, which accounts for 10%-60% of the genetic risk for MS [78]. Particularly, HLA-DR15 haplotype in Caucasian population (DRB1*1501, DRB5*0101, DQA1*0102, DQB1*0602) is considered the strongest risk allele for MS [79]. Other risk genes include tumour necrosis factor (TNF) cluster, transforming growth factor (TGF)-β family members, CTLA-4, IL1RA, IL-1 and ESR [77]. More recently, genome-wide association studies have discovered number susceptibility genes in MS; including IL7RA, IL2RA, TNFRSF1A, CLEC16A, IRF8, CD58 and CD6 [80]. The functional mapping of

GWAS genes in MS is still in an early stage. Although fine mapping and functional studies will be required to define the functionally relevant variants responsible for determining susceptibility to MS, the over-representation of immunological genes near associated SNPs is already evident. For instance, CD40, IL12A, IL2RA, STAT3 *and* TNFRSF1A, CBLB , CD6, CD58, CD226, SH2B3 *and* TNFAIP3 are involved in T-cell activation and IL7 signalling [81]. It is believed that the associations identified so far are likely to represent only the tip of the iceberg, with available evidence suggesting that possibly hundreds of other variants are also involved [82]. Application of systems biology may help to overcome the limited outcomes from gwas studies. Notably, Baranzini and collegues have analyzed nominally significant SNPs (*P*-value< 0.05) from two MS GWAS using a network approach based on experimental annotations of protein–protein interactions and found immunological networks, neural pathways such as those related to axon guidance and synaptic potentiation, which were significantly associated with these variants [83]. Similar approaches are within the scope of this thesis and can be found in the *results section*.

2.1.5.4.2 Environmental Factors

The medium-low concordance of MS in identical twins suggests a strong contribution from environmental factors in triggering MS [77]. Even though several environmental agents have been investigated as potential causative factors in MS pathogenesis, sunlight (in the context of vitamin D production) and viruses are the most widely

investigated and documented. It has been shown that the risk of MS among whites significantly decreased with increasing serum levels of 25-hydroxyvitamin D [84]. Among the viruses, Epstein-Barr virus (EBV) is hypothesized to be associated with MS. This virus leads to a latent lifelong infection of B cells. EBV affects almost 100% of MS patients in comparison to 94% of age-matched controls [85]. More recently, recent report has showed that there is an abnormal accumulation of EBV-infected B cells in MS brains [86]. However, research in this area needs to be developed to have more strong evidences. Other environmental risk factors include: severe stress, smoking, exposures to toxins, diet and hormone intake [72, 85, 87]. Notably, the fact that MS incidence is about 2-fold higher in women is suggestive of the role of gender-specific hormones in MS[76].

2.1.5.5 MS pathogenesis

2.1.5.5.1 Peripheral immune response in MS

Generally, inflammation in the CNS is considered to be the cause of neurodegeneration. But, recently, the independence of neurodegeneration and inflammation is highly debated, especially in progressive forms of MS [88]. Unfortunately, the specific elements that trigger inflammation are still unknown. There are two general types of immune responses: namely, *innate* and *adaptive* immune responses. The innate immune response, also known as first line of defense, is basically initiated by an infection by other organisms (eg: microbes) that activate specific receptors, mainly toll-like

receptors (TLRs) in an antigen nonspecific manner. Activation of specific subsets of TLR is done by pathogen products that are unique for different groups of pathogens and binding of these molecules to TLRs results in the production of cytokines that modulate the adaptive immune response [89]. It has been shown that this mechanism plays an important role in MS by influencing the effector function of T and B cells response [89]. For instance, when activated through TLRs, dendritic cells (DCs) become semi-mature and induce regulatory T cells to produce inhibitory cytokines such as IL-10 or TGF-β [90]. The maturation of dendritic cells makes them to polarize $CD4^+$ T cells to differentiate into Th1, Th2 phenotypes or Th17 phenotypes [91]. Importantly, when T cells differentiate to a Th1 phenotype, inflammation is promoted. Interestingly, in recent studies of the animal model of MS, the Experimental autoimmune encephalomyelitis (EAE)., it has been shown that glatiramer acetate (GA, a drug used in MS) may induce type II monocytes which promote Th2 cell production and development of regulatory T cells, which then decrease inflammation [92].

On the other hand, the adaptive response (or acquired immune system) is initiated by the presentation of a specific antigen to T lymphocytes by antigen presenting cells (APCs). APCs include B cells, dendritic cells, microglia and macrophages. The interaction between the APC and T cell is a central point of initiation the adaptive immune response, and T cell subsets such as $CD4^+$ and $CD8^+$ phenotype can be activated by APCs. Th1, Th2 cells and

Th17 are $CD4^+$ effector cells that are polarized in response to exposure to specific interleukins [91]. Once polarized to Th1, Th2 or Th17, these effector T cells secrete specific cytokines. The cytokines produced by Th1 cells are proinflammatory cytokines (eg: interferon gamma), but Th2 cells secrete anti-inflammatory cytokines (eg: IL-4, IL-13). On the other hand, Th17 produces IL-17, IL-21, IL-22 and IL-26. Like Th1 cells, Th17 cells promote inflammation in MS. IL-17 receptors are seen in acute and chronic MS plaques [91]. It has been shown that in IL-17 deficient mice, there is a reduction of clinical severity [93]. Th1 cells and Th17 cells migrate to the central nervous system from periphery and causes demyelination and axonal loss [89].

In addition, the role of regulatory T cells (T reg) has been also studied in the pathogenesis of MS [94]. T reg cells regulate effector Th1, Th2 and Th17 cells. The number of T reg cells is the same between MS patients and controls, however, it has been shown that patients with MS have reduced T reg function [94]and that GA can increase regulatory T cell function by increasing the expression of naive $CD4^+CD25^+FoxP3^+CD31^+$ T cells [95].

Besides $CD4^+$ T cells, studies have shown that $CD8^+$ T cells are present in MS lesions and may have regulatory function in the progression of disease. $CD8^+$ cells mediate suppression of $CD4^+$ T cell proliferation through the secretion of perforin, which is cytotoxic on $CD4^+$ T cells, leading to their inactivation [90, 96]. Moreover, $CD8^+$ T cells is capable of transecting axons,

promote vascular permeability and activate oligodendrocyte death [90].

Finally, role of B cells in the pathogenesis of MS is also documented. B cells are important for the production of antibodies against myelin antigens. It has to be noted that B cells produce proinflammatory and anti-inflammatory cytokines [97].. In addition, it has been shown that B cell follicles are found in the brains of MS patients, and these follicles express CD20 [98].

2.1.5.5.2 Infiltration of autoreactive T cells in CNS

The CNS is an immune-privileged site which is protected by the blood-brain barrier (BBB), a specialized barrier made of tight junctions between endothelial cells of blood vessels, the basement membrane of vessels, astrocyte feet and microglial cells [99] . However, it has been shown that in physiological conditions, activated T cells and memory T cells express adhesion molecules and chemokine receptors necessary for crossing these barriers and perform CNS immune surveillance [100]. A key molecule regulating the entry of $CD4^+$ T cells in the CNS is the integrin very late antigen (VLA)-4. VLA-4 is shown to be present on perivascular T cells of acute MS lesions, and VLA-4 binds vascular cell adhesion molecule (VCAM)-1, that is expressed by activated endothelial cells and mediates adhesion of $CD4^+$ T cell [101]. Once the blood-CSF and blood-brain barriers is crossed, immune cells can diffuse into the white matter of the CNS, and target myelin sheath as

observed in EAE and MS. In this regard, a model has been proposed based on the studies in the animal model of MS, namely experimental animal encephalomyelitis (EAE), as shown in the Fig.7 [102].

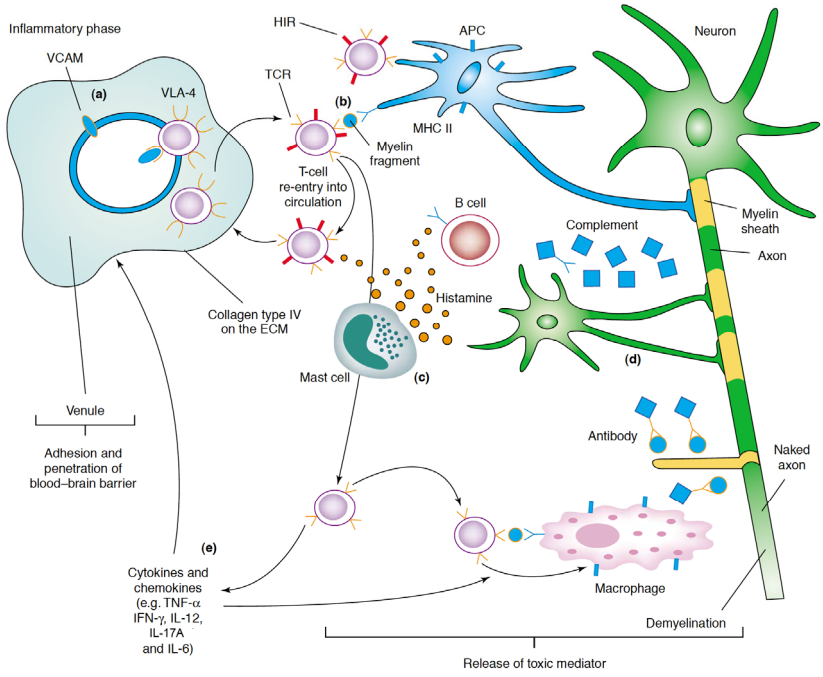


Figure 6. EAE serves as a model for proof of concept of new MS therapies. Interventions at various checkpoints in the pathophysiology of EAE, and presumably MS, are shown. T and B cells penetrate the blood vessel endothelium. (a) The key molecule in adhesion is α4β1 integrin on T and B cells. This integrin binds to VCAM and the lymphocytes diapedese, crawling through and penetrating the extracellular matrix. Matrix metalloproteases (MMPs) are crucial for this process. MMPs can be inhibited with IFN-β. Antibodies to α4β1 integrin inhibit the adhesion step. (b) Once inside the brain, T cells recognize myelin fragments in association with class II molecules of the MHC. The expression of these molecules, including MHC II and co-stimulatory molecules, such as CD80, CD86 and CD23, can be inhibited by statins and

PPAR agonists. APLs and glatiramer can inhibit the interaction of MHC II with T cells. (c) Mast cells also have a role in the inflammatory response in MS. In EAE, histamine antagonists and platelet-activating factor (PAF) antagonists can prevent EAE. (d) Antibodies to protein and lipid components of the myelin sheath can activate complement, culminating in the production of membrane attack complexes, which damage the oligodendrocyte and lead to the stripping of myelin by activated macrophages. (e) Destructive cytokines, such as IL-6, osteopontin, TNF and IFN-γ, amplify the inflammatory response in the brain. Some of these cytokines have Janus-like activities, both inducing pathology but also having key roles in recovery (derived from: [102])

2.1.5.6 Neurodegeneration in MS

The pathological hallmark of MS consists of multiple plaques of demyelination, located more often in the optic nerves, periventricular white matter, corpus callosum, brain stem, cerebellum and spinal cord [69]. According to structural and immunopathological features, MS lesions have been distinguished and classified into *acute active, chronic active, smoldering active* and c*hronic silent* plaques [103]. There is a correlation between CNS inflammation and and the frequency of transected axons in MS [104]. There are a number of substances that could injure axons, such as proteolytic enzymes, cytokines, oxidative products, and free radicals produced by activated immune and glial cells [105]. One of the key enzymes is iNOS, which is involved in synthesis of nitric oxide (NO). iNOS is upregulated in acute inflammatory MS lesions and elevated levels of NO can have a detrimental effect on axonal survival by modifying the action of key ion channels,

transporters, and glycolytic enzymes [106]. Further, NO can limit axon's ability to generate ATP. It was shown that cytotoxic CD8+ T cells can mediate axonal transection in active MS lesions [107]. Inflammation reduces energy metabolism in demyelinated axons and inflammatory intermediates may act directly on the mitochondria, and local inflammatory edema. Axoplasmic changes resulting from chronic demyelination can reduce ATP production [108].

Even though axonal loss is extensive in acute MS lesions, during early stages of RRMS only minor permanent axonal loss is seen. This is due to the plasticity of the human CNS which compensates for neuronal dysfunction and loss. Probably, the conversion of RR-MS to SP-MS happens when the CNS can no longer compensate for additional neuronal loss [104]. However, at later stages of MS, progressive and irreversible disability and brain atrophy often occur in the absence of new inflammatory demyelinating lesions [109]. Therefore, mechanisms other than inflammatory demyelination of white matter could be contributing to axonal degeneration.

2.1.5.7 MS therapeutics

All the available therapies are either immunomodulating and immunosuppressive treatments. These treatments differ in their selectivity for the aspects of the immune system that they target and in their selectivity for immune dysfunction specific to multiple sclerosis.

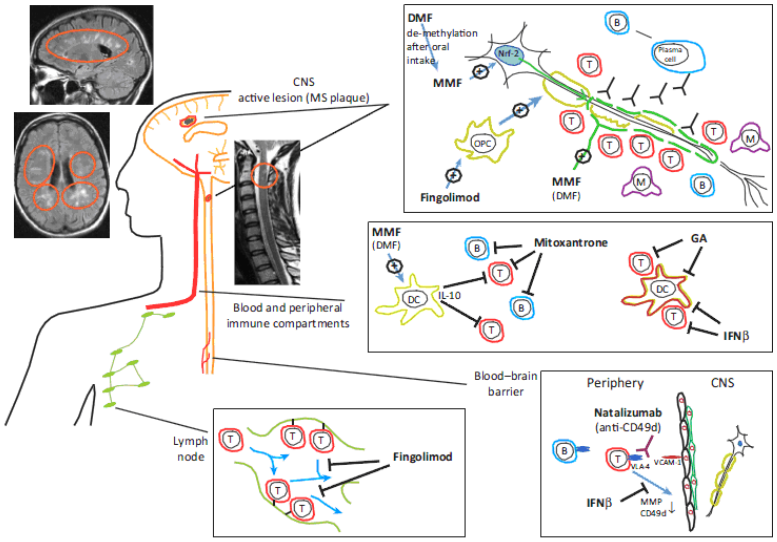


Figure 7. MS patho-mechanism and targets of established drugs. The torso on the left side schematically represents the different compartments involved in MS pathogenesis, as shown enlarged with the candidates involved on the right side, including the central nervous system (CNS) with autoinflammatory lesions or plaques, the blood–brain barrier, the peripheral blood/immune compartments (e.g., thymus, spleen), and lymph nodes. The images exemplify how inflammatory lesions are detected by MRI (red circles): lower left shows a transversal image (T2-FLAIR, fluid attenuated inversion recovery image), upper left a sagittal image of a MS-CNS; right side image shows a sagittal spinal cord lesion (T2). The modes and sites of action of currently established MS therapeutics are illustrated in the respective compartments (derived from : [110])

The new era in MS therapy began with interferon beta-1b (IFNb), and then GA. Many targets have been proposed for GA and despite conflicting evidence, the inhibition or modulation of antigen-presenting cells is the currently favored model, altering

effector T and B cell responses as well as regulatory T cells that inhibit T cell autoreactivity [111, 112]. In the case of IFNb, evidence suggests the induction of anti-inflammatory effects such as reduced BBB permeability and the inhibition of autoreactive lymphocytes (Fig.7) [113, 114]. Mitoxantrone is an anthracenedione that causes DNA strand breaks by intercalation, delays DNA repair by inhibiting topoisomerase II, and reduces lymphocyte numbers (Fig.7). Despite these severe adverse events, mitoxantrone remains a therapeutic option, particularly for SPMS, for which there are few alternatives. Natalizumab is a monoclonal antibody (mAb) against the a4 subunit of the VLA-4, and its therapeutic potential was first shown in the EAE model and then in MS patients [115]. VLA-4 is expressed on virtually all activated leukocytes and enables them to cross the BBB, an activity that is blocked by natalizumab (Fig.7). Fingolimod is the first oral drug approved for RR-MS. It binds to spingosine-1-phosphate (S1P) receptors, inhibits S1P1-mediated lymphocyte egress from secondary lymphoid organs, and prevents lymphocytes from encountering antigens within the CNS. Experimental evidence also suggests beneficial roles within the CNS, where it promotes the survival of oligodendrocytes and may lead to remyelination in damaged white matter via auxiliary S1P receptor subtypes (Fig.7) [116]. DMF was recommended for approval by the EMA and the FDA in March 2013, final approval is expected in June 2013. BG-12, a fumaric acid ester, is an oral drug approved as first-line therapy for RR-MS. Experimental evidence implies that DMF

may have immunomodulatory and neuroprotective effects by activating antioxidative pathways, ultimately protecting neuronal and glial cells from inflammation-induced cell death (Fig.7) [117].

The immunological, clinical and pathological data suggest that early treatment of MS patients with immunomodulatory drugs is advisable. Further, a better understanding of the mechanisms underlying axonal damage during both the acute inflammatory phase and the chronic degenerative phase can help in the development of neuroprotective drugs.

2.1.5.8 Animal model of MS - Experimental Animal Encephalomyelitis (EAE)

Over the years EAE has been induced in a wide range of species, including mouse, rat, guinea pig, rabbit, hamster, dog, sheep and marmoset [118]. The most common EAE models currently employ rats and mice and different clinical courses have been reported, according to the strain of mice and peptide used [119]. In the majority of the models, EAE clinically manifests as ascending-flaccid paralysis starting from the tail and progressing to hind and forelimbs [119]. In C57BL/6 mice strain (bearing H-2^b haplotype of MHC), EAE can be induced actively by subcutaneous administration of MOG_{35-55} peptide in CFA and by intravenous or intraperitoneal injection of PTX. These mice develop EAE with a chronic clinical course of paralysis [120]. Active immunization with

$PLP_{139-151}$ induces a relapsing-remitting EAE in SJL-J mice (H-2^s), while in other strains such as PL/J (H-2^u) an acute monophasic disease is observed [120]. Actively induced EAE consists of an induction phase, which involves the priming and activation of myelin-specific $CD4^+$Th1/Th17cells, and an effector phase. During the effector phase, encephalitogenic $CD4^+$ T cells migrate to the CNS, where they are re-activated by APCs, and orchestrate an immune-mediated attack of myelin. Immune cells recruited in EAE lesions include macrophages, $CD8^+$ T cells, B cells and plasma cells, thus resembling the neuroinflammatory milieu observed in MS plaques [119].

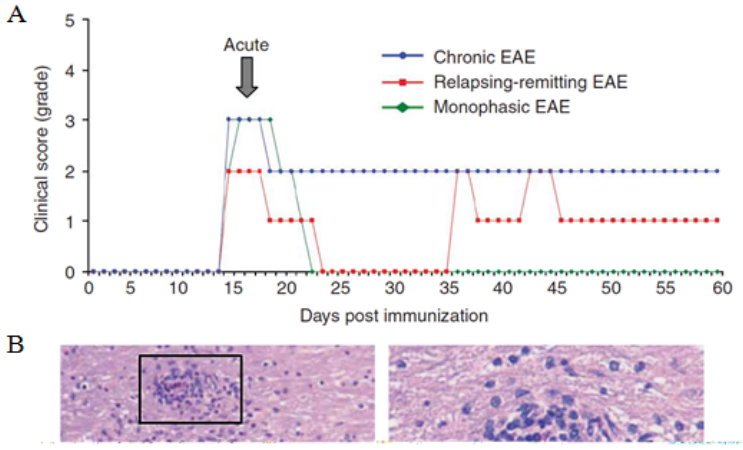


Figure 8. Examples of different clinical courses of EAE. (A) Induction of EAE in different mouse strains and with different myelin epitopes result in chronic, relapsing-remitting or acute monophasic disease. Clinical signs are graded according to a 0-5 point scale (B) Representative tissue section of spinal cord from mice at the onset of EAE, stained with haematoxylin and eosin. Inflammatory cells infiltrates in the white matter are boxed, whereas a perivascular lesion is presented at higher magnification (derived from: [119]).

EAE has been useful for better characterizing the effector mechanisms of CNS autoimmune responses, and exploring new therapeutic strategies for MS. However some limitations have emerged as a model of MS. Infact, EAE is induced by administration of non-physiological stimuli and consequently is not useful to understand the initiating events occurring in MS [121].

2.2 Genome-wide association studies

A genome-wide association study (GWAS) scans markers across genomes of large cohorts to find genetic variations associated with a particular disease. Such studies are extremely useful in finding genetic variations that contribute to complex diseases [122]. With the completion of the Human Genome Project in 2003 and the International HapMap Project in 2005, researchers now have a set of research tools that make it possible to find the genetic contributions to common diseases. Till date, more than 1,200 human GWA studies have examined over 200 diseases and traits, and almost 4,000 SNP associations have been found [122].

2.2.1 GWAS methods

In general, GWA studies are conducted in a case-control setup which compares two large groups of individuals: one healthy control group and a diseased group. All individuals in each group are genotyped for the majority of common known SNPs, based on the genotyping technology [122]. Each of these SNPs is then investigated if the allele frequency is significantly altered between the case and the control group [123]. The basic unit for reporting effect sizes is the odds ratio, which is the ratio between proportion of individuals in the case group having a specific allele, and the proportions of individuals in the control group having the same allele. The odds ratio will be greater than 1 if the allele frequency

in the case group is higher than in the control group, and vice versa for a lower allele frequency. Further, a statistical test is performed (chi-squared test is widely used) to obtain the significance of the odds ratio. Further, the p-values are corrected for multiple testing [122, 123]. After determining odds ratios and p-values have been calculated for all SNPs, data are represented by a Manhattan plot, in which negative logarithm of the p-value will be shown as a function of genomic location [124]. So, most significant association will have the highest peak in the plot.

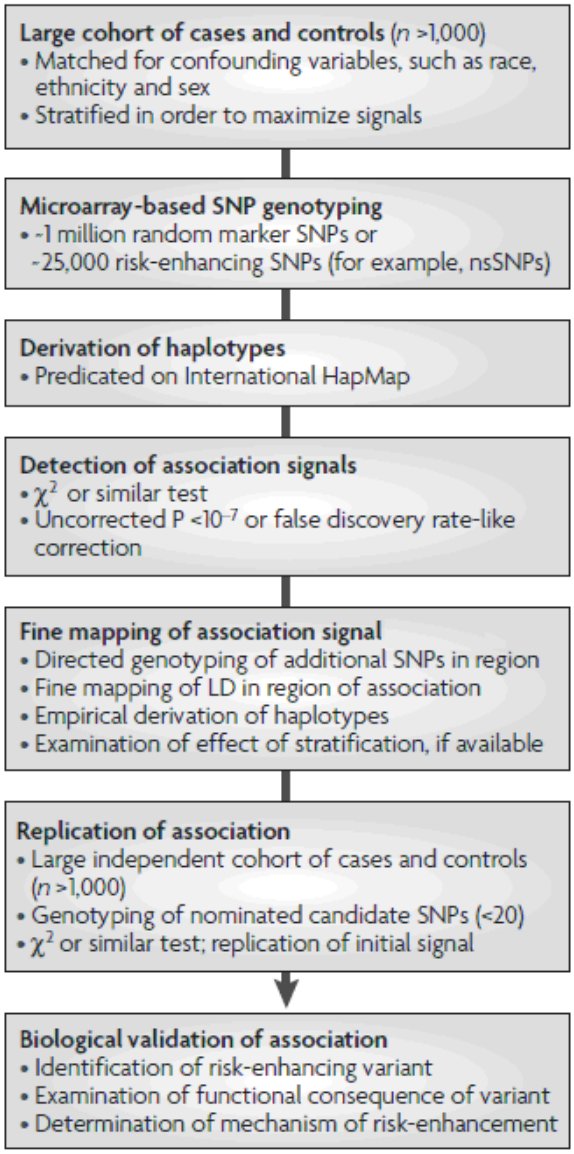


Figure 9. *Overview of the general design and workflow of a genome-wide association (GWA) study using microarrays.* The discovery phase entails genotyping many case and control DNA samples and evaluation for significant associations. The replication phase involves fine mapping of association signals and independent confirmation in a second cohort. Biological validation is important for translation of GWA findings into diagnostic or therapeutic discoveries. (Derived from: [125])

2.2.2 The past, present and future of GWAS

The first successful GWAS was of age-related macular degeneration, with 100,000 SNPs tested for association in 96 cases and 50 healthy controls [126]. In the year 2007, a landmark study was conducted in seven common diseases, including bipolar disorder, coronary artery disease, Crohn's disease, hypertension, RA, and diabetes, at the Wellcome Trust Case Control Consortium (WTCCC). For each disease, around 500,000 SNP genotypes of 1500–2000 cases were compared to 3000 control samples. This study discovered several risk loci, and confirmed previously implicated ones [127]. Following WTCCC study several studies were reported in a short span of time. A series of remarkable studies in T1D accounted for about 80% of genetic variation in T1D [128-130]. In the recent times gwas doesn't restrict to pathological conditions. A recent GWAS meta-analysis of nearly 180,000 individuals identified 200 loci explained approximately 14% of height variation and another study was elaborated the genetic basis of body mass index [131, 132]. Notably, the National Human Genome Research Institute has catalogued the results from all published gwa studies. The database is frequently updated, and is available at http://www.genome.gov/gwastudies/. We have utilized data from this resource for our study.

On the other hand, GWA studies have several limitations. The main issues are: insufficient sample size, lack of proper statistical methods, and control for population stratification [133].

The future of gwas will be to apply the findings in a way to better integrate genetic studies into the drug-development process and diagnostics. Notably, a recent technical advancement happened in the gwa studies through the application of next-generation sequencing technology (eg: 454 sequencing, Illumina Hi-seq etc), which may enable the discovery of novel SNPs that are not included in the microarrays. From a functional point of view, the gwas data are limited to offer the downstream effects of the associated genes, because of the lower number of reported associations [134]. However in the last two years some studies has demonstrated that the network analysis can group genes into modules that participate in specific pathways or biological processes [135, 136]. Another evolving area is systems genetics, which is an approach for studying complex genetic traits in which genomic data and clinical phenotypes are obtained using global "omic" technologies such as gene expression arrays, mass spectrometry, and sequencing [137].

2.3 TRANCRIPTOME ANALYSIS

2.3.1 Oligonucleotide microarray

A microarray is a high-throughput technology consisting of an arrayed series of thousands of microscopic spots of oligonucleotides (*probes*). Each feature contains picomoles of a specific DNA sequence. This can be a short section of a gene or other DNA element that are used as probes to hybridize a cDNA or cRNA sample (*target*s). Before the experiment the targets are of fluorescence labeled using dyes. The *probe-target* hybridization is usually detected and quantified by utilizing the intensity from the fluorescence and the abundance is determined [138]. Microarrays are typically used in molecular biology and in medicine to monitor levels of thousands of genes simultaneously and study the effects of certain treatments, diseases, and developmental stages on gene expression. The quantified data (*raw data*) is to be processed before analysis. The commonly used microarray platforms are Affymetrix Sentrix arrays and Illumina beadarrays.

2.3.2 Illumina Beadarray Technology

We have used Illumina microarray for the transcriptome analysis in peripheral blood mononuclear cells (PBMC) from MS patients and healthy controls. Illumina's beadarray technology is based on color-coded 3-micron silica beads that randomly self assemble in either a fiber-optic bundle substrate that then themselves assemble

into arrays, or a silica slide substrate. Each bead is covered with hundreds of thousands of copies of a specific oligonucleotide that act as the capture sequences in one of Illumina's assays. Each bead has a 23-mer oligo *address* attached to it, which then anchors a 50-mer sequence-specific *probe* (Fig 10).

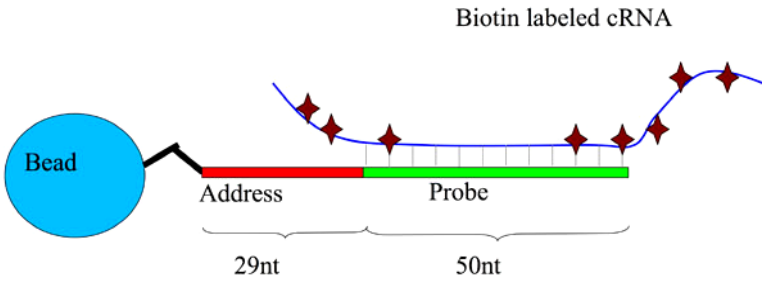


Figure 10. **Illumina bead and probe**. The beads are randomly scattered across etched substrates during the array production process, with each array bundle containing about 50,000 beads. With this platform design, a specific oligonucleotide sequence is assigned to each bead type, but is replicated about 30 times on the array at random positions. Each gene is represented by two probe sequences. A series of decoding hybridizations are used to determine which oligos are present at each matrix coordinate for every array.

2.3.2.1 Illumina array data analysis

2.3.2.1.1 Data processing

For microarray data processing, Illumina provides *GenomeStudio* software (earlier called as *BeadStudio*) along with the instrument. The raw data obtained from the probe intensities need to be processed before differential gene expression analysis. A background correction calculation is then used to correct for non-specific or random contributions to the overall signal. Further, the software reports a detection p-value, which represents the confidence that a given transcript is expressed above the background defined by negative control probes (non-specific hybridization). The p-value represents the significance that a given probe is detected beyond the background level. A commonly used threshold is detection p-value of 0.05. An important step in in data processing is normalization. The GenomeStudio software provides normalization methods such as *CubicSpline, Rank-Invariant*, and *Quantile* normalization. The normalization process is essential to remove systematic biases due to sample preparation, variability in hybridization, spatial effects of the chip, settings of the scanner etc. Finally, the data is transformed into a log2scale or log2ratio in order to reduce the variance.

It is a good practice in microarray analysis to perform a quality control of the processed data in order to assess the consistency of the samples, in which the cluster analysis and principal component analysis are useful methods. Among

different cluster models (such as connectivity models, centroid models and graph models and subspace models), connectivity models are widely used. A cluster is built using a bottom-up (agglomerative) or top-down (divisive) approaches. The most popular cluster method is hierarchical clustering (connectivity model, usually agglomerative), the output is a tree classification composed by increasing number of nested classes that can be easily viewed and interpreted. The hierarchical cluster structure is easy to understand and provides very useful information about the relationships between samples/genes. Principle component analysis (PCA) is used to visualize the variability in the data by means of principal components. To each principal component a certain fraction of the overall variability of the data is attributed such that each successive component determined accounts for less of the variability than the previous one. In a typical PCA map, only the first two or three principal components are represented. Together, hierarchical cluster analysis and PCA provides a enables a highly efficient quality control in microarray experiment. In addition, these tools are invaluable in downstream analysis, in terms of data representation, class identification etc.

2.3.2.1.2 Differential expression analysis

Transcriptional profiling is aimed to identify the genes whose levels of expression change significantly between two or

more conditions or time-points. A rejection threshold must be defined, and only those having a p-value lower to the threshold will be selected for the downstream analysis. In the last decade, a number of methods have been developed, which includes both parametric and non-parametric statistics. Some commonly used methods are t-test, ANOVA, Bayesian methods – Limma/BADGE, SAM, Mann-Whitney test etc. As microarray contains several thousands of probes, a multiple test correction is usually used for the calculated p-values. Commonly used approaches are Benjamini-Hochberg's (FDR) method and Bonferroni's correction. In addition, generally, the fold change criteria are also imposed along with the statistical significance to determine the differentially expressed genes. However, expression data containing far extreme expression values in few samples may wrongly influence the accurate estimation of fold changes. Fold change values can be calculated by taking the ratio expression (in case unlogged data) or the difference between the expressions (in case of data in log scale).

2.4 FUNCTIONAL ANNOTATION AND NETWORK-BIOLOGY ANALYSIS

2.4.1 Functional annotation

2.4.1.1 Gene Ontology

In order to understand and interpret the biological or clinical relevance of the enormous data generated by high-throughput technologies, the Gene Ontology (GO) consortium has been established annotations linking primary data to expressions in controlled, structured vocabularies of genes [139]. The Gene Ontology project provides ontology of defined terms representing gene product properties. The ontology covers three domains: *biological process*, *molecular function*, and *cellular component*. The *biological process* comprises the molecular events pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms. *Molecular function* describes the elemental activities of a gene product at the molecular level, such as binding or catalysis. *Cellular component* refers to the part of a cell or its extra-cellular environment. Altogether, they describe all attributes of genes or gene products resulting from a typical high-throughput experiment. Finally, the GO content is not static. Additions, corrections and alterations suggested by research communities, as well as by those directly involved in the GO project contribute to the growth of GO database.

2.4.1.2 Pathway databases

A biological pathway is a series of actions among molecules/genes that leads to a certain product or a functional change in a cell. There are many types of biological pathways. Some of the most common are involved in *metabolism* (eg: glycolysis pathway, oxidative phosphorylation), *regulation* (eg: Ribosome biogenesis, Ubiquitin mediated proteolysis) and *signal transduction pathways* (eg: Calcium signaling pathway, MAPK signaling pathway). There are several databases that provide high-quality information about these pathways and genes involved in those pathways, for a number of species. Many of them are freely available (eg: KEGG database, PANTHER database) and some are commercial (eg: Ingenuity pathways, MetaCore pathways).

2.4.1.3 Transcription factor binding sites

Transcription factors (TFs) are regulatory proteins that bind to genomic DNA. Typically, TFs bind in close proximity to a gene, in order to activate or repress the gene expression. TFs bind to specific short DNA sequences of about 4 to 30 base pairs long (in most cases). A collection of such DNA binding sites, referred to as a DNA binding motif, can be represented by a consensus sequence. Few specialized databases (e.g. TRANSFAC, JASPER) contain the information about such sites for each TF, or a group of closely related TFs. In these databases, the binding motifs are aligned and

stored in the form of a matrix called as position-specific scoring matrix (PSSM) [140].

2.4.1.4 Functional enrichment analysis

Typically, a genome-wide high-throughput experiment gives a list of genes/proteins that may be functionally relevant to the biological/clinical conditions of the samples analyzed. As single genes are less informative, a functional enrichment analysis is typically performed on the outcome of those experiments. As described above, the GO, pathways and transcription factors provide much detailed information about the underlying biology in a specific condition or disease. Several tools are for functional enrichment study which carries out GO, pathway and TF analyzes in a single run. Given the subset of significant genes from the experiment (e.g. all differentially expressed genes from a microarray experiment), these programs identify which GO terms, pathways, or TFs which are most highly associated with this subset. This is done by performing a statistical test (e.g.: Hypergeometric test, Fisher's test) to prove that the association (enrichment) is significantly different from one expected by chance. Typically, a post-hoc test (e.g.: FDR, Bonferroni's correction) is also performed for multiple testing correction. Some of the free tools are Genecodis, ToppGene suite, DAVID etc, and the popular commercial tools include: Ingenuity pathway analysis and MetaCore.

2.4.2 NETWORK BIOLOGY ANALYSIS

The concept of systems biology emerged over the last two decade, which aims to characterize biological systems comprehensively as a complex network of interactions between the system's components. Network biology has become a core research domain of systems biology, in which cells are envisioned as complex webs of macromolecular interactions, referred as *interactome network* [141].

2.4.2.1 Protein-protein Interaction (PPI) databases and analysis

The study of PPIs are becoming increasingly important in our effort to understand complex diseases on a system-wide level. Traditionally, PPIs have been measured using a variety of assays, such as immuno-precipitation and yeast two-hybrid (Y2H). Recently, high-throughput techniques have been employed to identify protein complexes using affinity pull-down followed by mass spectrometry [142]. Further, systematically constructed double-knockout strains in yeast have proven to be useful for constructing a large-scale view of genetic-interaction networks [143]. However, the experimental reconstruction of the entire network of PPIs within a cell still remains a challenge. To complement these experimental techniques, a number of computational methods have been developed that include algorithms that are capable of predicting interactions based on gene co-expression data (expression of genes that are spatially and temporally coordinated)

[143]. Additional prediction techniques include the association method, the Maximum Likelihood Estimation technique, parsimony-explanation approach etc. [144]. Both experimental and computational methods for identifying interacting proteins have defined hundreds of thousands of protein interactions, which are systematically collected and stored in specialized databases. In the year 2000, the first among these databases was released for public access, namely, *the database of interacting proteins* - DIP [145]. Since then, a large number of further interaction databases have evolved such as BioGRID, MINT, BIND, HPRD etc, each containing some unique features (Table 1).

PPI database	Number of Interactions	Salient features
HPRD	>38,000	Information on phosphorylation motifs, signaling pathways, domain architecture, protein functions, enzyme–substrate relationships, sub-cellular localization, tissue expression and disease association of genes
BioGRID	>200,000	Interaction directionality, phenotype, post-translational modification, domains and motifs are being added
BIND	>100,000	Includes information on genetic, biopolymer–biopolymer and protein–small molecule interactions and DNA, RNA and protein sequences
MINT	>100,000	Represents complexes, bio-molecules, detailed experimental descriptions and protein structure information
DIP	>57,000	Provides experimental-quality assessments to identify the most reliable interactions; represents complexes
Reactome	>2900 reactions	Extensive cover of human pathways in 46 domains of human biology; hypergeometric testing is used to display statistically over-represented events in the event hierarchy

Table 1 : Major PPI databases (derived from: [144])

More recently, bioinformatics tools have evolved which provides a unified platform by integrating data from these PPI databases and provides tools to visualize the networks. Among these tools, VisANT is a highly user-friendly free tool which offers navigation of database-driven interaction and association networks, as well as

manipulation and storage of uploaded user-defined data [146]. The core interface of VisANT is a workbench for network analysis and visualization. In addition to simple networks, interactions in VisANT can also be defined as higher-level connections between groups of proteins, complexes, pathways or sub-networks. These modular connections can be viewed simultaneously with connections between subcomponents, such as individual protein interactions (Fig 11). Interaction networks and protein complexes can be viewed, e.g. within the context of GO annotations or KEGG pathway assignments. VisANT, also have developed a preliminary standard for exchanging files that have visual markup and annotation of network layouts, called visML. As a network specification format, visML extends concepts of similar graph languages, such as graph markup language, but contains additional features for complex and compound network components.

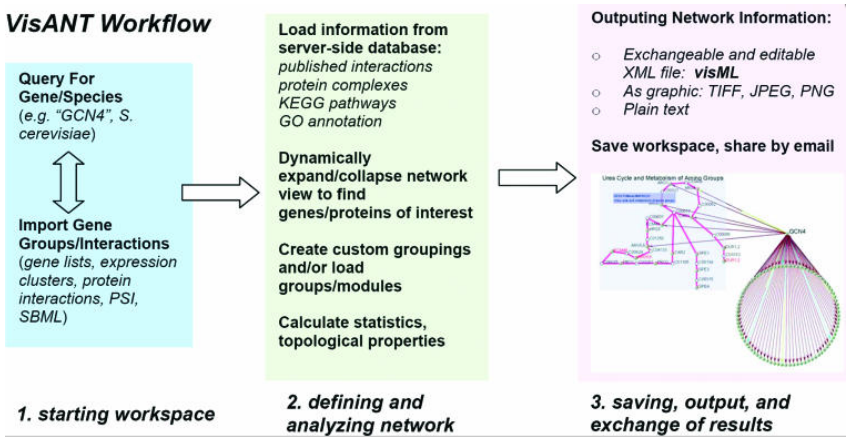


Figure 11. Workflow of VisANT

Users of VisANT can input several basic data types, including data in standardized network and interaction data exchange formats, such as PSI-MI, BioPAX and SBML pathway formats. Output from VisANT can be saved at any point once a network has been loaded, annotated and analyzed. There are three types of output: *Graphical* (JPEG, PNG and TIFF), *VisML* and *Network statistics*.

3. AIM OF THE STUDY

3.1 Analysis of interactomes linked to susceptibility genes in five complex human diseases

a. To derive the genetic interactome in MS, T1D, RA, PD and AD.

b. To understand the shared molecular pathways among these diseases.

3.2 Gender-based transcriptomics in MS

a. To understand the contribution of gender to global gene expression in PBMC from MS and healthy individuals**.**

b. Sex-specificity of differentially expressed genes in MS**.**

c. To understand the biological themes of differentially regulated genes in RR-MS patients in a systems biology context.

d. To identify new molecular drug targets in MS, and to validate them in the animal model of the disease.

4. MAIN RESULTS

4.1 Shared interactomes among five diseases

4.1.1 Link between MS interactome and genes predisposing to other neurodegenerative or autoimmune diseases

Initially we identified the susceptibility genes that were linked to MS and to other neurodegenerative (Park, Alz) or autoimmune (T1D, RA) disorders. We utilized results from 39 published genome-wide association studies on these five human diseases available at GWAS catalog and found genetic mutations in 179 genes passing the statistical significance threshold of 10^{-5} (Table 1). There were 6 studies available for MS, Park and RA, 8 for T1D, and 13 for Alz. Notably, the number of studies carried out in each disease did not seem to influence the total number of disease-associated genes. For instance, despite the highest number of genome-wide association studies in Alz, the total number of susceptibility genes reported was the lowest among the five diseases. The ratios between the number of the reported genes and the number of studies in each disease accounted for this observation (Table 1).

	MS	Park	Alz	T1D	RA	Total
No. of studies	6	6	13	8	6	39
Unique genes	54	35	17	52	21	179
Gene/study ratio	9	5.83	1.30	6.50	3.50	4.58

TABLE 1: Summary of gwas data.

Interestingly, MS displayed the highest number of susceptibility genes and the highest ratio, suggesting a greater genetic heterogeneity in MS than in the other four diseases. Then, as a first step towards the definition of the genetic network underlying these five diseases, we reconstructed the MS genetic interactome and checked possible links with the genes predisposing to the other neurodegenerative or autoimmune diseases. Using the VisANT tool we derived 376 first-degree interacting partners for the 54 MS gwas-genes. Among these MS interactors, 141 were connected with at least one among 17 T1D, 11 RA, and 10 Park or Alz gwas-genes (Figure 1A). Notably, 4 MS interactors were RA (HLA-DQA1, TRAF1) or T1D (IL2, PTPN11) gwas-genes themselves. It was also evident that several gwas-genes could contact two or more MS interactors (Figure 1A), and 24 MS interactors were connected with both neurodegenerative and

autoimmune genes (red nodes in Figure 1A), including the PTPN11 gene.

Now, the analysis of disease relatedness with MS is more effective when considering the genetic interactome rather than the direct interactions at the gwas-gene level. In fact, for example, no interactions existed between gwas-genes predisposing to MS and Alz, but there were several shared interactions within the genetic interactomes linked to (but not including) the respective gwas-genes (Figure 1B). Therefore, the introduction of the first-degree interactors in the definition of a disease-related molecular framework may lead to the discovery of relatedness among distinct complex disorders.

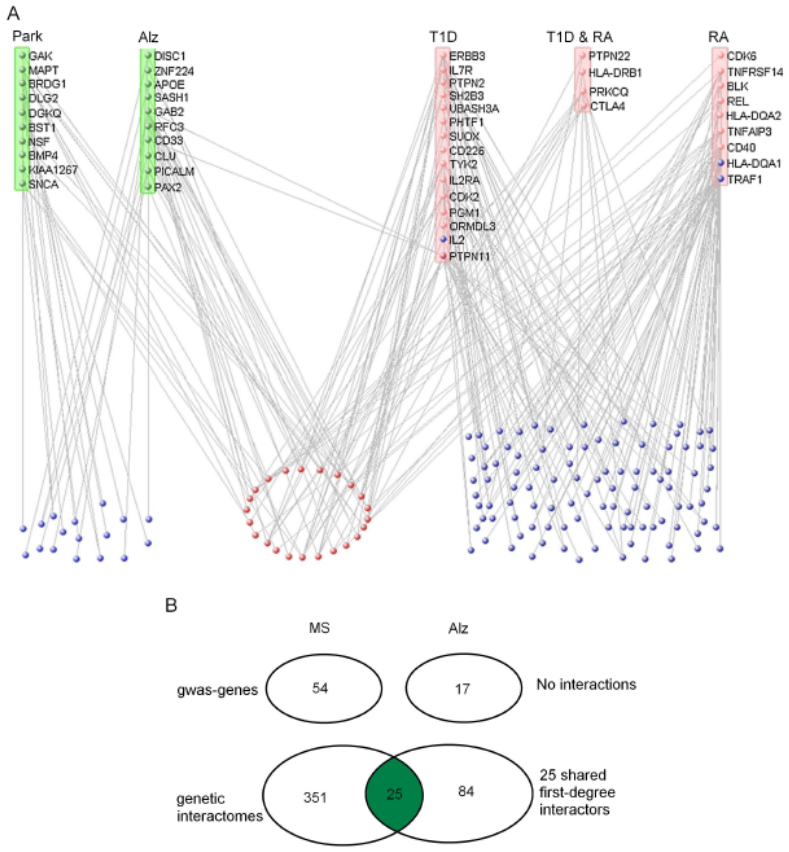


Figure 1. A. MS interactors shared by genes predisposing to neurodegenerative (green) and/or autoimmune (pink) diseases. The blue nodes indicate MS interactors contacted by either neurodegenerative or autoimmune susceptibility genes. Red nodes indicate MS interactors contacted by both neurodegenerative and autoimmune susceptibility genes. Note that four MS interactors are present among the T1D or RA gwas-genes. B. Comparison between MS and Alz at gwas-gene and genetic interactome levels. Despite the absence of a direct interaction at gwas level, the genetic interactome highlights the shared molecular networks (green section).

4.1.2 Molecular relatedness among autoimmune and neurodegenerative genetic interactomes.

To perform a global comparative analysis among autoimmune and neurodegenerative genetic interactomes, we derived the list of interacting partners for each gwas-geneset (Table 2). As expected, the number of interactors was higher in MS and T1D, due to the more abundant number of reported gwas-genes. However, when normalizing the number of interactors for the total number of gwas-genes in each disease, RA reported the highest ratio among the five diseases, indicating that at least some of the RA susceptibility genes were highly interacting. In contrast, Park displayed the lowest interactor/gene ratio, despite the discrete number of described gwas-genes.

	Genes	Interactors	Interactor/ gene ratio
Park	35	160	4.57
Alz	17	109	6.41
T1D	52	316	6.07
RA	21	249	11.85
MS	54	376	6.96

TABLE 2: Genetic interactomes for neurodegenerative and/or autoimmune disorders based on VisANT database.

Next, we examined the relationships among the five diseases by comparing the genetic interactomes in a pairwise fashion. Notably, higher concordances were found among the autoimmune diseases, with MS-RA, MS-T1D and T1D-RA having 84, 61 and 60 interactors shared respectively. Among the neurodegenerative

diseases, the sharing of interactors was limited to 10 elements for Park-Alz, 20 for Park-MS and 25 for MS-Alz. Most surprisingly, 26 interactors were shared between the neurodegenerative Alz and the autoimmune T1D. The statistical significances for these observations were calculated using hypergeometric test and are shown in Figure 2 (see bars) and Supplementary file 5. MS-RA had the lowest p-value (P= 1.02E-67), followed by T1D-RA (P= 4.42E-43) and MS-T1D (P= 3.09E-33). The T1D-Alz pair, with the p-value of 6.48E-19, was more significant than the MS-Alz pair (P= 5.21E-16). Comparatively higher p-values were found among T1D-Park (P= 1.18E-04), Alz-Park (P= 8.72E-07), RA-Park (P= 3.95E-07) and MS-Park (P= 3.36E-08), exhibiting the distant relatedness for all possible combinations with Park (Figure 2). Altogether, the results showed the close relatedness among autoimmune disorders and within Alz pairs.

The degree of relatedness may be due to the real biological properties of the gwas-genes, i.e. the autoimmune gwas-genes participate in more biological pathways and processes than the neurodegenerative ones, and therefore the probability of abundant interactor sharing among autoimmune diseases is higher. However, we hypothesized that the observations could be partly biased by the difference in numbers of studies carried out on distinct gwas-genes, resulting consequently in higher or lower interactor information. To critically assess this aspect, we considered that the average number of interactions per gene in the VisANT database was equal to 7.26 (93684 interactions for

12888 human genes). Consequently, we derived a *normalization factor* by normalizing each disease interaction ratio for the one in the VisANT database. For example, the highest interaction ratio seen in RA resulted in a *normalization factor* equal to 1.88, signifying a nearly two fold increase in the interaction ratio compared to the database. Vice versa, Park normalization factor was 0.71, indicating that the interaction ratio was only 71% than the expected.

For paired analyses we multiplied the *normalization factors* relative to the two diseases, and used the *paired normalization factor* to optimize the observed/expected ratio of interactor sharing. This resulted in an *interactor score* for each disease pair (Figure 2). So, if on the one hand the p-values relative to the inter-disease relatedness may reflect the shared interactome among diseases, on the other hand the interactor score might flatten or enhance some of these observations as it corrects for annotation bias. Therefore, both the p-values and the interactor scores have to be considered while interpreting the results. The interactor scores confirmed the close association among disease pairs in the autoimmune group (Figure 2). Surprisingly, the highest scores appeared in Alz-Park and T1D-Alz pairs. Therefore, even after eliminating the possible bias introduced by database annotation, the T1D-Alz pair maintained high association.

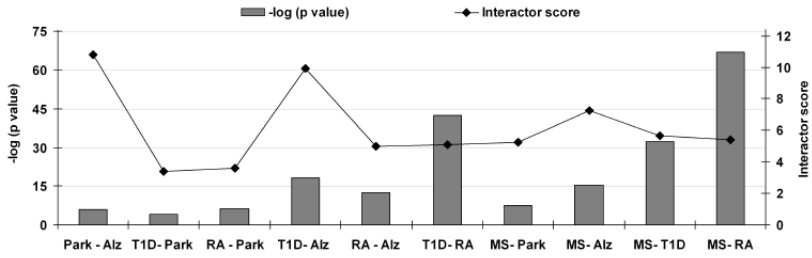


Figure 2. Molecular relatedness based on shared interactome. Statistical significance for each paired analysis is given.

4.1.3 The biological themes in autoimmune and/or neurodegenerative interactomes

In order to identify the biological themes related to each interactome, we used the ToppGene suite, an online tool for functional enrichment analysis. A Bonferroni corrected p-value of 0.05 was used to extract the significant biological pathways reported by three distinct databases (CGAP-BioCarta, KEGG and Panther). The highest number of pathways was reported in T1D followed by MS, RA, Alz and Park respectively. Further, we tabulated the shared pathways among diseases and grouped them according to the database in seven main categories: *Growth factor/Hormone signaling*, *Innate/Adaptive immunity*, *Cell cycle and apoptosis*, *Cancer*, *Adhesion*, *Host response* and *Other* (Table 3). The combined statistical significances of the pathway enrichments are also listed in Table 3. The majority of the shared pathways were categorized in *Growth factor/Hormone signaling,* followed by *Innate/Adaptive immunity* and *Cell cycle and apoptosis,* while categories like *Cancer, Host response* and *Other* appeared in single databases.

In the *Growth factor/Hormone signaling* category Panther indicated *EGF, FGF* and *PDGF signaling pathways* as shared by three diseases, Alz and T1D with MS or RA. The other pathways were exclusively shared by the autoimmune diseases, with the exception of *Angiogenesis* that appeared both in T1D and Alz. In contrast to Panther, KEGG evidenced predominant pathway sharing among autoimmune disorders, except for the *ErbB signaling* present in Alz and in the three autoimmune diseases. CGAP-BioCarta exposed high concordances among Alz, T1D and MS. There were 10 pathways exclusively shared by these three diseases, among which *EGF signaling* and *PDGF signaling* were already reported by the Panther database. Many of the pathways common to these trios were related to tyrosine kinase signaling, such as *Trka receptor signaling* and *Sprouty regulation of tyrosine kinases*. Other remarkable observation was the sharing of *NGF pathway* by MS and Alz. *IGF-1 signaling*, *IL3 signaling*, *Insulin signaling* and *Growth hormone signaling* pathways were shared exclusively by T1D and Alz. In addition, Park shared only three pathways with the autoimmune diseases, namely *Ceramide signaling, Trefoil factors initiate mucosal signaling* and *Phosphoinositides and their downstream targets.* Overall, pathways shared by the five diseases in the *Growth factor/Hormone signaling* category portrayed an undisputable association within the autoimmune group with the predominant presence of T1D. It also revealed that several biological themes related to tyrosine kinase signaling were shared among Alz, T1D

and MS. Most unexpectedly, the analysis exposed numerous growth factor related pathways common to T1D and Alz.

The second category *Innate and adaptive immunity* contained 20 pathways derived from the three pathway databases. Although expected for the autoimmune diseases, there was consistent sharing of immunity related pathways also for Alz. For example, pathways related to B and T cell activation appeared in all databases as shared by T1D, RA, MS and Alz. Furthermore, in the KEGG database Alz and T1D exclusively shared the *Fc epsilon RI signaling pathway* and *Natural killer cell mediated cytotoxicity* pathway, whereas CGAP-BioCarta emphasized predominant pathway sharing between MS and RA.

In the third category *Cell cycle and apoptosis* Panther and KEGG contributed with 2 pathways each, and Park shared the *Apoptosis signaling pathway* with RA and MS. Among the CGAP-BioCarta results, Alz shared 3 pathways with autoimmune diseases, while 4 pathways were exclusively shared by T1D-RA or RA-MS disease pairs.

The fourth category contained pathways derived from KEGG database related to *Cancer*. It resulted that many genes appearing in the autoimmune and /or neurodegenerative interactomes played a role in cancer related pathways.

In the fifth category (named *Adhesion*), the *Focal adhesion pathway* (KEGG) was common to Park and the autoimmune group and the *Integrin signaling pathway* (CGAP-BioCarta) was shared among neurodegenerative disorders and T1D. In addition,

Adherens junction pathway in KEGG was shared among Alz, T1D and MS.

The sixth category contained pathways related to *Host response*, which was reported exclusively by the KEGG database. Park and MS shared two pathways related to *E. coli* infection, and the autoimmune diseases shared the *Epithelial cell signaling in Helicobacter pylori infection* pathway. Lastly, the category *Other* contained the *Parkinson's pathway* reported by Panther, which was common to Park and RA.

Overall, pathway analysis revealed predominant sharing of functions among autoimmune diseases. Moreover, many of these pathways appeared also in Alz, which was associated with T1D in most cases.

Disaeses	Pathway	p-value*
Category 1 : Growth factor/Hormone signaling		
Alz, T1D, RA	EGF receptor signaling	6.6E-13
Alz, T1D, RA	FGF signaling	3.6E-10
Alz, T1D, MS	PDGF signaling	4.2E-14
T1D, MS	Interleukin signaling	1.0E-16
T1D, MS	JAK/STAT signaling	3.6E-06
T1D, RA	PI3 kinase	1.5E-06
Alz, T1D	Angiogenesis	1.1E-18
Alz,T1D,RA, MS	ErbB signaling pathway	3.1E-33
T1D, RA, MS	Adipocytokine signaling pathway	7.1E-15
T1D, RA	Cytokine-cytokine receptor interaction	2.4E-07
T1D, MS	Insulin signaling pathway	1.5E-06
T1D, MS	Jak-STAT signaling pathway	3.8E-25
RA, MS	MAPK signaling pathway	1.1E-06
Alz, T1D, MS	EGF Signaling	8.9E-17
Alz, T1D, MS	EPO Signaling	1.8E-13
Alz, T1D, MS	IL 6 signaling	4.1E-06
Alz, T1D, MS	PDGF Signaling	6.5E-12
Alz, T1D, MS	Signaling of Hepatocyte GF Receptor	4.3E-12
Alz, T1D, MS	Angiotensin II mediated activation of JNK via Pyk2 dependent signaling	5.7E-08
Alz, T1D, MS	Sprouty regulation of Trk signals	9.4E-09
Alz, T1D, MS	Links between Pyk2 and Map Kinases	1.4E-13
Alz, T1D, MS	Trka Receptor Signaling	1.7E-09
Alz, T1D, MS	Bioactive Peptide Induced Signaling	1.1E-07

Park, RA, MS	Ceramide Signaling	1.1E-12
Park, T1D, MS	Trefoil Factors Initiate Mucosal Healing	3.6E-10
T1D, RA, MS	Chaperones modulate interferon Signaling	1.5E-07
T1D, RA, MS	Keratinocyte Differentiation	3.2E-15
Alz, MS	Nerve growth factor (NGF)	2.7E-05
Alz, T1D	IGF-1 Signaling	3.5E-05
Alz, T1D	IL 3 signaling	1.4E-06
Alz, T1D	Insulin Signaling	1.1E-04
Alz, T1D	Growth Hormone Signaling	9.1E-12
T1D, MS	IL 2 signaling	1.5E-12
T1D, MS	IL 4 signaling	4.4E-10
T1D, MS	IL-7 Signal Transduction	3.8E-09
T1D, MS	CBL mediated ligand-induced downregulation of EGF receptors	5.6E-07
T1D, MS	CXCR4 Signaling	3.3E-08
T1D, MS	TPO Signaling	4.3E-10
T1D, MS	Ca Signaling by HBx of Hepatitis B virus	5.2E-05
RA, MS	MAPKinase Signaling	8.0E-11
RA, MS	Signal transduction through IL1R	1.3E-05
RA, MS	SODD/TNFR1 Signaling	1.3E-11
RA, MS	TNF/Stress Related Signaling	2.8E-21
RA, MS	TNFR1 Signaling	2.4E-12
RA, MS	TNFR2 Signaling	5.5E-19
Park, T1D	Phosphoinositides and their downstream targets.	7.1E-04
Category 2 : Innate and adaptive immunity		
Alz,RA,T1D,MS	B cell activation	1.0E-09
Alz,RA,T1D,MS	T cell activation	1.1E-31
Alz,T1D,RA,MS	B cell receptor signaling pathway	2.5E-25
Alz,T1D,RA,MS	T cell receptor signaling pathway	3.3E-38
T1D, RA, MS	Toll-like receptor signaling pathway	7.2E-13
Category 2 : Innate and adaptive immunity (continued)		
Alz, T1D	Natural killer cell mediated cytotoxicity	7.0E-16
Alz, T1D	Fc epsilon RI signaling pathway	2.6E-09
RA, MS	Antigen processing and presentation	1.7E-13
Alz,T1D,RA,MS	T Cell Receptor Signaling	7.9E-15
Alz, T1D, MS	IL-2R Beta Chain in T cell Activation	4.3E-24
Alz, T1D, RA	BCR Signaling	1.5E-07
RA, MS	CD40L Signaling	1.6E-13
RA, MS	The 4-1BB-dependent immune response	2.6E-10
RA, MS	TACI and BCMA stimulation of B cell immune responses.	3.9E-16
RA, MS	NFkB activation by Nontypeable *H.influenzae*	5.9E-07
RA, MS	NF-kB Signaling	1.4E-15
RA, MS	Acetylation-Deacetylation of RelA in Nucleus	8.4E-13
Alz, MS	Fc Epsilon Receptor I Signaling in Mast Cells	3.3E-05
T1D, RA	Lck and Fyn tyrosine kinases in initiation of TCR Activation	1.9E-07
T1D, RA	Co-Stimulatory Signal During Tcell Activation	6.3E-10
Category 3 :Cell cycle and apoptosis		
Park, RA, MS	Apoptosis signaling	1.0E-34
T1D, RA	p53	1.2E-05
RA, MS	Apoptosis	3.1E-23

T1D, RA	Cell cycle	7.6E-23
Alz,T1D,RA,MS	PTEN dependent cell cycle arrest and apoptosis	9.7E-15
Alz, T1D, MS	Inhibition of Cellular Proliferation by Gleevec	2.4E-11
T1D, RA, MS	Telomeres,Telomerase, Aging, Immortality	1.2E-12
T1D, RA	Cyclins and Cell Cycle Regulation	1.0E-16
T1D, RA	Influence of Ras and Rho proteins on G1 to S Transition	6.5E-14
T1D, RA	Cell Cycle: G1/S Check Point	8.1E-11
T1D, RA	p53 Signaling	4.5E-07
Alz, T1D	Multiple antiapoptotics from IGF-1R signaling lead to BAD phosphorylation	8.5E-06
RA, MS	HIV-I Nef: negative effector of Fas and TNF	2.0E-29
RA, MS	Caspase Cascade in Apoptosis	1.1E-03
RA, MS	FAS signaling (CD95)	2.3E-08
RA, MS	Induction of apoptosis through DR3 and DR4/5	3.2E-18
Category 4 :Cancer		
Park, Alz, T1D, MS	Colorectal cancer	7.1E-10
Alz, T1D, RA, MS	Prostate cancer	3.1E-19
Alz, T1D, RA, MS	Pancreatic cancer	4.4E-24
Alz, T1D, RA, MS	Pathways in cancer	1.1E-46
Alz, T1D, RA, MS	Chronic myeloid leukemia	8.6E-48
T1D, RA, MS	Acute myeloid leukemia	7.7E-13
T1D, RA, MS	Small cell lung cancer	4.8E-26
Alz, T1D, RA	Glioma	1.0E-18
Alz, T1D, RA	Non-small cell lung cancer	5.9E-17
Category 5 :Adhesion		
Park, T1D, RA, MS	Focal adhesion	8.8E-14
Alz, T1D, MS	Adherens junction	1.9E-08
Park, Alz, T1D, MS	Integrin Signaling	6.7E-20
Category 6 :Host response		
Alz, RA, MS	Epithelial cell signaling in *H. pylori* infection	8.2E-11
Park, MS	Pathogenic *E. coli* infection - EHEC	6.3E-09
Park, MS	Pathogenic *E. coli* infection - EPEC	6.3E-09
Category 7 :Other		
Park, RA	Parkinson's disease	1.0E-13

Panther KEGG CGAP-BioCarta

* Combined p-value by Fisher's method.

Table 3: Shared pathways among interactomes

4.1.4 Quantitative analysis of functional relatedness among neurodegenerative and autoimmune interactomes

We tabulated the shared pathways for all disease pairs and quantified the significance of the observed pathway overlaps by performing hypergeometric test for each database separately

(Figure 3a). According to Panther, the most significant pathway overlap was in T1D-Alz pair (P= 2.10E-07), followed by the T1D-RA (P= 1.48E-05) and RA-Alz (P= 8.69E-05). MS-Park pair displayed no significant pathway overlap (P= 0.184). Among the KEGG results, the most highly significant pathway overlap was between MS and RA (P= 1.03E-13), followed by T1D-RA (P= 51.22E-12), T1D-Alz (P= 5.91E-11), MS-T1D (P= 8.58E-10) etc. Yet, the analysis revealed that pathway overlaps with Park could be due to chance, except for the MS-Park pair (P= 5.44E-03). Finally, the CGAP-BioCarta analysis recognized T1D-Alz as the most significant association (P= 3.71E-12), followed by MS-RA (P= 1.60E-11), MS-T1D (P= 2.16E-09), MS-Alz (P= 1.60E-08) etc. To summarize, the p-values based on KEGG and CGAP-BioCarta databases were more significant among the disease pairs in the autoimmune group, especially MS-RA. Moreover, through the consistent statistical results with respect to all the three databases, T1D-Alz pair could be considered as the most significant disease pair in the context of shared pathways.

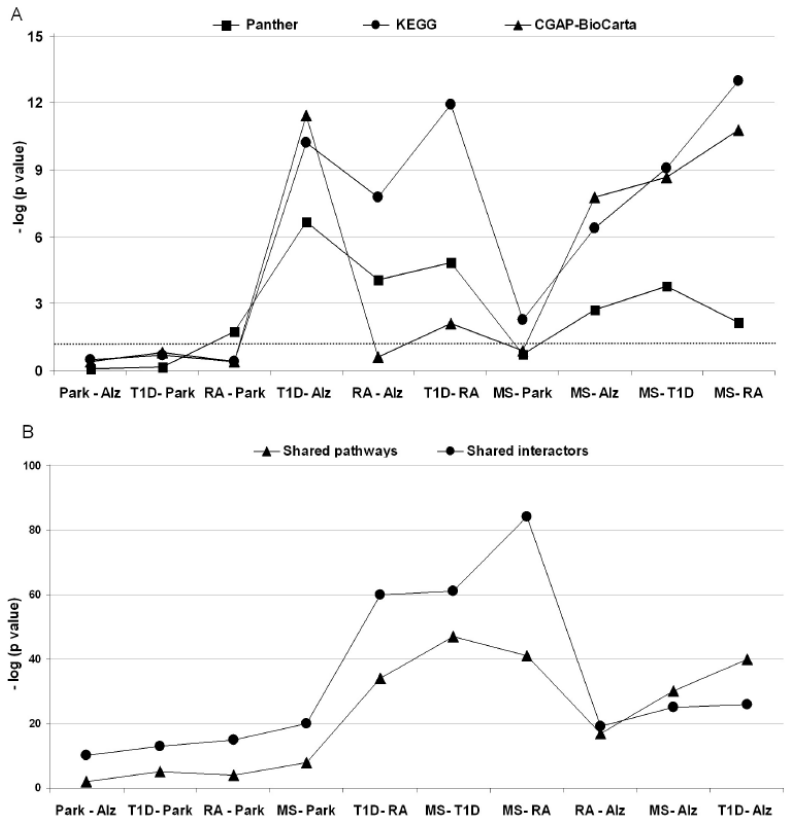


Figure 3. A. Functional relatedness based on shared pathways. Statistical significance for each paired analysis is given as relative to the database, Dotted line indicates the p-value threshold of 0.05. B. Shared pathways and interactors among disease pairs.

Theoretically, pathway sharing was partly anticipated as consistent interactor sharing was present in several disease pairs. Clearly, the more interactors were shared, the more common pathways were expected. We critically evaluated this issue and measured the Pearson's correlation between the interactors and the pathways (as reported by the three databases together) shared among the disease pairs. Consistently, we observed that the number of shared pathways was proportional to the number of

shared interactors in most cases (Figure 3B). However, the Alz pairs with autoimmune diseases were not following this trend. The overall Pearson's correlation coefficient was 0.79, which increased remarkably to 0.95 when Alz-autoimmune disease pairs were excluded. These observations demonstrate that the number of shared biological pathways among Alz and autoimmune diseases is higher than the expected on the basis of interactor sharing.

Finally, we ranked the associations among diseases based on the statistical significances of shared pathways in each database. Scores starting from 1 were assigned from the highest significant association to the lowest. Those with insignificant p-values were assigned the rank 10. Then, we derived the cumulative ranks for each disease pair by summing their ranks in the three databases. As expected, T1D-Alz pair had the highest rank, followed by the autoimmune disease pairs MS-RA and T1D-RA, and then by MS-Alz and RA-Alz. Park pairs scored all very low. These results were depicted in Figure 4 displaying a network summary of relationships among the five diseases.

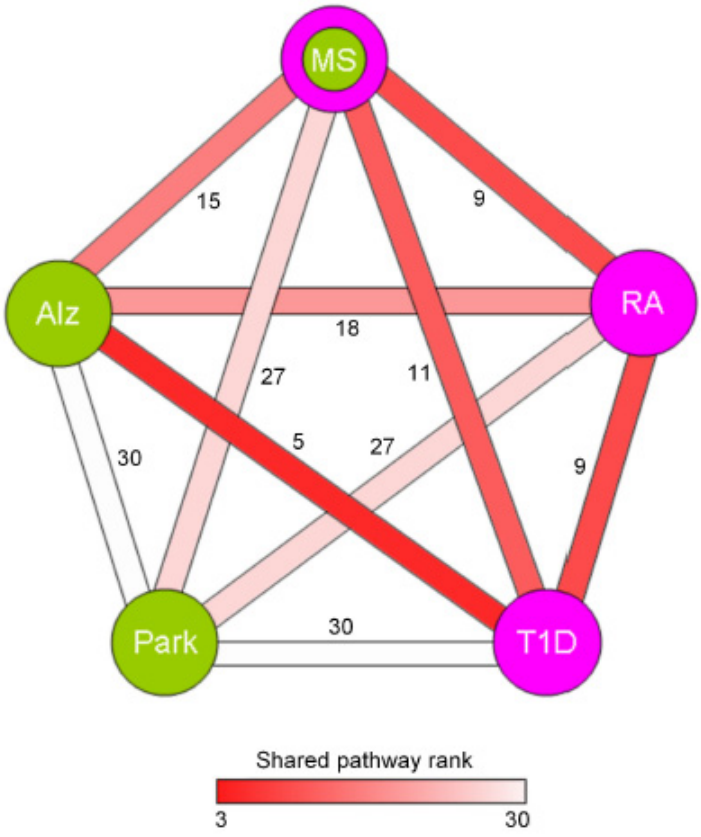


Figure 4. Overall disease relatedness based on shared pathways in the Panther, KEGG and CGAP-BioCarta databases. Green nodes indicate neurodegenerative disorders, whereas pink nodes highlight autoimmune diseases. The color of the edges connecting the nodes reflects the shared pathway rank ranging from 3 (highest relatedness) to 30 (lowest relatedness).

Altogether, the relationships among diseases identified at the level of molecular interactions were found to be sustained at the level of biological pathways, especially among autoimmune diseases. Surprisingly, the same approaches evidenced strong relatedness between Type 1 diabetes and Alzheimer's disease.

4.2 GENDER-BASED TRANSCRIPTOMICS IN MS

4.2.1 A comparison of conventional and gender-based transcriptome analysis

Conventional analysis strategy in transcriptome studies foresees that differentially expressed genes (DEG) which are significant in the disease state are derived from the comparison between healthy and diseased populations. We applied this approach in our study which included initially 22 healthy individuals and 23 patients with relapsing-remitting MS. All the patients were clinically stable and none of them had started any immunomodulatory therapy till sampling. Peripheral blood drawings were performed at least 4 weeks after relapse treatment with steroids, in order to exclude the variations in gene expression due to drugs. We labeled RNA extracted from PBMC and hybridized Illumina microarray chips that contained about 22,000 probes. After normalization and data filtering, the sample set was analyzed by PCA and hierarchical sample clustering methods. Three samples (1 healthy control and 2 RR-MS patients) were identified by both methods as outliers and removed from further analysis. For the remaining 21 diseased subjects there were no significant differences in age, disease duration, EDSS score and annual relapse rate between the two gender groups (see materials and methods). Control subjects were sex and age-matched.

The most widely used strategy of deriving genesets is through single statistical tests and/or fold change criteria. However, in a preliminary analysis it was observed for several genes that the mean expression values were considerably determined by far

extreme expression values in few samples within the group. This could result in fold changes that did not reflect reproducible expression in most of the samples within each group. Therefore, to facilitate higher reproducibility and reliability of results we used three statistical tests in parallel to find differential expressions and selected the genes which passed through at least two tests. This way we controlled the potential false positives without losing vital information in the transcriptomics data. The statistical tests were Welch t-test, Significance Analysis of Microarrays (SAM) and Bayesian analysis of differential gene expression (BADGE). The BADGE analysis also performed Leave One out Cross Validation. The DEG identified by these methods were 162, 80 and 289 for t-test, SAM and BADGE respectively, with 167 genes common to any two statistical tests (Figure 1A). Hierarchical clustering of this geneset, represented in Figure 1B, showed 65 downregulated and 102 upregulated genes in the relapsing-remitting course of the disease.

However, while observing the heatmap generated with the conventional (hereon called “couple”) case-control comparison, we noticed a considerable amount of intra-group variability which was independent from the health status, as it appeared both in the MS and CTRL groups. We hypothesized that this could be due to the inherent heterogeneity in the global gene expression between women and men. To verify this aspect, we measured the coefficient of variation (CV) for sets of probes that were grossly selected (median fold change threshold of 1.3 between healthy and diseased) in the couple, male and female case-control comparisons,

and then calculated the average CV in each sample group. Here, we used the median fold differences because median values are least affected by the presence of few extreme expression values. As shown in Figure 1C, high heterogeneity was found in the diseased and control groups when genelists were generated with the couple approach. On the contrary, gender-based case-control comparisons led to gene lists which exhibited a significant reduction in the variability.

Finally, we assessed whether MS pathology was associated with alterations in the expression of sex-specific genes. We defined the naturally occurring sex-specific genes as those genes that were normally differentially expressed in men versus women in the healthy population. Among the 10,390 filtered probes, 1083 (10.42%) passed the relaxed t-test with p-value threshold of 0.05 and were differentially regulated between healthy women and healthy men (Figure 1D left circle). Interestingly, when repeating the same analysis on the diseased population (Figure 1D right circle), only 206 probes (1.98%) were sex-specific (p-value 2.3E-140). This significant reduction in the percentage of the sex-specific genes was accompanied also by qualitative changes, as only 32 out of the 206 MS sex-specific probes appeared among the natural sex-specific probes, indicating that in diseased subjects almost all natural sex-specific genes were not differentially expressed between women and men any more, while a new (smaller) subset of genes performed as sex-specific.

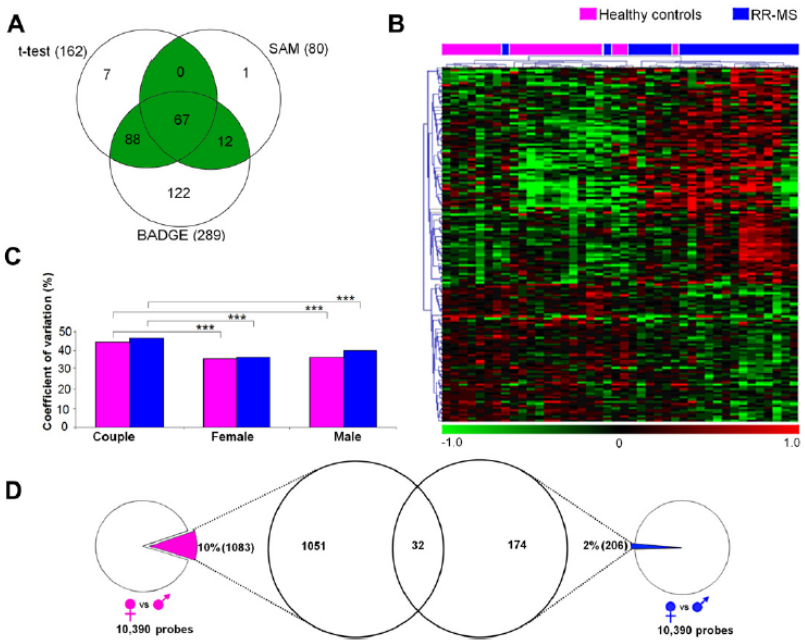


Figure 1. The heterogeneity in human populations displayed by conventional transcriptome output is reduced in gender-based genelists. A. Number of differentially expressed genes detected by three statistical tests. The MS signature is defined by the DEG common to at least two tests (green sections). B. Heatmap showing unsupervised clustering of DEG (in rows) and of the samples (in columns). Bar above the heatmap indicates sample classification (healthy/pink vs. MS/blue subjects). Bar below the heatmap indicates expression intensity. C. Coefficients of variation (%) of grossly selected probes in couple and gender-based groups. *** p value < 0.001. D. Left and right pie charts indicate the fraction of natural (pink) or MS (blue) sex-specific probes among the 10,390 filtered genes. The Venn diagram in the middle highlights the number of sex-specific probes overlapping among healthy and diseased populations.

4.2.2 The gender-based MS signatures

Based on the previous observations, we switched to a gender-based strategy as opposed to the couple approach and compared MS women or men with their respective healthy counterpart. The female population comprised of 9 controls and 12 RR-MS, while the male group consisted of 12 controls and 9 RR-MS samples. The statistical approaches and parameters for identifying DEG were the same as in the couple analysis. In the female sample group, the t-test, SAM and BADGE identified 197, 139 and 301 genes respectively, with 208 probes (207 genes) common to any two of them (Figure 2A and supplementary file WOMEN, sheet 1). Similarly, in the male group t-test identified 73, SAM 19 and BADGE 126 genes (Figure 2D). There were 72 genes common in any of these two tests (supplementary file MEN, sheet 1).

Hierarchical clustering of 208 MS female probes (207 genes, as SNRPN gene was represented by two probes in all the tests) showed 131 downregulated and 77 upregulated probes in the disease (Figure 2B). Unsupervised sample clustering showed that such signature was extremely efficient in classifying the healthy and diseased subjects, as 18/21 samples were properly assigned (Figure 2B pink and blue bar). The relative gene expressions in the male sampleset were also checked and the cluster analysis is shown in the Figure 2C. It is apparent that the MS female DEG had poor performance in male sampleset, as it resulted in a substantial rise in misclassifications in the unsupervised classification. The MS male geneset was characterized by 72 genes with 47 downregulated and 25 upregulated genes in the disease.

Hierarchical clustering of these genes is shown in the Figure 2E**.** Similar to the MS female DEG, the male DEG classified almost perfectly the male (20/21 correct classifications, Figure 2E) but not the female (Figure 2F) specimens. Overall, when compared to the couple heatmap, the gender-based heatmaps gained substantial clarity in differentiating healthy and diseased groups.

We compared the MS couple, female and male genesets to make out the commonalities and differences between them. Forty-three genes out of 207 genes in the female and 16 out of 72 genes in the male genesets were present in the couple genelist. Most unexpectedly, there was only one gene (LOC196752) common between the male and female genesets. This gene, located at 10q22.2 and coding for a protein with unknown function, appeared also in the couple geneset. Further, to verify the reliability of the female and male DEG, we added two new groups of RR-MS patients comprising of 10 women and 8 men to the existing sampleset and the corresponding log ratios were calculated without adjusting the average with the expression values of newly recruited samples. Consistent with the previous results (Figure 2), the global unsupervised cluster analysis (hierarchical clustering) classified correctly 29 out of 31 female and 27 out of 29 male samples. In addition, differential expression was evident also by real-time PCR for selected DEG from male and female groups.

In summary, gender-based analyses allowed to limit the natural heterogeneity existing within human populations and unraveled

distinct changes in blood gene expression in MS men or women although associated to the same disease form.

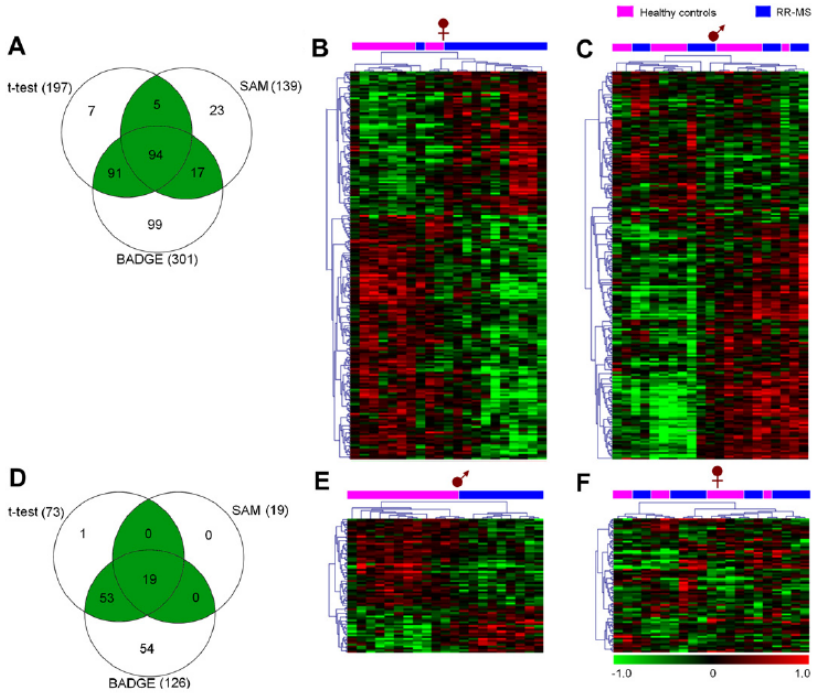


Figure 2. Gender-based analyses in multiple sclerosis improve sample classification. A. Number of differentially expressed genes detected by three statistical tests between healthy and RR-MS women. Green sections contain genes common to at least two tests. B-C. Heatmaps showing unsupervised clustering of female DEG in female (B) or male (C) samples. D. Number of differentially expressed genes detected by three statistical tests between healthy and RR-MS men. Green sections contain genes common at least two tests. E-F. Heatmaps showing unsupervised clustering of male DEG in male (E) and female (F) samples. Bars above the heatmaps indicate sample classification (healthy/pink vs. MS/blue subjects). Bar below the heatmap indicates expression intensity.

4.2.3 Biological themes of the MS signatures

In order to explore the biological information contained in the gender-related MS genesets, we applied the Genecodis program which uses the Gene Ontology (GO) database for annotations and identifies the corresponding GO-ID for the significantly enriched terms.

Out of the 30 enriched biological process categories in the MS female geneset, the top scores were *interspecies interaction between organisms*, *transcription* and r*egulation of transcription-DNA dependent*. In MS male geneset, due to the lower number of DEG, only 8 biological process categories were significant and *oxidation reduction- fatty acid metabolism, transcription* and *regulation of transcription- DNA dependent* were the most significantly enriched.

Most surprisingly, there was strong concordance in biological processes between the female and male geneset. The statistical significance and the number of genes for each common item are shown in Table 1. The commonalities were salient with the presence of 5 biological processes among the 8 enriched terms in men. The common GO terms were *regulation of transcription-DNA dependent*, co-occurrences of *regulation of transcription-DNA dependent* and *transcription*, *modification-dependent-protein catabolic process*, *transcription* and *chromatin modification*. As the genesets were derived from independent datasets through independent analyses, the combined p-values were calculated (Table 1). They ranged between 4.6E-06 and 5.6E-03, ruling out the possibility that these findings were due to chance. Therefore,

even if the DEG were different in MS women and men, many of them played a role in the same biological processes. Furthermore, 4 out of 5 highly enriched biological process categories referred to transcription and chromatin modification, indicating that epigenetic events may be the common basis for MS disease in women and men.

Next, we reasoned that if the DEG were involved in the same processes, then they could exert similar functions. We tested this hypothesis by searching for recurrent molecular functions in the genelists. Among the 41 enriched molecular functional categories in female geneset, the top scores were *protein binding*, *hydrolase activity*, and *transferase activity*. In the male genelist, only 10 significant functions were enriched, however, similarly to the concordance seen in biological processes and consistent with our hypothesis, 6 out of 10 enriched categories in men were present in women as well. Table 2 illustrates them together with the corrected and the combined p-values. The shared molecular functions were *protein binding*, co-occurrences of *protein binding* and *RNA binding*, *DNA binding*, *metal ion binding*, *zinc ion binding* and *RNA binding*. Furthermore, most of the common molecular functions dealt with interactions with nucleic acids, once more affirming the prominence of epigenetic mechanisms in multiple sclerosis.

Gene Ontology : Biological process	Gender	No. of Genes	Corrected p-value	Combined p-value
GO:0006350 : transcription	F	23	0.00012	4.6E-06
	M	10	0.00224	
GO:0006355 : regulation of transcription, DNA-dependent & GO:0006350 : transcription	F	20	0.0003	1.8E-05
	M	8	0.00423	
GO:0006355 : regulation of transcription, DNA-dependent	F	23	0.00104	7.4E-05
	M	10	0.00543	
GO:0016568 : chromatin modification	F	4	0.01767	9.0E-04
	M	3	0.00482	
GO:0019941 : modification-dependent protein catabolic process	F	6	0.02187	5.6E-03
	M	3	0.03103	

F= Female, M= Male

TABLE I. Biological processes (GO terms) shared by gender-based MS DEG.

Gene Ontology : Molecular Function	Gender	No. of Genes	Corrected p-value	Combined p-value
GO:0005515 : protein binding	F	73	2.142E-13	1.7E-14
	M	21	0.0022441	
GO:0003677 : DNA binding	F	16	0.0056011	3.0E-04
	M	8	0.0053933	
GO:0046872 : metal ion binding	F	20	0.0464598	9.0E-04
	M	14	0.0019222	
GO:0003723 : RNA binding & GO:0005515 : protein binding	F	5	0.0206388	9.0E-04
	M	4	0.0039642	
GO:0008270 : zinc ion binding	F	13	0.0456205	1.0E-03
	M	9	0.0021955	
GO:0003723 : RNA binding	F	9	0.0486291	3.1E-03
	M	5	0.0069794	

F= Female, M= Male

TABLE 2. Molecular functions (GO terms) shared by gender-based MS DEG.

These high concordances in biological processes and molecular functions might indicate that the diverse genesets share some interacting partners. We tested this hypothesis by a systems biology approach using the VisANT program, a web-enabled tool for data-mining, visualizing, analyzing and modeling biological networks from user given input of genes or proteins. VisANT extrapolates the interacting partners for each gene by querying databases such as Biogrid, MIPS, BIND and HPRD, represents then each gene as a node and connects the interacting nodes by a straight line. We found that there were 1486 and 453 interactors for female and male genesets respectively. Notably, 171 out of 453 male interactors (about 38%) were common to female interactor

set. In order to get further insights into disease related mechanisms, we checked the molecular functions and biological processes enriched in the common 171 interactors. Among the 134 significant biological processes, *interspecies interaction between organisms*, *positive regulation of I-kappaB kinase/NF-kappaB cascade* and *positive regulation of transcription from RNA polymerase II promoter* were the most significant, while *regulation of transcription-DNA dependent* was enriched with 33 genes and *transcription* with 29 genes, and were the GO terms defined by the highest numbers of interactors. Furthermore, 49 categories contained ontology terms related to transcription and 3 to chromatin modification. Similarly, 130 categories defined molecular functions for the shared interactors, with *protein binding*, *transcription factor activity* and *transcription factor binding* as the most significantly enriched terms. Among the 130 categories, 40 groups were related to transcription and 29 to DNA binding. As highlighted in Figure 3, a remarkable part of the DEG sharing the interactome (72/89 female and 27/35 male DEG, green nodes in upper and lower left circles) were involved in epigenetic processes and contacted 62/171 common interactors related to transcription and chromatin modification (green nodes in right circles).).

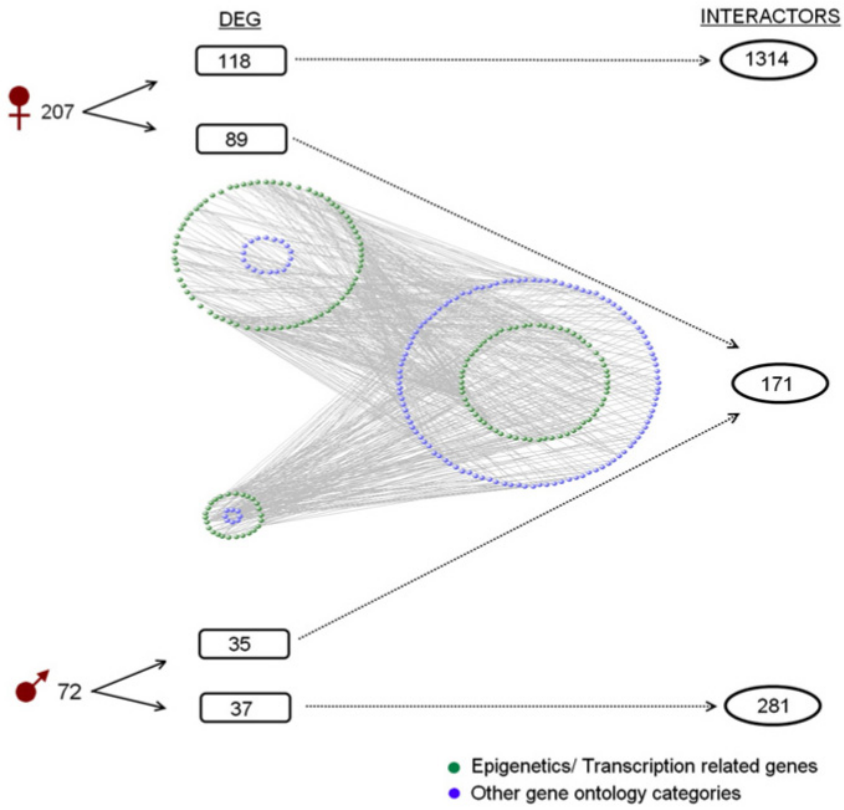


Figure 3. Gender related MS interactomes identify shared interactors among male and female MS DEG. Female and male DEG have unique and shared interacting genes. Network reconstruction identified 171 shared interactors contacted by 89/207 female and 35/72 male DEG. Green nodes highlight epigenetic and transcription related genes among the DEG and the interactors.

Further, additional epigenetic factors appear also among the female or male specific interactors, such as DNMT1 in the female specific interactome and HGMA2 in the male specific interactome. Further, many of the 62 epigenetic interactors (green nodes in Figure 4A) were connected with more than one female and male DEG (pink and blue nodes respectively). The most connected interactor was

HNF4A (with 36 female and 12 male DEG), followed by TP53 (11 female and 2 male DEG), IKBKG (7 female and 2 male DEG) and RELA (5 female and 4 male DEG). Finally, we verified a further network level by reconstructing the interactions among the 62 epigenetic interactors. Interestingly, with two exceptions (KLF5 and TSC22D4), all the other interactors were found to be involved in a complex network, where some nodes, such as histone deacetylase (HDAC)1, HDAC2, RELA, TP53, SP1 and AR, were highly interactive (Figure 4B, Supplementary file INTERACTORS, sheet5). Interestingly, most of the epigenetic factors fell in three main groups related to the HDAC, AR or NF-kappaB complexes. These groups could interact each other either directly or through a central group of transcription factors.

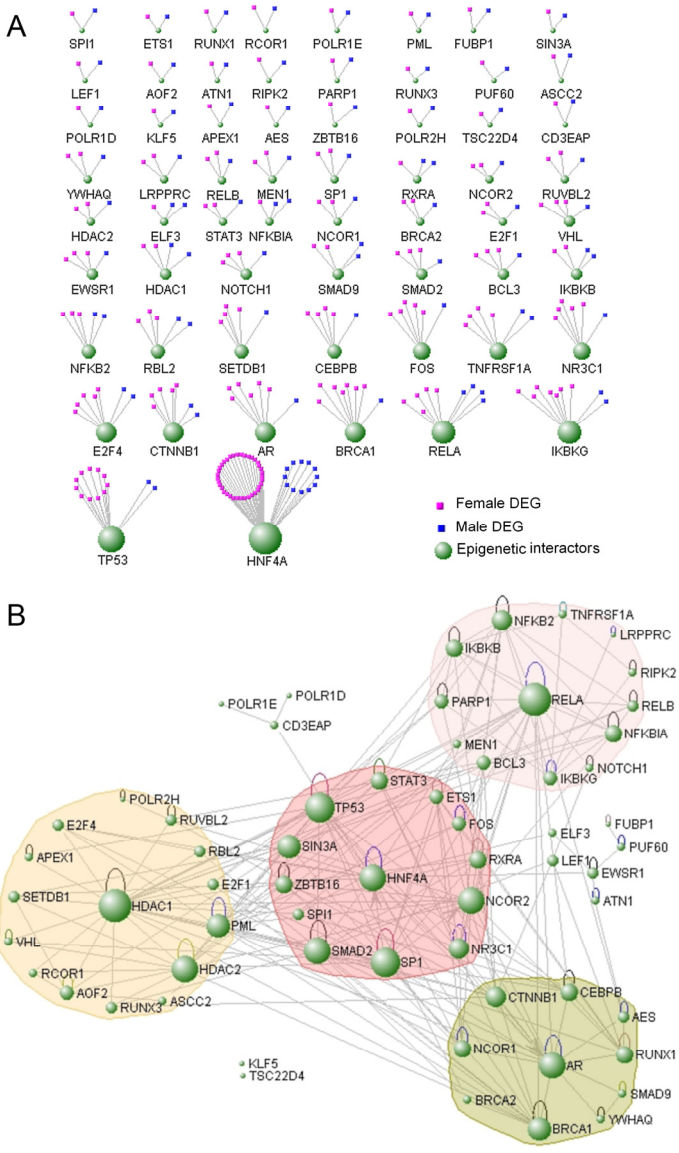


FIGURE 4. Epigenetic interactors shared by female and male MS DEG. A. Shared epigenetic interactors (green) of female (pink) and male (blue) MS DEG. B. Interaction network among the shared epigenetic interactors. Most of the epigenetic factors fall in three main groups related to the HDAC, AR or NF-kappaB complexes. These groups can interact each other either directly or through a central group of transcription factors. In A and B the sizes of the interactor nodes are proportional to the number of interactions.

Overall, the distinct MS female and male genesets shared biological and molecular functions as well as interactors. Themes related to epigenetics were predominant in both female and male MS signatures and in the shared interactome, suggesting they might give shape to and/or sustain pathogenic processes in multiple sclerosis.

4.2.4 Identification of the transcription factor SP1 as a modulator of experimental autoimmune encephalomyelitis

Finally, we checked whether the two MS genesets were potentially regulated by common transcription factors by querying the GeneCodis database. As shown in Table 3, both female and male MS signatures were significantly enriched for genes containing binding sites for four transcription factors (SP1, LEF1, NFY and ELK1) and their combinations. In particular, several DEG contained binding sites for SP1 alone or together with NFY and ELK1 (Table 3). We applied the same analysis tools to two published signatures for the RR-MS population and found that the binding sites for these transcription factors were enriched also among the genes described in other studies (Table 3). As SP1 appeared among the epigenetic interactors connected with both female and male DEG (Figure 4A) and belonged to the central core of highly connected epigenetic factors (Figure 4B), we

hypothesized that SP1-dependent transcription could play a role in MS pathogenesis.

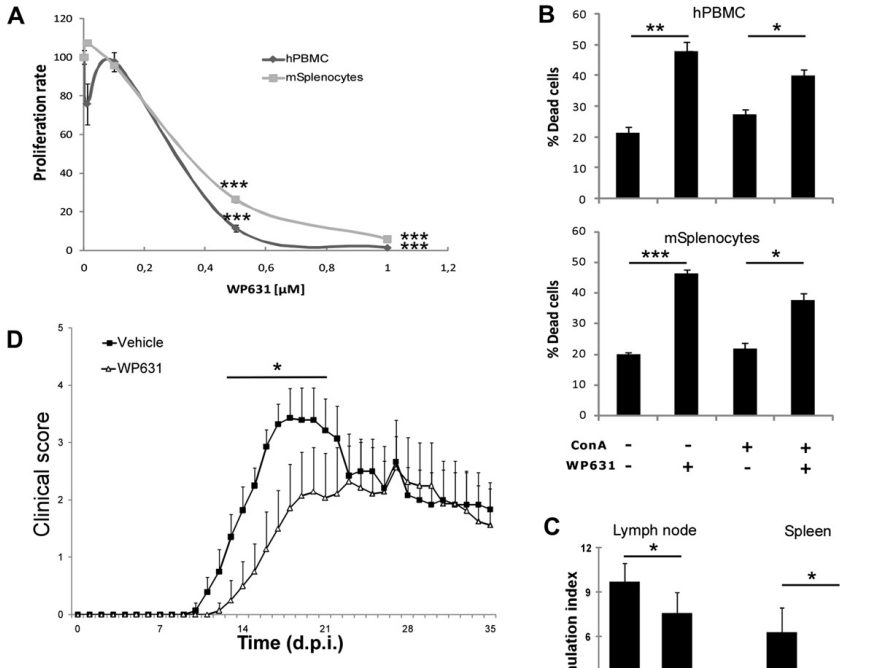
To validate the involvement of SP1 in immune responses, we initially targeted SP1-dependent gene transcription in peripheral blood mononuclear cells with the specific inhibitor WP631 and induced T cell proliferation with the mitogen concanavalin A (ConA). Interestingly, WP631 concentrations above 0.5μM strongly blocked proliferation of human T cells (Figure 5A). Similar results were obtained with ConA-activated mouse splenocytes (Figure 5A). As WP631 belongs to the family of anthracylines, cytotoxic agents used in cancer therapy, we checked WP631-dependent cytotoxicity in PBMC and splenocyte cultures by flow cytometry. As depicted in figure 5B, the drug increased the levels of cell death both in unstimulated and in ConA-stimulated cultures, indicating that part of the reduction in the proliferation was due to cytotoxicity. Finally, WP631 was administered *in vivo* to C57BL6 mice immunized with the encephalitogenic MOG_{35-55} peptide, and T cell responses to the autoantigen were tested ex vivo. Mice receiving the drug (3mg/kg/day) or its vehicle from day 3 to day 7 post-immunization were sacrificed at day 10 and lymphoid organs were extracted. Interestingly, T cell proliferation to MOG peptide was significantly reduced in WP631-treated animals (Figure 5C). Moreover, clinical expression of experimental autoimmune encephalomyelitis (EAE), the animal model of MS, was significantly altered by WP631 treatment. In fact, compared to the control group, in the WP631-treated group the incidence of disease was reduced, the onset was

delayed and the disease severity in the acute phase was milder (Figure 5D-E). Similar results were obtained in the EAE model in SJL mice (Figure 5F).

Unique TF	Multiple TF (if any)	Female p-val[a]	Male p-val[a]	Combined p-val[b]	*Gandhi et al.* p-val[a]	*Riveros et al.* p-val[a]
V$SP1_Q6	V$SP1_Q6	1.9E-05	1.7E-02	5.1E-06	4.7E-05	1.0E-06
	V$SP1_Q6 & V$NFY_Q6_01	2.7E-02	4.6E-02	9.6E-03	3.0E-02	*n.d.*
	V$SP1_Q6 & V$NFKB_Q6	1.5E-02	4.1E-03	6.7E-04	*n.d.*	*n.d.*
	V$SP1_Q6 & V$ELK1_02	3.4E-03	3.3E-03	1.4E-04	*n.d.*	*n.d.*
V$LEF1_Q2	V$LEF1_Q2	8.3E-03	6.9E-03	6.1E-04	*n.d.*	2.7E-02
	V$LEF1_Q2 & V$NFY_Q6_01	3.4E-02	7.6E-03	2.4E-03	*n.d.*	*n.d.*
	V$LEF1_Q2 & V$AR_Q6	6.3E-03	3.7E-03	2.8E-04	3.1E-02	2.8E-02
	V$LEF1_Q2 & V$TATA_01	3.1E-02	5.0E-02	1.1E-02	*n.d.*	*n.d.*
V$NFY_Q6_01	V$NFY_Q6_01	2.1E-02	5.4E-03	1.2E-03	4.8E-02	1.7E-03
V$ELK1_02	V$ELK1_02	2.9E-08	6.4E-03	4.2E-09	6.9E-16	4.9E-04

[a] Corrected p-values by Benjamini-Hochberg's method. [b] Combined p-values using Fisher's method.

TABLE 3. Transcription factors regulating MS DEG.



E

MOG_{35-55} EAE	Vehicle	WP631
Number of animals	14	14
Incidence (%)	100	71
Day of disease onset	12.1 ± 0.7	15.1 ± 1***
Maximal clinical score	3.7 ± 0.4	2.4 ± 0.8*
Score at peak	3.4 ± 0.4	2.1 ± 0.7*
Cumulative disease score (d0-26)	38.1 ± 4.4	23.2 ± 8.1**

F

$PLP_{139-151}$ EAE	Vehicle	WP631 1.5 mg/kg	WP631 3 mg/kg
Number of animals	10	10	10
Incidence (%)	100	100	80
Mortality (%)	30	0	0
Day of disease onset	11.1 ± 0.5	12.9 ± 0.4*	13.3 ± 1
Maximal clinical score	3.8 ± 0.3	2.9 ± 0.3*	2.4 ± 0.5*
Cumulative disease score (d0-37)	53.9 ± 7.8	26.4 ± 5.2**	30.4 ± 7.1*

Figure 5. Inhibition of SP1 dependent transcription modulates EAE. A. WP631 blocks ConA-induced T cell proliferation. Data are given as percentage of proliferation in absence of WP631 (100%). hPBMC= human PBMC, mSplenocytes= mouse splenocytes. Data are shown as mean ± SEM of 5 independent experiments. B. WP631 induces immune cell death. WP631 concentration was 1 µM. Data are shown as mean ± SEM of at least 4 independent experiments. C. WP631 administration in MOG_{35-55} immunized animals reduces T cell responses to the encephalitogenic peptide. Proliferation to MOG_{35-55} peptide of draining lymph node (left) or spleen (right) cells from WP631 or

vehicle-treated EAE mice (n=5 per group). Data are given as average ± SEM. Similar observations were obtained in a second experiment. D-E. WP631 administration *in vivo* attenuates MOG_{35-55} (D-E) and $PLP_{139-151}$ (F) induced EAE. Clinical score chart (D) and clinical parameters (E) are given for MOG_{35-55} induced EAE (n=14 mice per group). WP631 was administered at a dose of 3 mg/kg/day. Mean values ± SEM of 2 pooled EAE experiments are shown. F. Clinical parameters relative to $PLP_{139-151}$ induced EAE in mice treated with two distinct doses of WP631.* p value < 0.05, ** p value < 0.01, *** p value < 0.001.

5 CONCLUSIONS AND FUTURE PROSPECTS

The aim of genome-wide association studies (gwas) is to discover the common genetic variants associated with susceptibility to complex diseases. However, the analysis of isolated genes is little informative about the biological processes underlying disease and offers limited rationale for the development of novel therapies. Theoretically, reconstruction of the molecular interaction networks linked to gwas-genes by systems biology approaches may help to elucidate the functional consequences related to each susceptibility allele and the combined effects of more genetic variants. The few reports available till date on this issue demonstrated that this approach can identify previously unseen relationships among human diseases at molecular level [147, 148]. The reconstruction of the molecular framework hosting the genetic variants associated with susceptibility to a complex human disorder has the potential to reveal the biological mechanisms underlying that disease and to highlight similarities with other diseases. Here we made a cross-disease comparison on autoimmune (T1D, RA) and neurodegenerative diseases (Park, Alz) initially from the view point of MS, as it presents both autoimmune and neurodegenerative facets [104]. Indeed, several nodes in MS interactome interacted with susceptibility genes of the other four diseases, with a subset of MS interactors taking contact with both autoimmune and neurodegenerative gwas-genes. These observations indicate that MS offers a framework that can be shared to certain extend by other diseases.

Further, we elucidated the interactors linked to the gwas-genes in all five diseases and the pathway enriched in each disease-related interactome using the ToppGene suite. Although a few previous studies revealed enriched biological pathways related to susceptibility genes in several human complex disorders, including MS [83, 135, 149, 150], no information was available on the pathways emerging from the global genetic framework. Differently, our study ascertained such issue in multiple sclerosis as well as in four additional autoimmune or neurodegenerative disorders. Numerous immune related pathways were enriched in autoimmune interactomes, which was expected as many of the susceptibility genes in RA, T1D and MS were immune related. Notably, the pathways *B-cell activation* and *T-cell activation* appeared in all autoimmune diseases. It is well known that both arms of adaptive immunity greatly contribute to autoimmunity. Our data suggest that the processes leading to alterations in immune tolerance may be caused or sustained by the genetic framework, and, hence, support the rationale for therapeutical approaches targeting T and B lymphocytes in autoimmune disorders. Surprisingly, the same pathways appeared in Alzheimer's disease. The role of adaptive immunity in Alz remains under-explored so far, however some studies suggest altered T cell phenotypes and responses in such patients (reviewed by [151]). Interestingly, regular use of anti-inflammatory drugs reduces the odds of developing Alz [152, 153]. Our observation that Alz genetic framework may have an impact on immune function questions the classical distinction between inflammatory

and non-inflammatory diseases and supports the hypothesis that, even though the primary insult is not inflammation but neurodegeneration, immunological pathways play a role in the etiopathogenesis of Alzheimer's disease [154].

The paired comparisons of genetic interactomes allowed measuring the degree of relatedness among the five disorders. Sharing of interactors and pathways was highly significant among the autoimmune group. This may be partly explained by the genetic overlap between autoimmune diseases [155, 156]. In fact, among all autoimmune gwas-genes analyzed in our study, HLA-DRB1 was associated with MS, T1D and RA, three genes (PTPN22, PRKCQ, CTLA4) were shared between T1D and RA, three (IL2RA, IL7R and CLEC16A) between MS and T1D, and one more (CD40) between MS and RA. However, analyses at the genetic network level showed also that several gwas-genes specific for single pathologies converged at the interactome level, meaning that, although the primary events may differ, the resulting functional cascades may come together and lead to alterations in the same pathways.

The most surprising observation was the strong relatedness in T1D-Alz pair in terms of shared interactome and pathways. Clinical and epidemiological data are available about associations between T1D and Alz. Type 1 (and Type 2) diabetic patients present deficits in numerous cognitive functions (reviewed in [157]) and diabetes is a risk factor for Alz [158]. In addition, biological evidences indicate that dysregulation of insulin metabolism may affect amyloid-β accumulation and degradation [159].

On the whole, we have performed an unprecedented comparison among the genetic interactomes derived from the genes predisposing to five human complex disorders. The shift in the network analysis from the gwas-genes to their first-degree interactors made the detection of shared molecules possible even when no interactions were present at the gwas-level.

Next, we focused our attention on transcriptomics in PBMC's from patients with RR-MS, and compared them with healthy individuals. Even though several past gene profiling studies in MS were successful in giving molecular insights [160-167], many of them did not effectively restrict the diverse parameters of the disease. For instance, Bomprezzi and colleagues [161] through PBMC transcriptomics in MS and healthy controls demonstrated that a few candidate genes could accurately classify MS patients and healthy controls. Unfortunately, out of the 24 MS patients used in the study 18 were RR-MS and 6 were in the secondary progressive course of the disease, moreover active or stable phase of disease was not specified. In addition, in this and in other studies [160, 161, 163, 164] the analyses were not unbiased but restricted to a selection of potentially interesting genes. Critically considering these issues, we recruited patients with the relapsing-remitting course of the disease; all were clinically stable, free from immunomodulatory therapy and from other inflammatory or autoimmune disorders. Moreover, the standardized Illumina gene expression platform containing more than 22,000 probes was used for all the

hybridizations. We assumed that a gender-based analysis of the transcriptome data could possibly eliminate a considerable amount of heterogeneity seen in the heatmap. The analysis of the coefficient of variation among selected probes from the couple and gender-based datasets strongly supported our hypothesis as significant reduction in variability both in the healthy and in the MS genelists was observed. Further, we investigated whether sexual dimorphism in MS was driven by sex-specific genes. Sex related gene expression differences are normally present in several tissues, including blood [168, 169], however they have never been specifically monitored under disease. Using relaxed statistics to maximize the number of differences existing in blood between genders, about 10% of the filtered genes resulted sex-specific in the healthy population, while only 2% in MS subjects. Moreover, only a minor part of the MS sex-specific genes were natural, pre-existing sex-specific genes. These data clearly demonstrated that MS pathology is associated with dysregulation in sex-specific genes. The factors leading to physiological gender differences are complex and include genetic, hormonal and environmental stimuli. It is well known that sex makes a difference in MS susceptibility and clinical course, and sex hormones may modulate disease [170]. However, the contribution of sex-related factors to the same clinical course in men and women remains still elusive. Certainly, our observations cannot be allocated to specific etiological factors. However, our data showed for the first time, that the genes accountable for the remission state of MS are different in men and women and that part of this diversity is driven by the sex specific

genes. On the other hand, the dissimilar DEG in the two genders were wrapped in common biological contexts and this phenomenon was associated with shared molecular functions. More interestingly, the common biological themes were dominated by ontology terms related to transcription, DNA binding and chromatin modification, strongly indicating that epigenetics may be the underlying pathogenic mechanism in multiple sclerosis. Further, we used system biology tools to explore the genes interacting with the DEG. We reconstructed the global interactome relative to the female and male genelists and found 171 distinct genes interacting with both the female and male DEG.

A recently published study proposes that differential expression in MS might be sustained by a network of regulatory transcription factors [171]. They found that using the TRANSFAC database several transcription factor binding motifs were overrepresented in the transcriptional signatures associated with MS. We analyzed this issue by querying the GeneCodis database, in which the annotations for human transcription factors is derived by the systematic catalogue described in [172]. This analysis unraveled that several genes in the female and male MS signatures could be regulated by common transcription factors. We focused the attention on SP1 which appeared also among our epigenetic interactors and belonged to the central core of highly connected epigenetic factors. In fact, it is implicated in chromatin remodeling [173] and may form transcriptional complexes with HDAC, NF-kappaB and AR [174, 175]. SP1 was not one of the transcription factors

described in [171], as TRANSFAC lacks high quality matrices to detect SP1 binding sites [176]. SP1 is the prototype member of a family of related transcription factors that recognize G/C rich tracts in DNA. Importantly, mitoxantrone, one of the approved drugs for the treatment of multiple sclerosis, is a type II topoisomerase inhibitor with intercalating properties at GC-rich sequences including SP1 binding sites [177], raising the hypothesis that SP1 targeting in multiple sclerosis could be beneficial. The drug WP631 is a specific inhibitor of SP1-dependent transcription [178, 179]. It belongs to the family of anthracyclines, antineoplastic compounds with potent cytotoxic effects after DNA intercalation. Compared to other anthracyclines, WP631 bisintercalates into DNA with a binding affinity close to that of transcription factors [180-182]. Treatment of Jurkat T lymphocytes with WP631 induces cell cycle arrest and death [179]. Consistently, we observed cell death and blockade of T cell proliferation in cultures of human PBMC and mouse splenocytes exposed *in vitro* to WP631. Moreover, *in vivo* administration of the drug reduced T cell responses to the encephalitogenic MOG peptide and significantly reduced EAE incidence and clinical expression, demonstrating that SP1 dependent transcription modulates autoimmune responses and that its blockade may represent a novel target for MS treatment.

In both the studies (genetic and transcriptomics), the usefulness of applying network biology approaches was evident. On the other hand, the downstream effects of the genetic modification in terms of transcriptional changes will be an interesting area to be

investigated. The transcriptome analysis needs to be extended to the progressive forms (SP and PP) of MS. Additional samples needs to be recruited in RR-MS and healthy groups. Importantly, transcriptome analysis of converted and non-converted CIS individuals will be very useful for the identification of biomarkers related to conversion. Finally, considering the present limitations in the proper diagnosis of MS sub-types, it will be very useful to apply machine learning approaches in the transcriptome data to identify the reliable gene signatures which can predict the clinical form of MS.

6 REFERENCES

1. Smith GD, Ebrahim S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32(1):1-22.

2. Dexter DT, Jenner P. Parkinson disease: From pathology to molecular disease mechanisms. Free Radic Biol Med 2013.

3. Schrag A, Ben-Shlomo Y, Quinn NP. Cross sectional prevalence survey of idiopathic parkinson's disease and parkinsonism in london. BMJ 2000;321(7252):21-22.

4. Jankovic J. Parkinson's disease: Clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008;79(4):368-376.

5. Langston JW. The parkinson's complex: Parkinsonism is just the tip of the iceberg. Ann Neurol 2006;59(4):591-596.

6. Kempster PA, O'Sullivan SS, Holton JL et al. Relationships between age and late progression of parkinson's disease: A clinico-pathological study. Brain 2010;133(Pt 6):1755-1762.

7. Popat RA, Van Den Eeden SK, Tanner CM et al. Effect of reproductive factors and postmenopausal hormone use on the risk of parkinson disease. Neurology 2005;65(3):383-390.

8. Hardy J. Genetic analysis of pathways to parkinson disease. Neuron 2010;68(2):201-206.

9. Houlden H, Singleton AB. The genetics and neuropathology of parkinson's disease. Acta Neuropathol 2012;124(3):325-338.

10. Goedert M. Parkinson's disease and other alpha-synucleinopathies. Clin Chem Lab Med 2001;39(4):308-312.

11. Goedert M, Spillantini MG. Lewy body diseases and multiple system atrophy as alpha-synucleinopathies. Mol Psychiatry 1998;3(6):462-465.

12. Dawson TM. Unraveling the role of defective genes in parkinson's disease. Parkinsonism Relat Disord 2007;13 Suppl 3:S248-9.

13. Gupta A, Dawson VL, Dawson TM. What causes cell death in parkinson's disease? Ann Neurol 2008;64 Suppl 2:S3-15.

14. International Parkinson's Disease Genomics Consortium (IPDGC), Wellcome Trust Case Control Consortium 2 (WTCCC2). A two-stage meta-analysis identifies several new loci for parkinson's disease. PLoS Genet 2011;7(6):e1002142.

15. International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V et al. Imputation of sequence variants for identification of genetic risks for parkinson's disease: A meta-analysis of genome-wide association studies. Lancet 2011;377(9766):641-649.

16. Schapira AH. Mitochondrial pathology in parkinson's disease. Mt Sinai J Med 2011;78(6):872-881.

17. Milne A. The 'D' word: Reflections on the relationship between stigma, discrimination and dementia. J Ment Health 2010;19(3):227-233.

18. Cummings JL. Alzheimer's disease. N Engl J Med 2004;351(1):56-67.

19. McKhann G, Drachman D, Folstein M et al. Clinical diagnosis of alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on alzheimer's disease. Neurology 1984;34(7):939-944.

20. Dubois B, Feldman HH, Jacova C et al. Research criteria for the diagnosis of alzheimer's disease: Revising the NINCDS-ADRDA criteria. Lancet Neurol 2007;6(8):734-746.

21. Williamson J, Goldman J, Marder KS. Genetic aspects of alzheimer disease. Neurologist 2009;15(2):80-86.

22. Rao AT, Degnan AJ, Levy LM. Genetics of alzheimer disease. AJNR Am J Neuroradiol 2013.

23. Grant WB, Campbell A, Itzhaki RF, Savory J. The significance of environmental factors in the etiology of alzheimer's disease. J Alzheimers Dis 2002;4(3):179-189.

24. Lauren J, Gimbel DA, Nygaard HB et al. Cellular prion protein mediates impairment of synaptic plasticity by amyloid-beta oligomers. Nature 2009;457(7233):1128-1132.

25. Ferri CP, Prince M, Brayne C et al. Global prevalence of dementia: A delphi consensus study. Lancet 2005;366(9503):2112-2117.

26. Blennow K, de Leon MJ, Zetterberg H. Alzheimer's disease. Lancet 2006;368(9533):387-403.

27. Bettens K, Sleegers K, Van Broeckhoven C. Genetic insights in alzheimer's disease. Lancet Neurol 2013;12(1):92-104.

28. Harold D, Abraham R, Hollingworth P et al. Genome-wide association study identifies variants at CLU and PICALM associated with alzheimer's disease. Nat Genet 2009;41(10):1088-1093.

29. Lambert JC, Heath S, Even G et al. Genome-wide association study identifies variants at CLU and CR1 associated with alzheimer's disease. Nat Genet 2009;41(10):1094-1099.

30. Anderson MS, Bluestone JA. The NOD mouse: A model of immune dysregulation. Annu Rev Immunol 2005;23:447-485.

31. Maahs DM, Rewers M. Editorial: Mortality and renal disease in type 1 diabetes mellitus--progress made, more to be done. J Clin Endocrinol Metab 2006;91(10):3757-3759.

32. Gillespie K. Type 1 diabetes - pathogenesis, genetics and immunotherapy. USA: InTechPublication, 2011.

33. Knip M, Veijola R, Virtanen SM et al. Environmental triggers and determinants of type 1 diabetes. Diabetes 2005;54 Suppl 2:S125-36.

34. Karjalainen J, Martin JM, Knip M et al. A bovine albumin peptide as a possible trigger of insulin-dependent diabetes mellitus. N Engl J Med 1992;327(5):302-307.

35. Shehadeh N, Shamir R, Berant M, Etzioni A. Insulin in human milk and the prevention of type 1 diabetes. Pediatr Diabetes 2001;2(4):175-177.

36. Bluestone JA, Herold K, Eisenbarth G. Genetics, pathogenesis and clinical interventions in type 1 diabetes. Nature 2010;464(7293):1293-1300.

37. Gillespie KM, Bain SC, Barnett AH et al. The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. Lancet 2004;364(9446):1699-1700.

38. Vafiadis P, Bennett ST, Todd JA et al. Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet 1997;15(3):289-292.

39. Pugliese A, Zeller M, Fernandez A,Jr et al. The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. Nat Genet 1997;15(3):293-297.

40. Nistico L, Buzzetti R, Pritchard LE et al. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. belgian diabetes registry. Hum Mol Genet 1996;5(7):1075-1080.

41. Ueda H, Howson JM, Esposito L et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 2003;423(6939):506-511.

42. Bottini N, Musumeci L, Alonso A et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 2004;36(4):337-338.

43. Vella A, Cooper JD, Lowe CE et al. Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. Am J Hum Genet 2005;76(5):773-779.

44. Qu HQ, Montpetit A, Ge B et al. Toward further mapping of the association between the IL2RA locus and type 1 diabetes. Diabetes 2007;56(4):1174-1176.

45. Dendrou CA, Plagnol V, Fung E et al. Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. Nat Genet 2009;41(9):1011-1015.

46. Zhou X, Bailey-Bucktrout SL, Jeker LT et al. Instability of the transcription factor Foxp3 leads to the generation of pathogenic memory T cells in vivo. Nat Immunol 2009;10(9):1000-1007.

47. McClymont SA, Putnam AL, Lee MR et al. Plasticity of human regulatory T cells in healthy subjects and patients with type 1 diabetes. J Immunol 2011;186(7):3918-3926.

48. Nokoff N, Rewers M. Pathogenesis of type 1 diabetes: Lessons from natural history studies of high-risk individuals. Ann N Y Acad Sci 2013;1281(1):1-15.

49. Pratt AG, Isaacs JD, Mattey DL. Current concepts in the pathogenesis of early rheumatoid arthritis. Best Pract Res Clin Rheumatol 2009;23(1):37-48.

50. Majithia V, Geraci SA. Rheumatoid arthritis: Diagnosis and management. Am J Med 2007;120(11):936-939.

51. John S, Myerscough A, Marlow A et al. Linkage of cytokine genes to rheumatoid arthritis. evidence of genetic heterogeneity. Ann Rheum Dis 1998;57(6):361-365.

52. MacGregor AJ, Snieder H, Rigby AS et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 2000;43(1):30-37.

53. Silman AJ, Newman J, MacGregor AJ. Cigarette smoking increases the risk of rheumatoid arthritis. results from a nationwide study of disease-discordant twins. Arthritis Rheum 1996;39(5):732-735.

54. Padyukov L, Silva C, Stolt P et al. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. Arthritis Rheum 2004;50(10):3085-3092.

55. Eerola E, Mottonen T, Hannonen P et al. Intestinal flora in early rheumatoid arthritis. Br J Rheumatol 1994;33(11):1030-1038.

56. Benson RA, Patakas A, McQueenie R et al. Arthritis in space and time--to boldly go! FEBS Lett 2011;585(23):3640-3648.

57. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. Nat Rev Rheumatol 2013;9(3):141-153.

58. Feitsma AL, van der Voort EI, Franken KL et al. Identification of citrullinated vimentin peptides as T cell epitopes in HLA-DR4-positive patients with rheumatoid arthritis. Arthritis Rheum 2010;62(1):117-125.

59. Kochi Y, Okada Y, Suzuki A et al. A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. Nat Genet 2010;42(6):515-519.

60. Smith KA, Popmihajlov Z. The quantal theory of immunity and the interleukin-2-dependent negative feedback regulation of the immune response. Immunol Rev 2008;224:124-140.

61. Orozco G, Hinks A, Eyre S et al. Combined effects of three independent SNPs greatly increase the risk estimate for RA at 6q23. Hum Mol Genet 2009;18(14):2693-2699.

62. Adrianto I, Wen F, Templeton A et al. Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. Nat Genet 2011;43(3):253-258.

63. Kool M, van Loo G, Waelput W et al. The ubiquitin-editing protein A20 prevents dendritic cell activation, recognition of apoptotic cells, and systemic autoimmunity. Immunity 2011;35(1):82-96.

64. Matmati M, Jacques P, Maelfait J et al. A20 (TNFAIP3) deficiency in myeloid cells triggers erosive polyarthritis resembling rheumatoid arthritis. Nat Genet 2011;43(9):908-912.

65. Raychaudhuri S. Mapping rare and common causal alleles for complex human diseases. Cell 2011;147(1):57-69.

66. Clanet M. Jean-martin charcot. 1825 to 1893. Int MS J 2008;15(2):59-61.

67. Kumar DR, Aslinia F, Yale SH, Mazza JJ. Jean-martin charcot: The father of neurology. Clin Med Res 2011;9(1):46-49.

68. Hauser SL, Oksenberg JR. The neurobiology of multiple sclerosis: Genes, inflammation, and neurodegeneration. Neuron 2006;52(1):61-76.

69. Noseworthy JH, Lucchinetti C, Rodriguez M, Weinshenker BG. Multiple sclerosis. N Engl J Med 2000;343(13):938-952.

70. Compston A, Coles A. Multiple sclerosis. Lancet 2002;359(9313):1221-1231.

71. Kurtzke JF. Multiple sclerosis: Changing times. Neuroepidemiology 1991;10(1):1-8.

72. Marrie RA. Environmental risk factors in multiple sclerosis aetiology. Lancet Neurol 2004;3(12):709-718.

73. Detels R, Visscher BR, Haile RW et al. Multiple sclerosis and age at migration. Am J Epidemiol 1978;108(5):386-393.

74. Polman CH, Reingold SC, Banwell B et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 2011;69(2):292-302.

75. Fazekas F, Barkhof F, Filippi M et al. The contribution of magnetic resonance imaging to the diagnosis of multiple sclerosis. Neurology 1999;53(3):448-456.

76. Ramagopalan SV, Dobson R, Meier UC, Giovannoni G. Multiple sclerosis: Risk factors, prodromes, and potential causal pathways. Lancet Neurol 2010;9(7):727-739.

77. Sospedra M, Martin R. Immunology of multiple sclerosis. Annu Rev Immunol 2005;23:683-747.

78. Haines JL, Terwedow HA, Burgess K et al. Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. the multiple sclerosis genetics group. Hum Mol Genet 1998;7(8):1229-1234.

79. Hillert J, Olerup O. HLA and MS. Neurology 1993;43(11):2426-2427.

80. De Jager PL, Jia X, Wang J et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet 2009;41(7):776-782.

81. Kemppinen A, Sawcer S, Compston A. Genome-wide association studies in multiple sclerosis: Lessons and future prospects. Brief Funct Genomics 2011;10(2):61-70.

82. International Multiple Sclerosis Genetics Consortium (IMSGC), Bush WS, Sawcer SJ et al. Evidence for polygenic

susceptibility to multiple sclerosis--the shape of things to come. Am J Hum Genet 2010;86(4):621-625.

83. Baranzini SE, Galwey NW, Wang J et al. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Hum Mol Genet 2009;18(11):2078-2090.

84. Munger KL, Levin LI, Hollis BW et al. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. JAMA 2006;296(23):2832-2838.

85. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. part I: The role of infection. Ann Neurol 2007;61(4):288-299.

86. Serafini B, Rosicarelli B, Franciotta D et al. Dysregulated epstein-barr virus infection in the multiple sclerosis brain. J Exp Med 2007;204(12):2899-2912.

87. Ghadirian P, Jain M, Ducic S et al. Nutritional factors in the aetiology of multiple sclerosis: A case-control study in montreal, canada. Int J Epidemiol 1998;27(5):845-852.

88. Frischer JM, Bramow S, Dal-Bianco A et al. The relation between inflammation and neurodegeneration in multiple sclerosis brains. Brain 2009;132(Pt 5):1175-1189.

89. Gandhi R, Laroni A, Weiner HL. Role of the innate immune system in the pathogenesis of multiple sclerosis. J Neuroimmunol 2010;221(1-2):7-14.

90. Kasper LH, Shoemaker J. Multiple sclerosis immunology: The healthy immune system vs the MS immune system. Neurology 2010;74 Suppl 1:S2-8.

91. Loma I, Heyman R. Multiple sclerosis: Pathogenesis and treatment. Curr Neuropharmacol 2011;9(3):409-416.

92. Weber MS, Prod'homme T, Youssef S et al. Type II monocytes modulate T cell-mediated central nervous system autoimmune disease. Nat Med 2007;13(8):935-943.

93. Komiyama Y, Nakae S, Matsuki T et al. IL-17 plays an important role in the development of experimental autoimmune encephalomyelitis. J Immunol 2006;177(1):566-573.

94. Haas J, Hug A, Viehover A et al. Reduced suppressive effect of CD4+CD25high regulatory T cells on the T cell immune response against myelin oligodendrocyte glycoprotein in patients with multiple sclerosis. Eur J Immunol 2005;35(11):3343-3352.

95. Haas J, Korporal M, Balint B et al. Glatiramer acetate improves regulatory T-cell function by expansion of naive CD4(+)CD25(+)FOXP3(+)CD31(+) T-cells in patients with multiple sclerosis. J Neuroimmunol 2009;216(1-2):113-117.

96. Friese MA, Fugger L. Pathogenic CD8(+) T cells in multiple sclerosis. Ann Neurol 2009;66(2):132-141.

97. Duddy M, Niino M, Adatia F et al. Distinct effector cytokine profiles of memory and naive human B cell subsets and implication in multiple sclerosis. J Immunol 2007;178(10):6092-6099.

98. Magliozzi R, Howell O, Vora A et al. Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology. Brain 2007;130(Pt 4):1089-1104.

99. Goverman J. Autoimmune T cell responses in the central nervous system. Nat Rev Immunol 2009;9(6):393-407.

100. Ransohoff RM, Kivisakk P, Kidd G. Three or more routes for leukocyte migration into the central nervous system. Nat Rev Immunol 2003;3(7):569-581.

101. Steinman L, Martin R, Bernard C et al. Multiple sclerosis: Deeper understanding of its pathogenesis reveals new targets for therapy. Annu Rev Neurosci 2002;25:491-505.

102. Steinman L, Zamvil SS. Virtues and pitfalls of EAE for the development of therapies for multiple sclerosis. Trends Immunol 2005;26(11):565-571.

103. Hu W, Lucchinetti CF. The pathological spectrum of CNS inflammatory demyelinating diseases. Semin Immunopathol 2009;31(4):439-453.

104. Trapp BD, Nave KA. Multiple sclerosis: An immune or neurodegenerative disorder? Annu Rev Neurosci 2008;31:247-269.

105. Hohlfeld R. Biotechnological agents for the immunotherapy of multiple sclerosis. principles, problems and perspectives. Brain 1997;120 (Pt 5)(Pt 5):865-916.

106. Smith KJ, Lassmann H. The role of nitric oxide in multiple sclerosis. Lancet Neurol 2002;1(4):232-241.

107. Babbe H, Roers A, Waisman A et al. Clonal expansions of CD8(+) T cells dominate the T cell infiltrate in active multiple sclerosis lesions as shown by micromanipulation and single cell polymerase chain reaction. J Exp Med 2000;192(3):393-404.

108. Lassmann H. Hypoxia-like tissue injury as a component of multiple sclerosis lesions. J Neurol Sci 2003;206(2):187-191.

109. Simon JH, Jacobs LD, Campion MK et al. A longitudinal study of brain atrophy in relapsing multiple sclerosis. the multiple sclerosis collaborative research group (MSCRG). Neurology 1999;53(1):139-148.

110. Haghikia A, Hohlfeld R, Gold R, Fugger L. Therapies for multiple sclerosis: Translational achievements and outstanding needs. Trends Mol Med 2013.

111. Farina C, Weber MS, Meinl E et al. Glatiramer acetate in multiple sclerosis: Update on potential mechanisms of action. Lancet Neurol 2005;4(9):567-575.

112. Lalive PH, Neuhaus O, Benkhoucha M et al. Glatiramer acetate in the treatment of multiple sclerosis: Emerging concepts regarding its mechanism of action. CNS Drugs 2011;25(5):401-414.

113. Prinz M, Schmidt H, Mildner A et al. Distinct and nonredundant in vivo functions of IFNAR on myeloid cells limit autoimmunity in the central nervous system. Immunity 2008;28(5):675-686.

114. Comabella M, Rio J, Espejo C et al. Changes in matrix metalloproteinases and their inhibitors during interferon-beta treatment in multiple sclerosis. Clin Immunol 2009;130(2):145-150.

115. Polman CH, O'Connor PW, Havrdova E et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. N Engl J Med 2006;354(9):899-910.

116. Miron VE, Ludwin SK, Darlington PJ et al. Fingolimod (FTY720) enhances remyelination following demyelination of organotypic cerebellar slices. Am J Pathol 2010;176(6):2682-2694.

117. Linker RA, Lee DH, Ryan S et al. Fumaric acid esters exert neuroprotective effects in neuroinflammation via activation of the Nrf2 antioxidant pathway. Brain 2011;134(Pt 3):678-692.

118. Baxter AG. The origin and application of experimental autoimmune encephalomyelitis. Nat Rev Immunol 2007;7(11):904-912.

119. Stromnes IM, Goverman JM. Active induction of experimental allergic encephalomyelitis. Nat Protoc 2006;1(4):1810-1819.

120. Miller SD, Karpus WJ. Experimental autoimmune encephalomyelitis in the mouse. Curr Protoc Immunol 2007;Chapter 15:Unit 15.1.

121. Zamvil S, Nelson P, Trotter J et al. T-cell clones specific for myelin basic protein induce chronic relapsing paralysis and demyelination. Nature 1985;317(6035):355-358.

122. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS Comput Biol 2012;8(12):e1002822.

123. Clarke GM, Anderson CA, Pettersson FH et al. Basic statistical analysis in genetic case-control studies. Nat Protoc 2011;6(2):121-133.

124. Gibson G. Hints of hidden heritability in GWAS. Nat Genet 2010;42(7):558-560.

125. Kingsmore SF, Lindquist IE, Mudge J et al. Genome-wide association studies: Progress and potential for drug discovery and development. Nat Rev Drug Discov 2008;7(3):221-230.

126. Klein RJ, Zeiss C, Chew EY et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308(5720):385-389.

127. 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(7145):661-678.

128. Hakonarson H, Grant SF, Bradfield JP et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 2007;448(7153):591-594.

129. Cooper JD, Smyth DJ, Smiles AM et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nat Genet 2008;40(12):1399-1401.

130. Barrett JC, Clayton DG, Concannon P et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41(6):703-707.

131. Lango Allen H, Estrada K, Lettre G et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 2010;467(7317):832-838.

132. Speliotes EK, Willer CJ, Berndt SI et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010;42(11):937-948.

133. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA 2008;299(11):1335-1344.

134. Farber CR. Identification of a gene module associated with BMD through the integration of network analysis and genome-wide association data. J Bone Miner Res 2010;25(11):2359-2367.

135. Baranzini SE. The genetics of autoimmune diseases: A networked perspective. Curr Opin Immunol 2009;21(6):596-605.

136. Farber CR. Systems-level analysis of genome-wide association data. G3 (Bethesda) 2013;3(1):119-129.

137. Mackay TF, Stone EA, Ayroles JF. The genetics of quantitative traits: Challenges and prospects. Nat Rev Genet 2009;10(8):565-577.

138. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. Nat Genet 1999;21(1 Suppl):33-37.

139. Gene Ontology Consortium. The gene ontology project in 2008. Nucleic Acids Res 2008;36(Database issue):D440-4.

140. Stormo GD. DNA binding sites: Representation and discovery. Bioinformatics 2000;16(1):16-23.

141. Chuang HY, Hofree M, Ideker T. A decade of systems biology. Annu Rev Cell Dev Biol 2010;26:721-744.

142. Pellegrini M, Haynor D, Johnson JM. Protein interaction networks. Expert Rev Proteomics 2004;1(2):239-249.

143. Tong AH, Evangelista M, Parsons AB et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 2001;294(5550):2364-2368.

144. Pattin KA, Moore JH. Role for protein-protein interaction databases in human genetics. Expert Rev Proteomics 2009;6(6):647-659.

145. Xenarios I, Rice DW, Salwinski L et al. DIP: The database of interacting proteins. Nucleic Acids Res 2000;28(1):289-291.

146. Hu Z, Mellor J, Wu J et al. VisANT: Data-integrating visual framework for biological networks and modules. Nucleic Acids Res 2005;33(Web Server issue):W352-7.

147. Barrenas F, Chavali S, Holme P et al. Network properties of complex human disease genes identified through genome-wide association studies. PLoS One 2009;4(11):e8090.

148. Feldman I, Rzhetsky A, Vitkup D. Network properties of genes harboring inherited disease mutations. Proc Natl Acad Sci U S A 2008;105(11):4323-4328.

149. Zhong H, Yang X, Kaplan LM et al. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. Am J Hum Genet 2010;86(4):581-591.

150. Li Y, Agarwal P. A pathway-based view of human diseases and disease relationships. PLoS One 2009;4(2):e4346.

151. Town T, Tan J, Flavell RA, Mullan M. T-cells in alzheimer's disease. Neuromolecular Med 2005;7(3):255-264.

152. McGeer PL, Schulzer M, McGeer EG. Arthritis and anti-inflammatory agents as possible protective factors for alzheimer's disease: A review of 17 epidemiologic studies. Neurology 1996;47(2):425-432.

153. Vlad SC, Miller DR, Kowall NW, Felson DT. Protective effects of NSAIDs on the development of alzheimer disease. Neurology 2008;70(19):1672-1677.

154. Zipp F, Aktas O. The brain as a target of inflammation: Common pathways link inflammatory and neurodegenerative diseases. Trends Neurosci 2006;29(9):518-527.

155. Seldin MF, Amos CI. Shared susceptibility variations in autoimmune diseases: A brief perspective on common issues. Genes Immun 2009;10(1):1-4.

156. Zenewicz LA, Abraham C, Flavell RA, Cho JH. Unraveling the genetics of autoimmunity. Cell 2010;140(6):791-797.

157. S Roriz-Filho J, Sa-Roriz TM, Rosset I et al. (Pre)diabetes, brain aging, and cognition. Biochim Biophys Acta 2009;1792(5):432-443.

158. Biessels GJ, Staekenborg S, Brunner E et al. Risk of dementia in diabetes mellitus: A systematic review. Lancet Neurol 2006;5(1):64-74.

159. Craft S, Watson GS. Insulin and neurodegenerative disease: Shared and specific mechanisms. Lancet Neurol 2004;3(3):169-178.

160. Ramanathan M, Weinstock-Guttman B, Nguyen LT et al. In vivo gene expression revealed by cDNA arrays: The pattern in relapsing-remitting multiple sclerosis patients compared with normal subjects. J Neuroimmunol 2001;116(2):213-219.

161. Bomprezzi R, Ringner M, Kim S et al. Gene expression profile in multiple sclerosis patients and healthy controls:

Identifying pathways relevant to disease. Hum Mol Genet 2003;12(17):2191-2199.

162. Achiron A, Gurevich M, Friedman N et al. Blood transcriptional signatures of multiple sclerosis: Unique gene expression of disease activity. Ann Neurol 2004;55(3):410-417.

163. Iglesias AH, Camelo S, Hwang D et al. Microarray detection of E2F pathway activation and other targets in multiple sclerosis peripheral blood mononuclear cells. J Neuroimmunol 2004;150(1-2):163-177.

164. Booth DR, Arthur AT, Teutsch SM et al. Gene expression and genotyping studies implicate the interleukin 7 receptor in the pathogenesis of primary progressive multiple sclerosis. J Mol Med 2005;83(10):822-830.

165. Achiron A, Gurevich M, Snir Y et al. Zinc-ion binding and cytokine activity regulation pathways predicts outcome in relapsing-remitting multiple sclerosis. Clin Exp Immunol 2007;149(2):235-242.

166. Gurevich M, Tuller T, Rubinstein U et al. Prediction of acute multiple sclerosis relapses by transcription levels of peripheral blood cells. BMC Med Genomics 2009;2:46.

167. Gandhi KS, McKay FC, Cox M et al. The multiple sclerosis whole blood mRNA transcriptome and genetic associations indicate dysregulation of specific T cell pathways in pathogenesis. Hum Mol Genet 2010;19(11):2134-2143.

168. Whitney AR, Diehn M, Popper SJ et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci U S A 2003;100(4):1896-1901.

169. Karlovich C, Duchateau-Nguyen G, Johnson A et al. A longitudinal study of gene expression in healthy individuals. BMC Med Genomics 2009;2:33.

170. Tomassini V, Pozzilli C. Sex hormones, brain damage and clinical course of multiple sclerosis. J Neurol Sci 2009;286(1-2):35-39.

171. Riveros C, Mellor D, Gandhi KS et al. A transcription factor map as revealed by a genome-wide gene expression analysis of whole-blood mRNA transcriptome in multiple sclerosis. PLoS One 2010;5(12):e14176.

172. Xie X, Lu J, Kulbokas EJ et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 2005;434(7031):338-345.

173. Cakouros D, Cockerill PN, Bert AG et al. A NF-kappa B/Sp1 region is essential for chromatin remodeling and correct transcription of a human granulocyte-macrophage colony-stimulating factor transgene. J Immunol 2001;167(1):302-310.

174. Lu S, Jenster G, Epner DE. Androgen induction of cyclin-dependent kinase inhibitor p21 gene: Role of androgen receptor and transcription factor Sp1 complex. Mol Endocrinol 2000;14(5):753-760.

175. Liu S, Wu LC, Pang J et al. Sp1/NFkappaB/HDAC/miR-29b regulatory network in KIT-driven myeloid leukemia. Cancer Cell 2010;17(4):333-347.

176. Chang JT, Nevins JR. GATHER: A systems approach to interpreting genomic signatures. Bioinformatics 2006;22(23):2926-2933.

177. Chiang SY, Azizkhan JC, Beerman TA. A comparison of DNA-binding drugs as inhibitors of E2F1- and Sp1-DNA complexes and associated gene expression. Biochemistry 1998;37(9):3109-3115.

178. Martin B, Vaquero A, Priebe W, Portugal J. Bisanthracycline WP631 inhibits basal and Sp1-activated transcription initiation in vitro. Nucleic Acids Res 1999;27(17):3402-3409.

179. Mansilla S, Portugal J. Sp1 transcription factor as a target for anthracyclines: Effects on gene transcription. Biochimie 2008;90(7):976-987.

180. Chaires JB, Leng F, Przewloka T et al. Structure-based design of a new bisintercalating anthracycline antibiotic. J Med Chem 1997;40(3):261-266.

181. Leng F, Priebe W, Chaires JB. Ultratight DNA binding of a new bisintercalating anthracycline antibiotic. Biochemistry 1998;37(7):1743-1753.

182. Hu GG, Shui X, Leng F et al. Structure of a DNA-bisdaunomycin complex. Biochemistry 1997;36(20):5940-5946.

7 ACKNOWLEDGEMENTS

My first debt of gratitude goes to my Ph.D supervisor, Dr. Cinthia Farina, for her invaluable support and help during these years. She nurtured my scientific curiosities, and with her support I could explore new ideas. She has been an inspiration. I also thank Prof. Enzo Medico for his scientific support and critical discussions. I thank Dr. David Horner for being my internal Ph.D supervisor.

I also thank my colleagues: Marco Di Dario and Emanuela Colombo for their support. Thanks also goes to Daniela Cantarella and Claudio Isella of Prof. Medico's lab for their time and support.

My special thanks to ALL who *collaborated* with me *outside work*: especially, *Massimo,* who supported me extensively throughout these years. Thanks to *Fabio* for his timely support during thesis writing.

Last but not the least, without the support from my family, it shall not be possible to make this long journey. I thank my wife *Dhanya*, and our son *Maanav (Manu)* for their encouragement and love. I thank all my relatives, friends and well-wishers who supported me.

PART II

OPEN ACCESS Freely available online

PLoS one

Shared Molecular and Functional Frameworks among Five Complex Human Disorders: A Comparative Study on Interactomes Linked to Susceptibility Genes

Ramesh Menon[1,2], **Cinthia Farina**[1,2]*

1 Neuroimmunology and Neuromuscular Disorders Unit, Foundation IRCCS Neurological Institute Carlo Besta, Milan, Italy, **2** Institute of Experimental Neurology (INSpe), San Raffaele Scientific Institute, Milan, Italy

Abstract

Background: Genome-wide association studies (gwas) are invaluable in revealing the common variants predisposing to complex human diseases. Yet, until now, the large volumes of data generated from such analyses have not been explored extensively enough to identify the molecular and functional framework hosting the susceptibility genes.

Methodology/Principal Findings: We investigated the relationships among five neurodegenerative and/or autoimmune complex human diseases (Parkinson's disease-Park, Alzheimer's disease-Alz, multiple sclerosis-MS, rheumatoid arthritis-RA and Type 1 diabetes-T1D) by characterising the interactomes linked to their gwas-genes. An initial study on the MS interactome indicated that several genes predisposing to the other autoimmune or neurodegenerative disorders may come into contact with it, suggesting that susceptibility to distinct diseases may converge towards common molecular and biological networks. In order to test this hypothesis, we performed pathway enrichment analyses on each disease interactome independently. Several issues related to immune function and growth factor signalling pathways appeared in all autoimmune diseases, and, surprisingly, in Alzheimer's disease. Furthermore, the paired analyses of disease interactomes revealed significant molecular and functional relatedness among autoimmune diseases, and, unexpectedly, between T1D and Alz.

Conclusions/Significance: The systems biology approach highlighted several known pathogenic processes, indicating that changes in these functions might be driven or sustained by the framework linked to genetic susceptibility. Moreover, the comparative analyses among the five genetic interactomes revealed unexpected genetic relationships, which await further biological validation. Overall, this study outlines the potential of systems biology to uncover links between genetics and pathogenesis of complex human disorders.

Citation: Menon R, Farina C (2011) Shared Molecular and Functional Frameworks among Five Complex Human Disorders: A Comparative Study on Interactomes Linked to Susceptibility Genes. PLoS ONE 6(4): e18660. doi:10.1371/journal.pone.0018660

Editor: Jason D. Barbour, University of Hawaii Manoa, United States of America

Received December 2, 2010; **Accepted** March 8, 2011; **Published** April 21, 2011

Funding: This study was supported by Fondazione Cariplo [Project 2006.0721/10.4898 to CF] and by Italian Ministry for Health [Ricerca Finalizzata RFPS-2007-1-637146 to CF]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: farina.cinthia@hsr.it

Introduction

The aim of genome-wide association studies (gwas) is to discover the common genetic variants associated with susceptibility to complex diseases. In a typical experiment, hundreds of thousands of markers are tested simultaneously in cases and controls, and the allelic frequencies of each marker in the two groups are compared, so that the contribution of single genes to the disease is quantified. However, complex diseases do not originate from changes in single genes, but from the interactions between several genetic and environmental factors. Therefore, the analysis of isolated genes is not overly informative about the biological processes underlying disease and offers limited rationale for the development of novel therapies. Theoretically, reconstruction of the molecular interaction networks linked to gwas-genes by systems biology approaches may help to elucidate the functional consequences related to each susceptibility allele and the combined effects of more genetic variants. The few reports available to date on this issue demonstrated that this approach can identify previously unseen relationships among human diseases at molecular level [1,2].

In this study, we elaborated the genetic interactomes relative to five complex human diseases. Our lab has strong interest in multiple sclerosis (MS), a chronic disorder of the central nervous system presumably of autoimmune etiology, characterized by inflammation of the white matter, demyelination and neurodegeneration [3]. Initially, we reconstructed the MS interactome and searched for interactions with genes predisposing to either neurodegenerative (Parkinson's disease Park, Alzheimer's disease Alz) or autoimmune (Type 1 diabetes T1D, Rheumatoid Arthritis RA) disorders. Then, disease interactomes of all five disorders were analyzed at the functional level by independent pathway enrichment studies. Finally, paired comparisons elucidated relatedness among the diseases. Interestingly, the shift from single genes to molecular frameworks via system biology unraveled novel functional relationships among the five complex diseases.

Table 1. Summary of gwas data.

	MS	Park	Alz	T1D	RA	Total
No. of studies	6	6	13	8	6	39
Unique genes	54	35	17	52	21	179
Gene/study ratio	9	5.83	1.30	6.50	3.50	4.58

doi:10.1371/journal.pone.0018660.t001

Results

Link between the MS interactome and the genes predisposing to other neurodegenerative or autoimmune diseases

Initially we identified the susceptibility genes that were linked to MS and to other neurodegenerative (Park, Alz) or autoimmune (T1D, RA) disorders. We utilized results from 39 published genome-wide association studies on these five human diseases available in the GWAS catalog [4] and found genetic mutations in 179 genes passing the statistical significance threshold of 10^{-5} (Table 1). There were 6 studies available for MS, Park and RA, 8 for T1D, and 13 for Alz (Table S1). Notably, the number of studies carried out for each disease did not seem to influence the total number of disease-associated genes. For instance, despite the highest number of genome-wide association studies performed for Alz, the total number of susceptibility genes reported was the lowest among the five diseases. The ratios between the number of the reported genes and the number of studies in each disease accounted for this observation (Table 1). Interestingly, MS displayed the highest number of susceptibility genes and the highest ratio, suggesting a greater genetic heterogeneity in MS than in the other four diseases. Then, as a first step towards the definition of the genetic network underlying these five diseases, we reconstructed the MS genetic interactome and checked possible links with the genes predisposing to the other neurodegenerative or autoimmune diseases. Using the VisANT tool we derived 376 first-degree interacting partners for the 54 MS gwas-genes. Among these MS interactors, 141 were connected with at least one among 17 T1D, 11 RA, and 10 Park or Alz gwas-genes (Figure 1A and Table S2). Notably, 4 MS interactors were RA (HLA-DQA1, TRAF1) or T1D (IL2, PTPN11) gwas-genes themselves. It was also evident that several gwas-genes could come into contact with two or more MS interactors (Figure 1A), and that 24 MS interactors were connected with both neurodegenerative and autoimmune genes (red nodes in Figure 1A), including the PTPN11 gene.

Now, the analysis of disease relatedness with MS is more effective when considering the genetic interactome rather than direct interactions at the gwas-gene level. In fact, no interactions existed between gwas-genes predisposing to MS and Alz, for example, but there were several shared interactions within the genetic interactomes linked to (but not including) the respective gwas-genes (Figure 1B). Therefore, the introduction of first-degree interactors in the definition of a disease-related molecular framework may lead to the discovery of relatedness among distinct complex disorders.

Molecular relatedness among autoimmune and neurodegenerative genetic interactomes

To perform a global comparative analysis among autoimmune and neurodegenerative genetic interactomes, we derived a list of interacting partners for each gwas-geneset (Table 2, Table S3). As expected, the number of interactors was higher in MS and T1D, due to the more abundant number of reported gwas-genes. However, when normalizing the number of interactors to the total number of gwas-genes in each disease, RA reported the highest

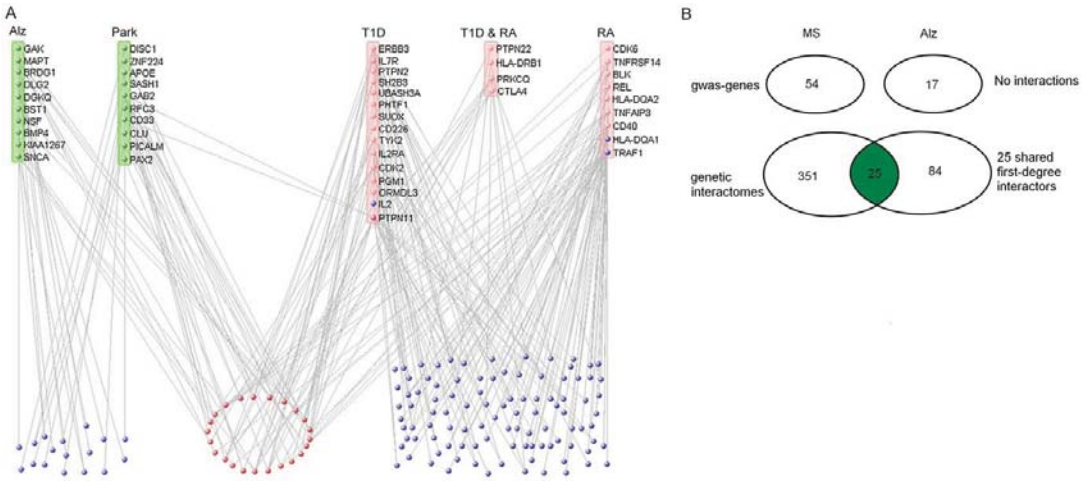


Figure 1. Relationship between MS interactome and gwas genes relative to other diseases. A. MS interactors shared by genes predisposing to neurodegenerative (green) and/or autoimmune (pink) diseases. Blue nodes indicate the MS interactors contacted by either neurodegenerative or autoimmune susceptibility genes. Red nodes indicate the MS interactors contacted by both neurodegenerative and autoimmune susceptibility genes. Note that four MS interactors are present among the T1D or RA gwas-genes. B. Comparison between MS and Alz at gwas-gene and genetic interactome levels. Despite the absence of a direct interaction at gwas level, shared molecular networks appear in the genetic interactomes (green section).
doi:10.1371/journal.pone.0018660.g001

Table 2. Genetic interactomes for neurodegenerative and/or autoimmune disorders based on VisANT database.

	Genes	Interactors	Interactor/gene ratio
Park	35	160	4.57
Alz	17	109	6.41
T1D	52	316	6.07
RA	21	249	11.85
MS	54	376	6.96

doi:10.1371/journal.pone.0018660.t002

ratio among the five diseases, indicating that at least some of the RA susceptibility genes were highly interactive. In contrast, Park displayed the lowest interactor/gene ratio, despite the discrete number of described gwas-genes.

Next, we examined the relationships among the five diseases by comparing the genetic interactomes in a pairwise fashion. The overlaps between interactomes are given in Table S4. Notably, higher concordances were found among the autoimmune diseases, with MS-RA, MS-T1D and T1D-RA having 84, 61 and 60 interactors shared respectively. Among the neurodegenerative diseases, the sharing of interactors was limited to 10 elements for Park-Alz, 20 for Park-MS and 25 for MS-Alz. Most surprisingly, 26 interactors were shared between the neurodegenerative Alz and the autoimmune T1D. The statistical significances for these observations were calculated using a hypergeometric test and are shown in Figure 2 (see bars) and Table S5. MS-RA had the lowest p-value (P = 1.02E-67), followed by T1D-RA (P = 4.42E-43) and MS-T1D (P = 3.09E-33). The T1D-Alz pair, with the p-value of 6.48E-19, was more significant than the MS-Alz pair (P = 5.21E-16). Comparatively higher p-values were found among T1D-Park (P = 1.18E-04), Alz-Park (P = 8.72E-07), RA-Park (P = 3.95E-07) and MS-Park (P = 3.36E-08), exhibiting the distant relatedness for all possible combinations with Park (Figure 2). Altogether, the results showed the close relatedness among autoimmune disorders and within Alz pairs.

The degree of relatedness may be due to the real biological properties of gwas-genes, i.e. the autoimmune gwas-genes participate in more biological pathways and processes than the neurodegenerative ones, and therefore the probability of abundant interactor sharing among autoimmune diseases is higher. However, we hypothesized that the observations could be partly biased by the difference in the number of studies carried out on distinct gwas-genes, consequently resulting in higher or lower interactor information. To critically assess this aspect, we considered that the average number of interactions per gene in the VisANT database was equal to 7.26 (93684 interactions for 12888 human genes). Consequently, we derived a *normalization factor* by normalizing each disease interaction ratio to the one in the VisANT database (Table S5). For example, the highest interaction ratio seen in RA resulted in a *normalization factor* equal to 1.88, signifying a nearly two fold increase in the interaction ratio compared to the database. Vice versa, Park normalization factor was 0.71, indicating that the interaction ratio was only 71% of that expected.

For paired analyses we multiplied the *normalization factors* relative to the two diseases (see [materials and methods]), and used the *paired normalization factor* to optimize the observed/expected ratio of interactor sharing. This resulted in an *interactor score* for each disease pair (Figure 2). So, if on the one hand the p-values relative to the inter-disease relatedness may reflect the shared interactome among diseases, on the other hand the interactor score might flatten or enhance some of these observations as it corrects for annotation bias. Therefore, both the p-values and the interactor scores have to be considered while interpreting the results. The interactor scores confirmed the close association among disease pairs in the autoimmune group (Figure 2). Surprisingly, the highest scores appeared in Alz-Park and T1D-Alz pairs. Therefore, even after eliminating the possible bias introduced by database annotation, the T1D-Alz pair maintained high association levels.

Biological themes in autoimmune and/or neurodegenerative interactomes

In order to identify the biological themes embedded in each interactome, we used the ToppGene suite, an online tool for functional enrichment analysis. A Bonferroni corrected p-value of 0.05 was used to extract the significant biological pathways reported by three distinct databases (CGAP-BioCarta, KEGG and Panther). The highest number of pathways was reported in T1D followed by MS, RA, Alz and Park respectively (Table S6). Furthermore, we tabulated the shared pathways among diseases and grouped them according to the database in seven main categories: *Growth factor/Hormone signaling*, *Innate/Adaptive immunity*, *Cell cycle and apoptosis*, *Cancer*, *Adhesion*, *Host response* and *Other* (Figure 3). The combined statistical significance of the pathway enrichments is also listed in Figure 3. The majority of shared pathways were categorized in *Growth factor/Hormone signaling*, followed by *Innate/Adaptive immunity* and *Cell cycle and apoptosis*,

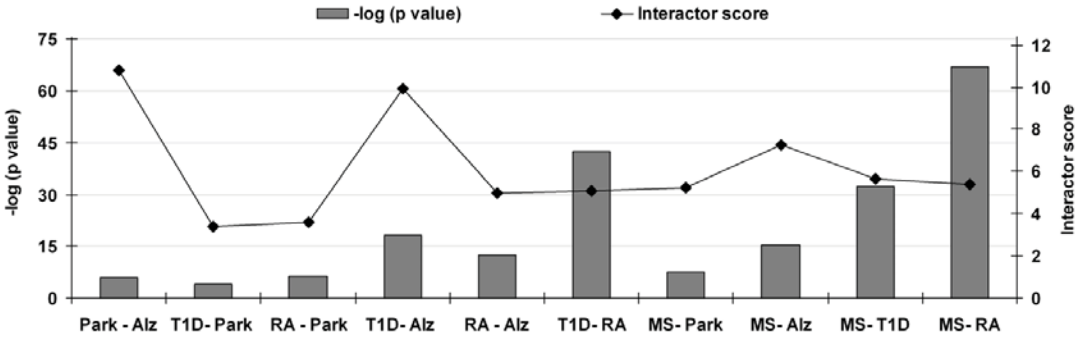


Figure 2. Molecular relatedness based on shared interactome. Statistical significance for each paired analysis is given.
doi:10.1371/journal.pone.0018660.g002

DISEASE	PATHWAY	P-VAL*
Category 1: Growth factor/ Hormone signaling		
Alz, T1D, RA	EGF receptor signaling	6.6E-13
Alz, T1D, RA	FGF signaling	3.6E-10
Alz, T1D, MS	PDGF signaling	4.2E-14
T1D, MS	Interleukin signaling	1.0E-16
T1D, MS	JAK/STAT signaling	3.6E-06
T1D, RA	PI3 kinase	1.5E-06
Alz, T1D	Angiogenesis	1.1E-18
Alz,T1D,RA,MS	ErbB signaling pathway	3.1E-33
T1D, RA, MS	Adipocytokine signaling pathway	7.1E-15
T1D, RA	Cytokine-cytokine receptor interaction	2.4E-07
T1D, MS	Insulin signaling pathway	1.5E-06
T1D, MS	Jak-STAT signaling pathway	3.8E-25
RA, MS	MAPK signaling pathway	1.1E-06
Alz, T1D, MS	EGF Signaling	8.9E-17
Alz, T1D, MS	EPO Signaling	1.8E-13
Alz, T1D, MS	IL 6 signaling	4.1E-06
Alz, T1D, MS	PDGF Signaling	6.5E-12
Alz, T1D, MS	Signaling of Hepatocyte GF Receptor	4.3E-12
Alz, T1D, MS	Angiotensin II mediated activation of JNK via Pyk2 dependent signaling	5.7E-08
Alz, T1D, MS	Sprouty regulation of Trk signals	9.4E-09
Alz, T1D, MS	Links between Pyk2 and Map Kinases	1.4E-13
Alz, T1D, MS	Trka Receptor Signaling	1.7E-09
Alz, T1D, MS	Bioactive Peptide Induced Signaling	1.1E-07
Park, RA, MS	Ceramide Signaling	1.1E-12
Park,T1D,MS	Trefoil Factors Initiate Mucosal Healing	3.6E-10
T1D, RA, MS	Chaperones modulate interferon Signaling	1.5E-07
T1D, RA, MS	Keratinocyte Differentiation	3.2E-15
Alz, MS	Nerve growth factor (NGF)	2.7E-05
Alz, T1D	IGF-1 Signaling	3.5E-05
Alz, T1D	IL 3 signaling	1.4E-06
Alz, T1D	Insulin Signaling	1.1E-04
Alz, T1D	Growth Hormone Signaling	9.1E-12
T1D, MS	IL 2 signaling	1.5E-12
T1D, MS	IL 4 signaling	4.4E-10
T1D, MS	IL-7 Signal Transduction	3.8E-09
T1D, MS	CBL mediated ligand-induced downregulation of EGF receptors	5.6E-07
T1D, MS	CXCR4 Signaling	3.3E-08
T1D, MS	TPO Signaling	4.3E-10
T1D, MS	Ca Signaling by HBx of Hepatitis B virus	5.2E-05
RA, MS	MAPKinase Signaling	8.0E-11
RA, MS	Signal transduction through IL1R	1.3E-05
RA, MS	SODD/TNFR1 Signaling	1.3E-11
RA, MS	TNF/Stress Related Signaling	2.8E-21
RA, MS	TNFR1 Signaling	2.4E-12
RA, MS	TNFR2 Signaling	5.5E-19
Park, T1D	Phosphoinositides and their downstream targets.	7.1E-04
Category 2 : Innate and adaptive immunity		
Alz,RA,T1D,MS	B cell activation	1.0E-09
Alz,RA,T1D,MS	T cell activation	1.1E-31
Alz,T1D,RA,MS	B cell receptor signaling pathway	2.5E-25
Alz,T1D,RA,MS	T cell receptor signaling pathway	3.3E-38
T1D, RA, MS	Toll-like receptor signaling pathway	7.2E-13

DISEASE	PATHWAY	P-VAL*
Category 2 : Innate and adaptive immunity (continued)		
Alz, T1D	Natural killer cell mediated cytotoxicity	7.0E-16
Alz, T1D	Fc epsilon RI signaling pathway	2.6E-09
RA, MS	Antigen processing and presentation	1.7E-13
Alz,T1D,RA,MS	T Cell Receptor Signaling	7.9E-15
Alz, T1D, MS	IL-2R Beta Chain in T cell Activation	4.3E-24
Alz, T1D, RA	BCR Signaling	1.5E-07
RA, MS	CD40L Signaling	1.6E-13
RA, MS	The 4-1BB-dependent immune response	2.6E-10
RA, MS	TACI and BCMA stimulation of B cell immune responses.	3.9E-16
RA, MS	NFkB activation by Nontypeable *H.influenzae*	5.9E-07
RA, MS	NF-kB Signaling	1.4E-15
RA, MS	Acetylation-Deacetylation of RelA in Nucleus	8.4E-13
Alz, MS	Fc Epsilon Receptor I Signaling in Mast Cells	3.3E-05
T1D, RA	Lck and Fyn tyrosine kinases in initiation of TCR Activation	1.9E-07
T1D, RA	Co-Stimulatory Signal During Tcell Activation	6.3E-10
Category 3 :Cell cycle and apoptosis		
Park, RA, MS	Apoptosis signaling	1.0E-34
T1D, RA	p53	1.2E-05
RA, MS	Apoptosis	3.1E-23
T1D, RA	Cell cycle	7.6E-23
Alz,T1D,RA,MS	PTEN dependent cell cycle arrest and apoptosis	9.7E-15
Alz, T1D, MS	Inhibition of Cellular Proliferation by Gleevec	2.4E-11
T1D, RA, MS	Telomeres,Telomerase, Aging, Immortality	1.2E-12
T1D, RA	Cyclins and Cell Cycle Regulation	1.0E-16
T1D, RA	Influence of Ras and Rho proteins on G1 to S Transition	6.5E-14
T1D, RA	Cell Cycle: G1/S Check Point	8.1E-11
T1D, RA	p53 Signaling	4.5E-07
Alz, T1D	Multiple antiapoptotics from IGF-1R signaling lead to BAD phosphorylation	8.5E-06
RA, MS	HIV-I Nef: negative effector of Fas and TNF	2.0E-29
RA, MS	Caspase Cascade in Apoptosis	1.1E-03
RA, MS	FAS signaling (CD95)	2.3E-08
RA, MS	Induction of apoptosis through DR3 and DR4/5	3.2E-18
Category 4 : Cancer		
Park,Alz,T1D,MS	Colorectal cancer	7.1E-10
Alz,T1D,RA,MS	Prostate cancer	3.1E-19
Alz,T1D,RA,MS	Pancreatic cancer	4.4E-24
Alz,T1D,RA,MS	Pathways in cancer	1.1E-46
Alz,T1D,RA,MS	Chronic myeloid leukemia	8.6E-48
T1D,RA,MS	Acute myeloid leukemia	7.7E-13
T1D,RA,MS	Small cell lung cancer	4.8E-26
Alz,T1D,RA	Glioma	1.0E-18
Alz,T1D,RA	Non-small cell lung cancer	5.9E-17
Category 5 : Adhesion		
Park,T1D,RA,MS	Focal adhesion	8.8E-14
Alz,T1D,MS	Adherens junction	1.9E-08
Park,Alz,T1D,MS	Integrin Signaling	6.7E-20
Category 6 :Host response		
Alz,RA,MS	Epithelial cell signaling in *H. pylori* infection	8.2E-11
Park,MS	Pathogenic *E. coli* infection - EHEC	6.3E-09
Park,MS	Pathogenic *E. coli* infection - EPEC	6.3E-09
Category 7 :Other		
Park,RA	Parkinson's disease	1.0E-13

Panther KEGG CGAP-BioCarta

* Combined p-values by Fisher's method.

Figure 3. Pathways shared among diseases. Colors refer to the three distinct databases used for the pathway enrichment analysis. The combined p-values for pathway overlap among diseases are given.
doi:10.1371/journal.pone.0018660.g003

while categories like *Cancer*, *Host response* and *Other* appeared in single databases.

In the *Growth factor/Hormone signaling* category, Panther indicated *EGF*, *FGF* and *PDGF signaling pathways* as shared by three diseases, Alz and T1D with MS or RA. The other pathways were shared exclusively by autoimmune diseases, with the exception of *Angiogenesis* that appeared both in T1D and Alz. In contrast to Panther, KEGG evidenced predominant pathway sharing among autoimmune disorders, except for the *ErbB signaling* present in Alz and in the three autoimmune diseases. CGAP-BioCarta exposed high concordances among Alz, T1D and MS. There were 10 pathways shared exclusively by these three diseases, among which *EGF signaling* and *PDGF signaling* were already reported by the Panther database. Many of the pathways common to these trios were related to tyrosine kinase signaling, such as *Trka receptor signaling* and *Sprouty regulation of tyrosine kinases*. Another remarkable observation was the sharing of the *NGF pathway* by MS and Alz. *IGF-1 signaling*, *IL3 signaling*, *Insulin signaling* and *Growth hormone signaling* pathways were shared exclusively by T1D and Alz. In addition, Park shared only three pathways with the autoimmune diseases, namely *Ceramide signaling*, *Trefoil factors initiate mucosal signaling* and *Phosphoinositides and their downstream targets*. Overall, pathways shared by the five diseases in the *Growth factor/Hormone signaling* category portrayed an undisputable association within the autoimmune group with the predominant presence of T1D. It also revealed that several biological themes related to tyrosine kinase signaling were shared among Alz, T1D and MS. Most unexpectedly, the analysis exposed numerous growth factor related pathways common to T1D and Alz.

The second category *Innate and adaptive immunity* contained 20 pathways derived from the three pathway databases. Although expected for autoimmune diseases, consistent sharing of immunity related pathways was also found in Alz. For example, pathways related to B and T cell activation appeared in all databases as shared by T1D, RA, MS and Alz. Furthermore, in the KEGG database Alz and T1D exclusively shared the *Fc epsilon RI signaling pathway* and *Natural killer cell mediated cytotoxicity* pathway, whereas CGAP-BioCarta emphasized predominant pathway sharing between MS and RA.

In the third category, *Cell cycle and apoptosis*, Panther and KEGG contributed with 2 pathways each, and Park shared the *Apoptosis signaling pathway* with RA and MS. Among the CGAP-BioCarta results, Alz shared 3 pathways with autoimmune diseases, while 4 pathways were shared exclusively by T1D-RA or RA-MS disease pairs.

The fourth category contained pathways derived from the KEGG database related to *Cancer*. It resulted that many genes appearing in autoimmune and/or neurodegenerative interactomes played a role in cancer related pathways.

In the fifth category (named *Adhesion*), the *Focal adhesion pathway* (KEGG) was common to Park and the autoimmune group and the *Integrin signaling pathway* (CGAP-BioCarta) was shared among neurodegenerative disorders and T1D. In addition, the *Adherens junction* pathway in KEGG was shared among Alz, T1D and MS.

The sixth category contained pathways related to *Host response*, which was reported exclusively by the KEGG database. Park and MS shared two pathways related to *E. coli* infection, and the autoimmune diseases shared the *Epithelial cell signaling in Helicobacter pylori infection* pathway. Lastly, the category *Other* contained the *Parkinson's pathway* reported by Panther, which was common to Park and RA.

Overall, pathway analysis revealed predominant sharing of functions among autoimmune diseases. Moreover, many of these pathways appeared also in Alz, which was associated with T1D in most cases.

Quantitative analysis of functional relatedness among neurodegenerative and autoimmune interactomes

We tabulated the shared pathways for all disease pairs (Table S7) and quantified the significance of observed pathway overlaps by performing a hypergeometric test for each database separately (Figure 4A). According to Panther, the most significant pathway overlap was in the T1D-Alz pair (P = 2.10E-07), followed by the T1D-RA (P = 1.48E-05) and the RA-Alz (P = 8.69E-05). The MS-Park pair displayed no significant pathway overlap (P = 0.184). Among the KEGG results, the most highly significant pathway overlap was between MS and RA (P = 1.03E-13), followed by T1D-RA (P = 51.22E-12), T1D-Alz (P = 5.91E-11), MS-T1D (P = 8.58E-10) etc. Yet, the analysis revealed that pathway overlaps with Park could be due to chance, except for the MS-Park pair (P = 5.44E-03). Finally, the CGAP-BioCarta analysis recognized T1D-Alz as the most significant association (P = 3.71E-12), followed by MS-RA (P = 1.60E-11), MS-T1D (P = 2.16E-09), MS-Alz (P = 1.60E-08) etc. To summarize, the p-values based on KEGG and CGAP-BioCarta databases were more significant among the disease pairs in the autoimmune group, especially MS-RA. Moreover, on the basis of consistent statistical results in all the three databases, the T1D-Alz pair can be considered as the most significant disease pair in the context of shared pathways.

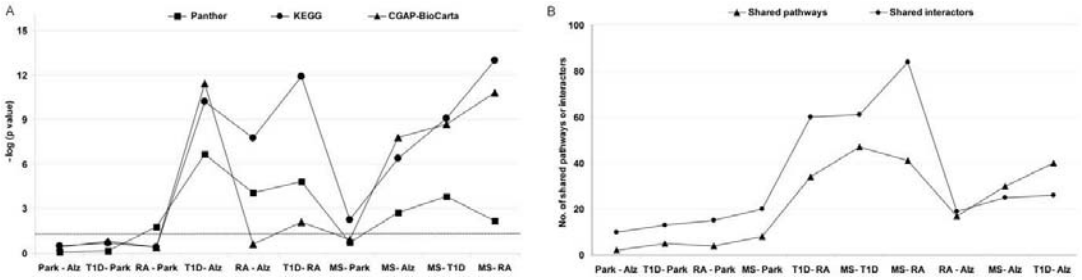


Figure 4. Molecular vs. functional relatedness. A. Functional relatedness based on shared pathways. Statistical significance for each paired analysis is given as relative to the database. The dotted line indicates the p-value threshold of 0.05. B. Shared pathways and interactors among disease pairs.
doi:10.1371/journal.pone.0018660.g004

Theoretically, pathway sharing was partly anticipated as consistent interactor sharing was present in several disease pairs. Clearly, the more interactors were shared, the more common pathways were expected. We critically evaluated this issue and measured the Pearson's correlation between the interactors and the pathways (as reported by the three databases together) shared among the disease pairs. Consistently, we observed that the number of shared pathways was proportional to the number of shared interactors in most cases (Figure 4B). However, the Alz pairs with autoimmune diseases were not following this trend. The overall Pearson's correlation coefficient was 0.79, which increased remarkably to 0.95 when Alz-autoimmune disease pairs were excluded. These observations demonstrate that the number of shared biological pathways among Alz and autoimmune diseases is higher than expected on the basis of interactor sharing.

Finally, we ranked the associations among diseases based on the statistical significance of shared pathways in each database (Table S6). Scores starting from 1 were assigned from the highest significant association to the lowest. Those with insignificant p-values were assigned the rank 10. Then, we derived the cumulative ranks for each disease pair by summing their ranks in the three databases. As expected, T1D-Alz pair had the highest rank, followed by the autoimmune disease pairs MS-RA and T1D-RA, and then by MS-Alz and RA-Alz. All Park pairs scored very low. These results were depicted in Figure 5 displaying a network summary of relationships among the five diseases.

Altogether, the relationships among diseases identified at the level of molecular interactions were found to be sustained at the level of biological pathways, especially among autoimmune diseases. Surprisingly, the same approaches evidenced strong relatedness between Type 1 diabetes and Alzheimer's disease.

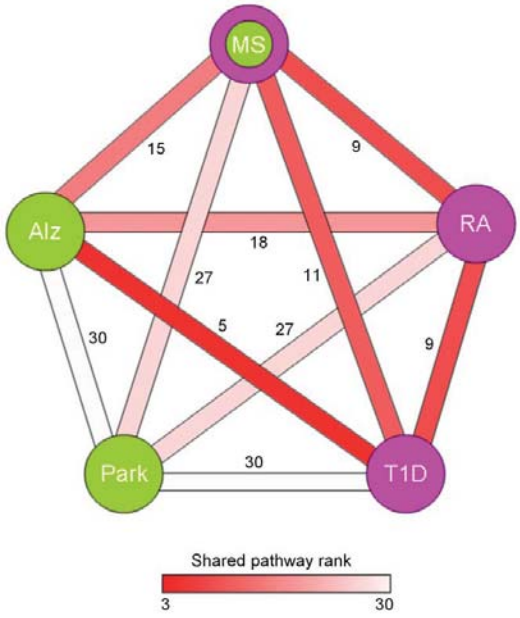


Figure 5. Overall disease relatedness based on shared pathways in the Panther, KEGG and CGAP-BioCarta databases. Green nodes indicate the neurodegenerative disorders, whereas pink nodes highlight the autoimmune diseases. The color of the edges connecting the nodes reflects the shared pathway rank ranging from 3 (highest relatedness) to 30 (lowest relatedness).
doi:10.1371/journal.pone.0018660.g005

Discussion

The reconstruction of the molecular framework hosting genetic variants associated with susceptibility to a complex human disorder has the potential to reveal the biological mechanisms underlying that disease and to highlight similarities with other diseases. Here we made a cross-disease comparison of autoimmune (T1D, RA) and neurodegenerative diseases (Park, Alz) starting from the point of view of MS, as it presents both autoimmune and neurodegenerative facets [3]. Indeed, several nodes in the MS interactome interacted with the susceptibility genes of the other four diseases, with a subset of MS interactors making contact with both autoimmune and neurodegenerative gwas-genes. These observations indicate that MS offers a framework that, to a certain extent, can be shared by other diseases. This is partly due to the fact that genetic susceptibility to autoimmune disorders may be driven by variations in some common genes [5,6]. In this regard, two scenarios have been demonstrated. In the first case, the same variant may be involved in susceptibility to distinct diseases. For instance, genetic variation in the CLEC16A gene is associated with MS as well as T1D in the Sardinian population [7]. In the second case, distinct variants in the same gene predispose to distinct diseases, i.e. polymorphisms in the IL2RA gene linked to MS or T1D susceptibility are different [8,9].

Previous studies have shown that molecular relatedness between distinct diseases may be found when searching for direct relationships among gwas-genes [10,11] and that the analysis of first-degree interactors may also lead to similar observations[1,2]. However, the advantage of the second approach compared to the first one has never been specifically addressed. Our study shows that the analysis of first-degree genetic interactomes may highlight shared molecular frameworks undetectable at the gwas-gene level. In fact, even when no interactions exist between gwas-genes, several shared interactions may be found when considering the genetic interactomes.

Furthermore, we elucidated the interactors linked to the gwas-genes in all five diseases and the pathway enriched in each disease-related interactome using the ToppGene suite. This tool contains a meta-database of annotated pathway databases such as Panther, KEGG, and CGAP-BioCarta. As the level of annotation and the number of genes related to a pathway may vary among databases [12], the outputs were not pooled but shown in relation to each database. Although a few previous studies revealed enriched biological pathways related to susceptibility genes in several complex human disorders, including MS [10,13–15], no information was available on pathways emerging from the global genetic framework. In contrast, our study ascertained such issue in multiple sclerosis as well as in four additional autoimmune or neurodegenerative disorders. Independent pathway analyses demonstrated several commonalities among distinct genetic interactomes. Numerous immune related pathways were enriched in autoimmune interactomes. This result was expected as many of the susceptibility genes in RA, T1D and MS were immune related. Notably, the pathways *B-cell activation* and *T-cell activation* appeared in all autoimmune diseases. It is well known that both arms of adaptive immunity greatly contribute to autoimmunity. Our data suggest that the processes leading to alterations in immune tolerance may be caused or sustained by the genetic framework, and, hence, support the rationale for therapeutical approaches

targeting T and B lymphocytes in autoimmune disorders. Surprisingly, the same pathways appeared in Alzheimer's disease. The role of adaptive immunity in Alz so far remains underexplored, however some studies suggest altered T cell phenotypes and responses in such patients (reviewed by [16]). Interestingly, regular use of anti-inflammatory drugs reduces the odds of developing Alz [17,18]. Our observation that the Alz genetic framework may have an impact on immune function questions the classical distinction between inflammatory and non-inflammatory diseases and supports the hypothesis that, even though the primary insult is not inflammation but neurodegeneration, immunological pathways play a role in the etiopathogenesis of Alzheimer's disease [19].

Most of the enriched pathways appeared in the category *Growth factor/Hormone signaling* consistently among the three pathway sources. Panther and CGAP-BioCarta reported the EGF related pathway as enriched in Alz, T1D and RA or MS respectively, while KEGG highlighted the ErbB signaling pathway in all four diseases. EGF receptor belongs to the ErbB gene family. Interestingly, EGF is decreased in liquor from multiple sclerosis patients [20], while increased in the synovial fluid of RA patients, where it may regulate the inflammatory process [21]. Moreover, EGF promotes the release of amyloid precursor protein (APP) [22], indicating that it might support amyloidogenesis in Alz. Finally, downregulation of the EGF receptor signaling in pancreatic islets causes diabetes [23]. Our data suggest that the involvement of the EGF pathway in the pathogenesis of various complex human disorders may be genetically determined. The FGF signaling pathway appeared in two databases as enriched in T1D and Alz. Interestingly, biological evidence links alterations in this pathway to the two diseases, e.g. attenuation of FGF signaling in mouse beta cells leads to diabetes [24], and aFGF levels in liquor are increased in Alzheimer's patients [25]. Insulin related pathways were also enriched in T1D and Alz interactomes. Obviously, the major defect in T1D is insulin deficiency caused by the autoimmune attack against pancreatic β-cells. Accumulating evidences indicate that alterations in insulin signaling may contribute to Alz pathology (reviewed in [26]). Moreover, tau phosphorylation is increased in diabetic animals [27], and mice with combined APP overexpression and diabetes show exacerbated histological features of Alz [28]. Another interesting overlap is the angiogenesis pathway shared exclusively between Alz and T1D, especially if interpreted in relation to blood brain barrier dysfunction due to increased vascular permeability induced by hyperglycemia [29]. Further sharing of *Growth factor/Hormone signaling* pathways between Alz and T1D strongly supports the hypothesis of a genetic and functional link between these two disorders.

CGAP-BioCarta highlighted TNF related pathways both in RA and MS genetic interactomes. TNF is an inflammatory mediator clearly involved in the RA and MS pathology [30,31]. Intriguingly, various biological compounds targeting TNF resulted effective in RA treatment [32], while detrimental in multiple sclerosis [33], indicating that TNF related pathways may play a dual role in autoimmune diseases.

The *Cell cycle and apoptosis* category was almost dedicated to autoimmune diseases. Such pathways regulate induction of immune tolerance and, indeed, changes in the balance between cell proliferation and death may lead to autoimmunity [34,35]. The same processes may be altered in neoplastic cells. Cancer related pathways also consistently appeared enriched in the KEGG database. Overall, our data indicate that the genetic framework predisposing to complex human disorders may contribute to changes in these basic cellular functions.

The paired comparisons of genetic interactomes allowed measuring the degree of relatedness among the five disorders. Sharing of interactors and pathways was highly significant among the autoimmune group. This may be partly explained by the genetic overlap between autoimmune diseases [5,6]. In fact, among all autoimmune gwas-genes analyzed in our study, HLA-DRB1 was associated with MS, T1D and RA, three genes (PTPN22, PRKCQ, CTLA4) were shared between T1D and RA, three (IL2RA, IL7R and CLEC16A) between MS and T1D, and one more (CD40) between MS and RA. However, analyses at the genetic network level also showed that several gwas-genes specific for single pathologies converged at the interactome level, meaning that, although the primary events may differ, the resulting functional cascades may come together and lead to alterations in the same pathways.

The most surprising observation was the strong correlation in the T1D-Alz pair in terms of shared interactome and functions. Moreover, this pair was found highly related by consistent performances in the three distinct pathway databases. Clinical and epidemiological data are available on associations between T1D and Alz. Type 1 (and Type 2) diabetic patients present deficits in numerous cognitive functions (reviewed in [36]) and diabetes is a risk factor for Alz [37]. In addition, biological evidence indicates that dysregulation of insulin metabolism may affect amyloid-β accumulation and degradation [38].

In conclusion, in this article we provided an unprecedented comparison among the genetic interactomes derived from genes predisposing to five complex human disorders. The shift in the network analysis from the gwas-genes to their first-degree interactors made the detection of shared molecules possible even when no interactions were present at the gwas-level. Furthermore, it revealed strong molecular and functional relatedness among autoimmune disorders. For example, the genetic interactomes pertaining to autoimmune diseases converged on numerous routes regarding immunity and growth factor signalling pathways. So, network generation and functional annotation highlighted several known pathogenic processes, indicating that changes in these functions might be driven or sustained by the framework linked to genetic susceptibility. Finally, the same tools underlined several inherent relationships among the five diseases at the level of genetic interactomes and biological pathways which went unnoticed until now. Type 1 diabetes and Alzheimer's disease were emblematic in this respect, as they appeared the most closely related disorders among all the disease pairs due to the extensive sharing of interactors and functions.

Overall, this study established that the reconstruction of the molecular framework hosting the genetic variants predisposing to complex human disorders can significantly contribute to our understanding of the biological functions linked to susceptibility genes. Many of these *in-silico* results highly correlate with the present experimental biology evidence, proving the reliability of systems biology tools. Furthermore, the study revealed unexpected genetic relationships, which await further biological validation.

Materials and Methods

Diseases and genetic association data

We collected genome wide association data related to five human diseases (T1D, RA, Park, Alz and MS) from the GWAS catalog (December 2009 version). The GWAS database is an online resource comprising of all published genome-wide association studies or meta-analyses of them, and provides information regarding the gene, SNP variations and their statistical significance in each study. We retrieved all the genes predisposing to each

disease with the default critical p-value (10^{-05}) of the database. The list of genes associated with each disease is summarized in Table 1 and the detailed records about each study such as sample size, journal name, authors, site of polymorphism etc. are summarized in the Table S1.

Interactome analysis

To derive the list of interacting partners of the gwas-genes, we used the VisANT tool (December 2009 version) [39], a web-enabled program for construction, visualization, and analysis of molecular and higher order networks based on functional and physical relations of the genesets. Further, VisANT retrieves the interaction information from databases such as BIND, MIPS, Biogrid, HPRD and CAGT. The lists of interactors are given in Table S3. The statistical significances for the interactor overlap among disease pairs were calculated using a hypergeometric test. In order to compensate the eventual biases in database annotation, we calculated the *interactor score* through the normalization of the observed/expected ratio. In the VisANT database there were 93684 interactions among 12888 *Homo sapiens* genes, resulting in an average of 7.26 interactions per gene (VisANT interaction ratio). This ratio was used for normalization, as follows:

Disease interaction ratio = No. of interactions/No. of gwas-genes.

Normalization factor = Disease interaction ratio/VisANT interaction ratio.

Paired normalization factor = (Disease1 normalization factor) x (Disease2 normalization factor).

Observed = No. of interactors shared within a disease pair.

Expected = (No. of interactors in disease1/12888) x (No. of interactors in disease2).

Interactor score = (Observed/Expected) x (Paired normalization factor).

Pathway enrichment study

We chose the ToppGene suite for determination of pathways enriched in the genetic interactomes. At the time of analysis, this tool maintained a meta-database of pathway information derived from KEGG (update: June 2009), Panther (update information not available), CGAP-BioCarta (update: August 2009). The analysis was performed with the stringent criteria of Bonferroni corrected p-value cut-off 0.05 in all cases. The combined p-values were calculated for the shared pathways among diseases using Fisher's method [40].

Supporting Information

Table S1 Detailed records about the studies in the GWAS catalog used in our analyses.
(XLS)

Table S2 Overlap of MS genetic interactome with genes predisposing to other neurodegenerative or autoimmune diseases.
(XLS)

Table S3 List of the gwas genes and their interactors.
(XLS)

Table S4 Shared interactors in disease pairs.
(XLS)

Table S5 Statistical significances for interactor sharing and interactor scores.
(XLS)

Table S6 Statistical significances for pathway sharing and ranking of disease associations.
(XLS)

Table S7 List of shared pathways in disease pairs.
(XLS)

Acknowledgments

We thank Dr. Zhenjun Hu at Boston University for help with the VisANT database.

Author Contributions

Conceived and designed the experiments: RM CF. Performed the experiments: RM. Analyzed the data: RM CF. Wrote the paper: RM CF.

References

1. Barrenas F, Chavali S, Holme P, Mobini R, Benson M (2009) Network properties of complex human disease genes identified through genome-wide association studies. PLoS One 4: e8090.
2. Feldman I, Rzhetsky A, Vitkup D (2008) Network properties of genes harboring inherited disease mutations. Proc Natl Acad Sci U S A 105: 4323–4328.
3. Trapp BD, Nave KA (2008) Multiple sclerosis: An immune or neurodegenerative disorder? Annu Rev Neurosci 31: 247–269.
4. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106: 9362–9367.
5. Seldin MF, Amos CI (2009) Shared susceptibility variations in autoimmune diseases: A brief perspective on common issues. Genes Immun 10: 1–4.
6. Zenewicz LA, Abraham C, Flavell RA, Cho JH (2010) Unraveling the genetics of autoimmunity. Cell 140: 791–797.
7. Zoledziewska M, Costa G, Pitzalis M, Cocco E, Melis C, et al. (2009) Variation within the CLEC16A gene shows consistent disease association with both multiple sclerosis and type 1 diabetes in sardinia. Genes Immun 10: 15–17.
8. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, et al. (2007) Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med 357: 851–862.
9. Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, et al. (2007) Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. Nat Genet 39: 1074–1082.
10. Baranzini SE (2009) The genetics of autoimmune diseases: A networked perspective. Curr Opin Immunol 21: 596–605.
11. Moskvina V, Craddock N, Holmans P, Nikolov I, Pahwa JS, et al. (2009) Gene-wide analyses of genome-wide association data sets: Evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. Mol Psychiatry 14: 252–260.
12. Huang R, Wallqvist A, Covell DG (2006) Comprehensive analysis of pathway or functionally related gene expression in the national cancer institute's anticancer screen. Genomics 87: 315–328.
13. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE (2010) Integrating pathway analysis and genetics of gene expression for genome-wide association studies. Am J Hum Genet 86: 581–591.
14. Li Y, Agarwal P (2009) A pathway-based view of human diseases and disease relationships. PLoS One 4: e4346.
15. Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, et al. (2009) Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Hum Mol Genet 18: 2078–2090.
16. Town T, Tan J, Flavell RA, Mullan M (2005) T-cells in alzheimer's disease. Neuromolecular Med 7: 255–264.
17. McGeer PL, Schulzer M, McGeer EG (1996) Arthritis and anti-inflammatory agents as possible protective factors for alzheimer's disease: A review of 17 epidemiologic studies. Neurology 47: 425–432.
18. Vlad SC, Miller DR, Kowall NW, Felson DT (2008) Protective effects of NSAIDs on the development of alzheimer disease. Neurology 70: 1672–1677.
19. Zipp F, Aktas O (2006) The brain as a target of inflammation: Common pathways link inflammatory and neurodegenerative diseases. Trends Neurosci 29: 518–527.
20. Scalabrino G, Galimberti D, Mutti E, Scalabrini D, Veber D, et al. (2010) Loss of epidermal growth factor regulation by cobalamin in multiple sclerosis. Brain Res 1333: 64–71.

21. Nah SS, Won HJ, Ha E, Kang I, Cho HY, et al. (2010) Epidermal growth factor increases prostaglandin E2 production via ERK1/2 MAPK and NF-kappaB pathway in fibroblast like synoviocytes from patients with rheumatoid arthritis. Rheumatol Int 30: 443–449.
22. Refolo LM, Salton SR, Anderson JP, Mehta P, Robakis NK (1989) Nerve and epidermal growth factors induce the release of the alzheimer amyloid precursor from PC 12 cell cultures. Biochem Biophys Res Commun 164: 664–670.
23. Miettinen PJ, Ustinov J, Ormio P, Gao R, Palgi J, et al. (2006) Downregulation of EGF receptor signaling in pancreatic islets causes diabetes due to impaired postnatal beta-cell growth. Diabetes 55: 3299–3308.
24. Hart AW, Baeza N, Apelqvist A, Edlund H (2000) Attenuation of FGF signalling in mouse beta-cells leads to diabetes. Nature 408: 864–868.
25. Mashayekhi F, Mirzajani E, Naji M, Azari M (2010) Expression of insulin-like growth factor-1 and insulin-like growth factor binding proteins in the serum and cerebrospinal fluid of patients with parkinson's disease. J Clin Neurosci 17: 623–627.
26. Freude S, Schilbach K, Schubert M (2009) The role of IGF-1 receptor and insulin receptor signaling for the pathogenesis of alzheimer's disease: From model organisms to human disease. Curr Alzheimer Res 6: 213–223.
27. Kim B, Backus C, Oh S, Hayes JM, Feldman EL (2009) Increased tau phosphorylation and cleavage in mouse models of type 1 and type 2 diabetes. Endocrinology 150: 5294–5301.
28. Jolivalt CG, Hurford R, Lee CA, Dumaop W, Rockenstein E, et al. (2010) Type 1 diabetes exaggerates features of alzheimer's disease in APP transgenic mice. Exp Neurol 223: 422–431.
29. Taguchi A (2009) Vascular factors in diabetes and alzheimer's disease. J Alzheimers Dis 16: 859–864.
30. Sharief MK, Hentges R, Ciardi M (1991) Intrathecal immune response in patients with the post-polio syndrome. N Engl J Med 325: 749–755.
31. Kollias G, Douni E, Kassiotis G, Kontoyiannis D (1999) The function of tumour necrosis factor and receptors in models of multi-organ inflammation, rheumatoid arthritis, multiple sclerosis and inflammatory bowel disease. Ann Rheum Dis 58(Suppl 1): I32–9.
32. Taylor PC (2009) How do the efficacy and safety of abatacept and infliximab compare in the treatment of active RA? Nat Clin Pract Rheumatol 5: 126–127.
33. (1999) TNF neutralization in MS: Results of a randomized, placebo-controlled multicenter study. the lenercept multiple sclerosis study group and the university of british columbia MS/MRI analysis group. Neurology 53: 457–465.
34. Chervonsky AV (1999) Apoptotic and effector pathways in autoimmunity. Curr Opin Immunol 11: 684–688.
35. Gatzka M, Walsh CM (2007) Apoptotic signal transduction and T cell tolerance. Autoimmunity 40: 442–452.
36. S Roriz-Filho J, Sa-Roriz TM, Rosset I, Camozzato AL, Santos AC, et al. (2009) (Pre)diabetes, brain aging, and cognition. Biochim Biophys Acta 1792: 432–443.
37. Biessels GJ, Staekenborg S, Brunner E, Brayne C, Scheltens P (2006) Risk of dementia in diabetes mellitus: A systematic review. Lancet Neurol 5: 64–74.
38. Craft S, Watson GS (2004) Insulin and neurodegenerative disease: Shared and specific mechanisms. Lancet Neurol 3: 169–178.
39. Hu Z, Snitkin ES, DeLisi C (2008) VisANT: An integrative framework for networks in systems biology. Brief Bioinform 9: 317–325.
40. Fisher, A R (1932) Statistical methods for research workers. Edinburgh: Oliver & Boyd.

Journal of Autoimmunity 38 (2012) J144–J155

Contents lists available at SciVerse ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

Gender-based blood transcriptomes and interactomes in multiple sclerosis: Involvement of SP1 dependent gene transcription

Ramesh Menon [a,b,1], Marco Di Dario [a,b,1], Chiara Cordiglieri [a,b], Silvia Musio [b], Loredana La Mantia [b], Clara Milanese [b], Anna Luisa Di Stefano [c], Massimo Crabbio [c], Diego Franciotta [d], Roberto Bergamaschi [c], Rosetta Pedotti [b], Enzo Medico [e], Cinthia Farina [a,b,*]

[a] Institute of Experimental Neurology (INSpe), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy
[b] Neuroimmunology and Neuromuscular Disorders Unit, Foundation IRCCS Neurological Institute Carlo Besta, Milan, Italy
[c] Department of Neurological Sciences, Foundation IRCCS Neurological Institute Casimiro Mondino, Pavia, Italy
[d] Laboratory of Neuroimmunology, Foundation IRCCS Neurological Institute Casimiro Mondino, Pavia, Italy
[e] Laboratory of Functional Genomics, Institute for Cancer Research and Treatment (IRCC), University of Turin Medical School, Candiolo, Italy

ARTICLE INFO

Article history:
Received 6 October 2011
Received in revised form
11 November 2011
Accepted 12 November 2011

Keywords:
Multiple sclerosis
Transcriptomics
Gender
Systems biology
SP1 transcription factor

ABSTRACT

In this study we investigated the contribution of gender to global gene expression in peripheral blood mononuclear cells from multiple sclerosis (MS) patients and healthy controls. We observed that, in contrast to the conventional approach, gender-based case-control comparisons resulted in genelists with significantly reduced heterogeneity in human populations. In addition, MS was characterized by significant changes both in the quantity and in the quality of the sex-specific genes. Application of stringent statistics defined gender-based signatures which classified a second independent MS population with high precision. The global unsupervised cluster analyses for 60 subjects showed that 29/31 female and 27/29 male samples were properly identified. Notably, MS was associated in women and in men with distinct gene signatures which however shared several molecular functions, biological processes and interactors. Issues regarding epigenetic control of gene expression appeared as the main common theme for disease, with a central role for the functional modules related to histone deacetylase, NF-kappaB and androgen receptor signaling. Moreover, *in silico* analyses predicted that the differential expression in MS women and men were depending on the transcription factor SP1. Specific targeting of this pathway by the bis-anthracycline WP631 impaired T cell responses *in vitro* and *in vivo*, and reduced the incidence and the severity of experimental autoimmune encephalomyelitis, indicating that SP1 dependent gene transcription sustains neuroinflammation.

Thus, the gender-based approach with its reduced heterogeneity and the systems biology tools with the identification of the molecular and functional networks successfully uncovered the differences but also the commonalities associated to multiple sclerosis in women and men. In conclusion, we propose gender-based systems biology as a novel tool to gain fundamental information on disease-associated functional processes.

1. Introduction

Multiple sclerosis (MS) is a chronic demyelinating disorder of the central nervous system (CNS) characterized by the presence of inflammatory cells and mediators within nervous tissue [1,2]. It is a chronic disease with onset in young adulthood, prevalence of 1:500 in some geographical regions and predominance in women (2:1 female/male ratio) with the relapsing-remitting (RR-MS) form of disease [3]. Diagnosis of MS is difficult due to the variability of clinical symptoms patients experience and to the resemblance with other neurological disorders of CNS. Currently, diagnostic criteria are mainly based on clinical and radiological examinations, as specific laboratory tests are not yet available. Furthermore, there is still no single biomarker that correlates accurately with clinical activity or treatment response [4,5].

Multiple sclerosis is a complex disease determined by both genetic and environmental factors. Intriguingly, many of the susceptibility genes play a role in immune system. For instance, classical linkage

* Corresponding author. Institute of Experimental Neurology (INSpe), DIBIT II - 1st Floor, San Raffaele Scientific Institute, via Olgettina 58, 20132 Milan, Italy. Tel.: +39 2 26436180.
E-mail address: farina.cinthia@hsr.it (C. Farina).
[1] These authors contributed equally to the work.

0896-8411/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jaut.2011.11.004

analysis in large families with MS pointed out mainly the importance of the major histocompatibility complex locus in determining predisposition to the disease [6]. More recently, application of high-throughput screening techniques in genetics association studies [6] identified additional polymorphisms in immune related genes such as interleukin (IL) 7 receptor and IL2 receptor [7]. Although MS affects the CNS, there are evidences of altered immunity in the periphery in MS patients [2]. Further, the most widely used therapeutic drugs in MS are either immunosuppressive or immunomodulatory agents [2,8], indicating that targeting peripheral immune system is beneficial to patients with this CNS disorder. These observations sustain the rationale for employing peripheral blood mononuclear cells (PBMC) as an easily accessible and informative source of biological material in transcriptome studies. Transcriptional profiling aims to comprehend complex molecular interactions and to identify disease related biomarker. Such information may lead to new hypotheses that either would not be considered based on current knowledge or which are too complex to be examined by conventional approaches. Here, we performed microarray-based gene expression profiling on peripheral blood mononuclear cells (PBMC) derived from 23 MS patients and 22 healthy controls (CTRL). To reduce variability, we included patients with the relapsing-remitting clinical course, who were free from immunomodulatory treatments and from any other acute or chronic inflammatory disorder. Distinct from the conventional disease transcriptomics approach, we applied a gender-based approach for selection of differentially expressed genes (DEG) in MS and validated them in a second independent MS population. Further, we extended bioinformatical annotation to the systems biology level and verified the involvement of a novel pathway in the animal model of multiple sclerosis.

2. Materials and methods

2.1. Inclusion criteria for RR-MS patients and healthy controls

Clinical investigations have been conducted according to the principles expressed in the Declaration of Helsinki and peripheral blood was drawn after signing of the institutional informed consent. Twenty-three MS patients were initially enrolled for this study. They were Italian adult subjects comprised of 13 women and 10 men. They had relapsing-remitting course of the disease diagnosed according to the McDonald criteria [9] and all of them were clinically stable at the time of blood sampling. Moreover, they were not suffering from any acute or chronic inflammatory diseases or other autoimmune disorders and they had not yet started any immunomodulatory therapy for MS. Sampling was performed at least 4 weeks after the last clinical attack or steroid treatment. All participants had peripheral blood counts within the reference range. A second MS population (comprised of 10 women and 8 men) was subsequently enrolled according to the same inclusion criteria. Twenty-two healthy subjects (comprised of 10 women and 12 men), who had no acute or chronic inflammatory diseases or autoimmune disorders, were included as controls for this study. All blood samplings were performed between 9 and 12 a.m.

2.2. PBMC isolation and RNA isolation

Peripheral blood mononuclear cells (PBMC) were isolated using a discontinuous density gradient (Lymphoprep, Nycomed, Oslo, Norway). Viable cells were counted by Trypan Blue (Sigma–Aldrich, Milan, Italy) exclusion. Then total RNA was extracted using TriReagent (Ambion, Applied Biosystems, Monza, Italy) and stored at −80 °C.

2.3. The microarray experiment and data processing

Total RNA extracted from PBMC was used for microarray experiments on Illumina Human_Ref-8_V2 arrays (Illumina, Son, Netherlands). Quantification and quality analyses of RNA were performed on a Bioanalyzer 2100 (Agilent, Milan, Italy). Reverse transcription and biotinylated cRNA synthesis were performed using the Illumina TotalPrep RNA Amplification Kit (Ambion), according to the manufacturer's protocol. Hybridization of the cRNAs was carried out on Illumina Human_Ref-8_V2 arrays. These arrays contain about 22,000 probes exploring the transcripts contained in the Refseq database. Array hybridization, washing, staining and scanning in the Beadstation 500 (Illumina) were performed according to standard Illumina protocols. The BeadStudio software (Illumina) was used to analyze raw data grouped by experimental condition. The data were normalized using cubic spline method as implemented in the software. The MIAMI compliant microarray data have been deposited in the EBI ArrayExpress database (Accession no. E-MTAB-380). The Beadstudio software reports a detection p-value, which represents the confidence that a given transcript is expressed above the background defined by negative control probes. This value clues the "absent" or "present" status of transcripts in the array. In this study, we considered a probe to be "present" in the array if at least 2/3 of the samples in both control and disease phenotypes had a detection p-value < 0.05. With this paradigm we removed the potential unfavorable probes from the analysis and proceeded with 10,390 probe sets for the study. In addition, we identified three outlier samples through both the principal component analysis (PCA) and hierarchical sample clustering methods using MeV package 4.3.01 [10], and removed those samples from further analysis. There were no significant differences in age (36.9 ± 11.3 vs. 35.5 ± 11.4 in female and male controls, 36.5 ± 7.5 vs. 36.3 ± 7.8 in female and male patients), disease duration (3.8 ± 3.2 vs. 6.1 ± 5.5 in female and male patients), EDSS score (1.7 ± 1.4 vs. 1.8 ± 1.5 in female and male patients) and annual relapse rate (1.5 ± 1.6 vs. 1.0 ± 0.9 in female and male patients) between the two gender groups.

For the validation group comprising of 10 female and 8 male RR-MS samples (Age: 40.5 ± 14.1 vs 36.5 ± 8.6; EDSS: 1.6 ± 0.7 vs. 1.0 ± 1.0; Disease duration: 6.9 ± 8.4 vs. 8.4 ± 6.3; Annual relapse rate: 1.1 ± 1.2 vs. 0.8 ± 0.4 in female and male patients respectively), the experimental protocols for RNA extraction, labeling and array hybridizations were performed as described above. In order to reduce the technical variation due to hybridizations made at distinct time points, batch correction was performed for the newly recruited samples using the dChip software [11].

2.4. Analysis of differential gene expression

Toward the identification of differentially expressed genes, the samples were divided into MS and healthy control groups, each comprising of 21 samples. We used three parallel statistical methods to identify the differentially expressed genes. A differentially expressed gene was the one that passed at least two tests. The first method was a two sample t-test performed on MS and healthy controls with a p-value threshold of 0.01. The t-test was performed with Welch approximation for unequal variances in the two groups using the MeV package. Secondly, we used *Significant Analysis of Microarrays* [12], a robust permutation based non-parametric method that relies on variance information present in measurements obtained from the probes. Unlike t-test, SAM outputs the q-value which represents the significance of the differential expression for a given gene. In our study, the q-value cut-off was 20% and used the SAM implementation in *Stanford tools for Excel* software, version 1.1. Finally, a bayesian approach to identify differentially

expressed genes was used as implemented in the *Bayesian Analysis of Differential Gene Expression, BADGE 1.0* [13]. The BADGE software computes the posterior probability that a gene is expressed more than one fold in the first condition than in the second condition. A false positive rate of 1% was selected for the bayesian analysis and a leave one out cross validation was carried out by which the prediction accuracy ranged from 70 to 100 percent. The unsupervised classification of samples using the differentially expressed genes was performed by means of Hierarchical clustering (Pearson correlation, average linkage), as implemented in the MeV package. For the validation experiments, we generated a heatmap where the new datasets did not contribute to the average expression value for each DEG used for the calculation of the Log_2 Ratio but were challenged with the previously identified average expression values.

2.5. Functional enrichment study

We choose the Genecodis program (as of November 2009) [14], a grid-based online tool that integrates different sources such as GO, KEGG, miRBase etc. to search for biological features that frequently co-occur in a set of genes and ranks them by statistical significance. Unlike most of the currently available tools which are designed to evaluate single annotations, Genecodis finds relationships among annotations based on co-occurrence patterns that can extend the understanding of the biological events associated to a given experimental system. Furthermore, the database was annotated with approx. 98% of the genes in our genelist. In the Genecodis analysis, we selected the hypergeometric test, corrected for multiple testing using Benjamini & Hochberg FDR method. The corrected *p*-value threshold was 0.05 and the minimum number of genes in a functional category was set to 3. Combined *p*-values were calculated using the Fisher's approach as implemented in MetaP program (http://people.genome.duke.edu/~dg48/metap.php).

2.6. Systems biology tool

To derive the interacting genes, we used the VisANT tool (as of December 2009) [15], a web-enabled program for construction, visualization, and analysis of molecular and higher order networks based on functional and physical relations of the genesets. Further, VisANT has references to several databases, such as KEGG, GenBank, and Gene Ontology database.

2.7. cDNA synthesis and real-time PCR

Total RNA was reverse transcribed using random hexamer primers and Superscript III reverse transcriptase (all from Invitrogen) following the manufacturers' instructions. Real-time PCR was performed using TAQMan Universal Master Mix (Applied BioSystem, Monza, Italy). Amplification sets were purchased at Applied BioSystem. GAPDH and ALG8 were used as housekeeping genes for high or low expression DEG respectively. mRNA levels of the target gene were expressed as percentage of the housekeeping gene.

2.8. In vitro immunological studies

Human PBMC were isolated from healthy donors by density gradient. Murine splenocytes were extracted from wild type C57BL/6N mice (Charles River Laboratories, Calco, Italy). Both human and mouse cells were seeded in 96-well round bottom plates at the concentration of 2×10^5 cells/well in RPMI medium supplemented with L-glutamine (200 mM), penicillin and streptomycin (all from Euroclone, Pero, Italy) and 5% FCS (PAA, M-Medical, Cornaredo, Italy). Immune cells were stimulated with Concanavalin A (ConA; 2.5 μg/ml; Sigma–Aldrich) in the presence of increasing concentrations of WP631-dimethansulfunate (Alexis, Enzo Life Sciences, Vinci, Italy) for 72 h. Then, cultures were pulsed for 18 h with 0.5 μCi/well of [^{3}H] thymidine, and proliferation was measured from quadruplicate samples on a micro-β counter (Perkin–Elmer, Monza, Italy). Parallel cultures were labeled with the vital dye 7-AAD marker (BD Biosciences, Buccinasco, Italy) and acquired at the flow cytometer FACS vantage (BD). The number of dead cells was calculated as relative to a known amount of beads (BD).

2.9. EAE experiments

EAE was induced in 8-week-old wild type C57BL/6N female mice by subcutaneous injection of 200 μg MOG_{35-55} emulsified in complete Freund's adjuvant (CFA) containing 5 mg/ml Mycobacterium tuberculosis (DIFCO, BD). Bordetella pertussis toxin (List, Quadratech, Epsom, England) was administered by intra-peritoneal (i.p.) injection on the day of immunization (400 ng/mouse) and 2 days later (200 ng/mouse). In SJL/J mice (Harlan Laboratories, Bolzano, Italy) EAE was induced by subcutaneous injection of 100 μg $PLP_{139-151}$ emulsified in CFA containing 2 mg/ml Mycobacterium tuberculosis. Animals were monitored daily and scored as follows: 0 = no disease; 1 = flaccid tail; 2 = gait disturbance; 3 = complete hind limb paralysis; 4 = tetraparesis; and 5 = death. C57BL6N mice received the following treatment schedule: from day 3 to day 7 p.i. a dose of 3 mg/kg/day of WP631 solubilized in 6% DMSO–PBS solution, divided into 2 daily i.p. injections. SJL/J mice received a dose of 1.5 or 3 mg/kg/day WP631 from day 5 to day 9 p.i.. Control mice received 6% DMSO–PBS solution in the same amount. All procedures involving animals were authorized by the Italian General Direction for Animal Health at the Ministry for Health.

2.10. Primary immune responses

Draining lymph nodes (popliteal, inguinal, para-aortal, axillary) and spleen cells of C57BL/6N mice receiving either vehicle or WP631 (5 animals/group) were harvested at day 10 after immunization with MOG_{35-55} peptide. Cells ($2x10^5$/well) were seeded in 96-well round bottom plates in complete RPMI medium and stimulated with 1 μg/ml MOG_{35-55} peptide. After 72 h of incubation cultures were pulsed for 18 h with 0.5 μCi/well of [^{3}H] thymidine, and proliferation was measured from quadruplicate cultures on a micro-β counter (Perkin–Elmer). The stimulation index was calculated as the ratio between stimulated and unstimulated wells for each mouse.

2.11. Statistical analysis of in vitro and in vivo experiments

Normality of data distribution was assessed by Kolmogorov–Smirnov statistics. ANOVA (in case of normal distribution) or non-parametric Mann–Whitney *U* test (in case of non-normal distribution) were performed to compare means. For statistic evaluation of EAE clinical score, the non-parametric Wilcoxon sign rank test was used. All *p*-values were two-sided and subjected to a significance level of 0.05.

3. Results

3.1. A comparison of conventional and gender-based transcriptome analysis

Conventional analysis strategy in transcriptome studies foresees that differentially expressed genes (DEG) which are significant in the disease state are derived from the comparison between healthy

and diseased populations. We applied this approach in our study which included initially 22 healthy individuals and 23 patients with relapsing-remitting MS. All the patients were clinically stable and none of them had started any immunomodulatory therapy till sampling. Peripheral blood drawings were performed at least 4 weeks after relapse treatment with steroids, in order to exclude the variations in gene expression due to drugs. We labeled RNA extracted from PBMC and hybridized Illumina microarray chips that contained about 22,000 probes. After normalization and data filtering, the sampleset was analyzed by PCA and hierarchical sample clustering methods. Three samples (1 healthy control and 2 RR-MS patients) were identified by both methods as outliers and removed from further analysis. For the remaining 21 diseased subjects there were no significant differences in age, disease duration, EDSS score and annual relapse rate between the two gender groups (see materials and methods). Control subjects were sex and age-matched.

The most widely used strategy of deriving genesets is through single statistical tests and/or fold change criteria. However, in a preliminary analysis it was observed for several genes that the mean expression values were considerably determined by far extreme expression values in few samples within the group. This could result in fold changes that did not reflect reproducible expression in most of the samples within each group (data not shown). Therefore, to facilitate higher reproducibility and reliability of results we used three statistical tests in parallel to find differential expressions and selected the genes which passed through at least two tests. This way we controlled the potential false positives without losing vital information in the transcriptomics data. The statistical tests were Welch *t*-test, Significance Analysis of Microarrays (SAM) and Bayesian analysis of differential gene expression (BADGE). The BADGE analysis also performed Leave One out Cross Validation. The DEG identified by these methods were 162, 80 and 289 for *t*-test, SAM and BADGE respectively, with 167 genes common to any two statistical tests (Fig. 1A). Hierarchical clustering of this geneset, represented in Fig. 1B, showed 65 downregulated and 102 upregulated genes in the relapsing-remitting course of the disease.

However, while observing the heatmap generated with the conventional (hereon called "couple") case-control comparison, we noticed a considerable amount of intra-group variability which was independent from the health status, as it appeared both in the MS and CTRL groups. We hypothesized that this could be due to the inherent heterogeneity in the global gene expression between women and men. To verify this aspect, we measured the coefficient of variation (CV) for sets of probes that were grossly selected (median fold change threshold of 1.3 between healthy and diseased) in the couple, male and female case-control comparisons, and then calculated the average CV in each sample group. Here, we used the median fold differences because median values are least affected by the presence of few extreme expression values. As shown in Fig. 1C, high heterogeneity was found in the diseased and control groups when genelists were generated with the couple approach. On the contrary, gender-based case-control comparisons led to genelists which exhibited a significant reduction in the variability.

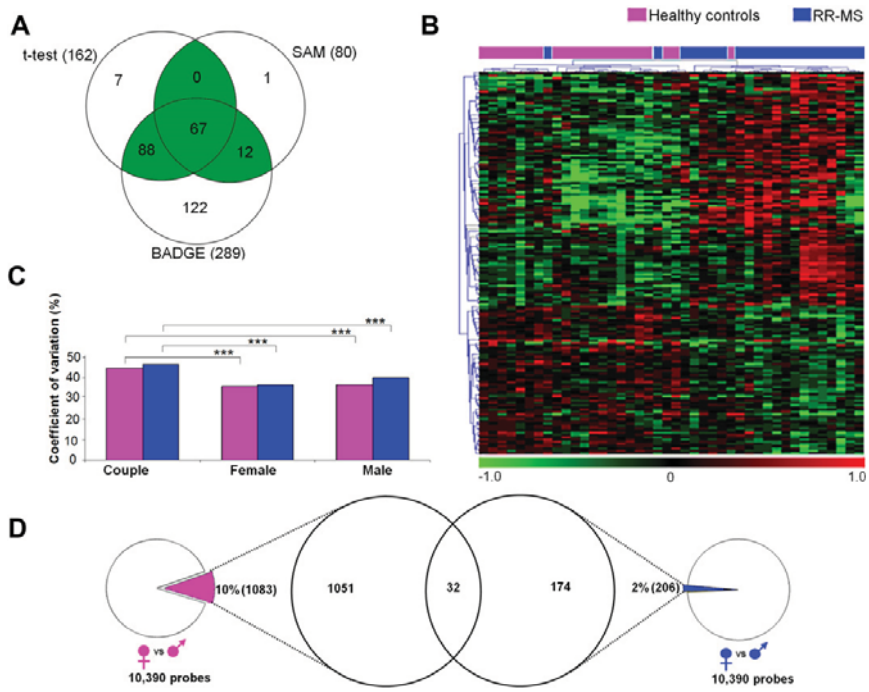


Fig. 1. The heterogeneity in human populations displayed by conventional transcriptome output is reduced in gender-based genelists. A. Number of differentially expressed genes detected by three statistical tests. The MS signature is defined by the DEG common to at least two tests (green sections). B. Heatmap showing unsupervised clustering of DEG (in rows) and of the samples (in columns). Bar above the heatmap indicates sample classification (healthy/pink vs. MS/blue subjects). Bar below the heatmap indicates expression intensity. C. Coefficients of variation (%) of grossly selected probes in couple and gender-based groups. ***p-value < 0.001. D. Left and right pie charts indicate the fraction of natural (pink) or MS (blue) sex-specific probes among the 10,390 filtered genes. The Venn diagram in the middle highlights the number of sex-specific probes overlapping among healthy and diseased populations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Finally, we assessed whether MS pathology was associated with alterations in the expression of sex-specific genes. We defined the naturally occurring sex-specific genes as those genes that were normally differentially expressed in men versus women in the healthy population. Among the 10,390 filtered probes, 1083 (10.42%) passed the relaxed *t*-test with *p*-value threshold of 0.05 and were differentially regulated between healthy women and healthy men (Fig. 1D left circle). Interestingly, when repeating the same analysis on the diseased population (Fig. 1D right circle), only 206 probes (1.98%) were sex-specific (*p*-value 2.3E-140). This significant reduction in the percentage of the sex-specific genes was accompanied also by qualitative changes, as only 32 out of the 206 MS sex-specific probes appeared among the natural sex-specific probes, indicating that in diseased subjects almost all natural sex-specific genes were not differentially expressed between women and men any more, while a new (smaller) subset of genes performed as sex-specific.

3.2. The gender-based MS signatures

Based on the previous observations, we switched to a gender-based strategy as opposed to the couple approach and compared MS women or men with their respective healthy counterpart. The female population comprised of 9 controls and 12 RR-MS, while the male group consisted of 12 controls and 9 RR-MS samples. The statistical approaches and parameters for identifying DEG were the same as in the couple analysis. In the female sample group, the *t*-test, SAM and BADGE identified 197, 139 and 301 genes respectively, with 208 probes (207 genes) common to any two of them (Fig. 2A and Supplementary file WOMEN, sheet 1). Similarly, in the male group *t*-test identified 73, SAM 19 and BADGE 126 genes (Fig. 2D). There were 72 genes common in any of these two tests (Supplementary file MEN, sheet 1).

Hierarchical clustering of 208 MS female probes (207 genes, as SNRPN gene was represented by two probes in all the tests) showed 131 downregulated and 77 upregulated probes in the disease (Fig. 2B). Unsupervised sample clustering showed that such signature was extremely efficient in classifying the healthy and diseased subjects, as 18/21 samples were properly assigned (Fig. 2B pink and blue bar). The relative gene expressions in the male sampleset were also checked and the cluster analysis is shown in the Fig. 2C. It is apparent that the MS female DEG had poor performance in male sampleset, as it resulted in a substantial rise in misclassifications in the unsupervised classification. The MS male geneset was characterized by 72 genes with 47 downregulated and 25 upregulated genes in the disease. Hierarchical clustering of these genes is shown in the Fig. 2E. Similar to the MS female DEG, the male DEG classified almost perfectly the male (20/21 correct classifications, Fig. 2E) but not the female (Fig. 2F) specimens. Overall, when compared to the couple heatmap, the gender-based heatmaps gained substantial clarity in differentiating healthy and diseased groups.

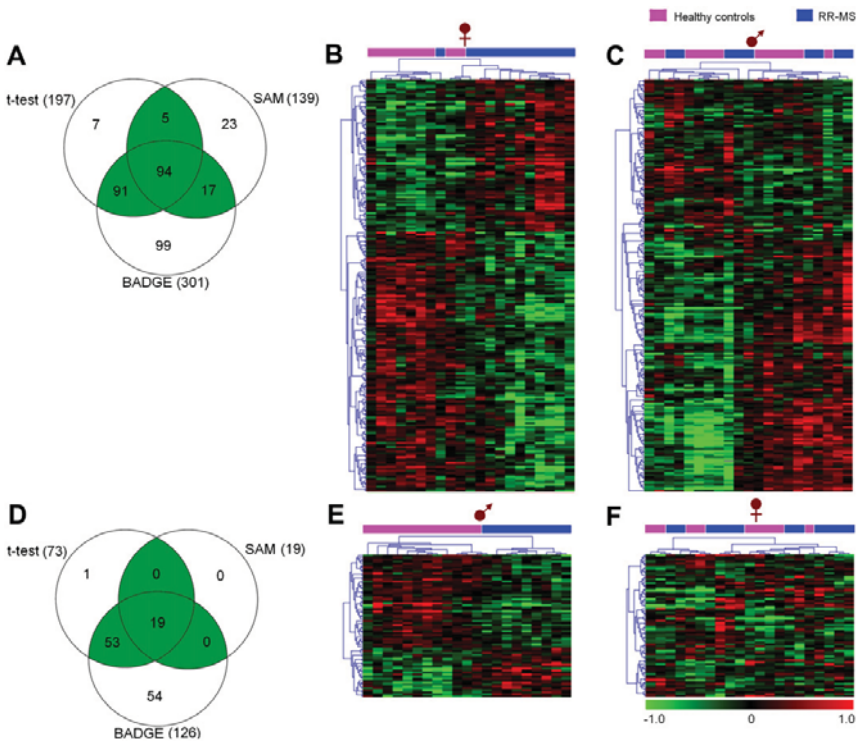


Fig. 2. Gender-based analyses in multiple sclerosis improve sample classification. A. Number of differentially expressed genes detected by three statistical tests between healthy and RR-MS women. Green sections contain genes common to at least two tests. B–C. Heatmaps showing unsupervised clustering of female DEG in female (B) or male (C) samples. D. Number of differentially expressed genes detected by three statistical tests between healthy and RR-MS men. Green sections contain genes common at least two tests. E–F. Heatmaps showing unsupervised clustering of male DEG in male (E) and female (F) samples. Bars above the heatmaps indicate sample classification (healthy/pink vs. MS/blue subjects). Bar below the heatmap indicates expression intensity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We compared the MS couple, female and male genesets to make out the commonalities and differences between them. Forty-three genes out of 207 genes in the female and 16 out of 72 genes in the male genesets were present in the couple genelist. Most unexpectedly, there was only one gene (LOC196752) common between the male and female genesets. This gene, located at 10q22.2 and coding for a protein with unknown function, appeared also in the couple geneset.

In summary, gender-based analyses allowed to limit the natural heterogeneity existing within human populations and unraveled distinct changes in blood gene expression in MS men or women although associated to the same disease form.

3.3. Validation of gender-based MS DEG in a new RR-MS population

To verify the reliability of the female and male DEG, we added two new groups of RR-MS patients comprising of 10 women and 8 men to the existing sampleset and the corresponding log ratios were calculated without adjusting the average with the expression values of newly recruited samples. Consistent with the previous results (Fig. 2), the global unsupervised cluster analysis (hierarchical clustering) classified correctly 29 out of 31 female and 27 out of 29 male samples (Supplementary Fig. 1). Moreover, we validated the differential expression of some MS female and male DEG by real-time PCR (Supplementary Fig. 1C).

3.4. Biological themes of the MS signatures

In order to explore the biological information contained in the gender-related MS genesets, we applied the Genecodis program which uses the Gene Ontology (GO) database for annotations and identifies the corresponding GO-ID for the significantly enriched terms.

Out of the 30 enriched biological process categories in the MS female geneset, the top scores were *interspecies interaction between organisms*, *transcription* and *regulation of transcription-DNA dependent* (Supplementary file WOMEN, sheet 2). In MS male geneset, due to the lower number of DEG, only 8 biological process categories were significant and *oxidation reduction- fatty acid metabolism, transcription* and *regulation of transcription-DNA dependent* were the most significantly enriched (Supplementary file MEN, sheet 2).

Most surprisingly, there was strong concordance in biological processes between the female and male geneset. The statistical significance and the number of genes for each common item are shown in Table 1. The commonalities were salient with the presence of 5 biological processes among the 8 enriched terms in men. The common GO terms were *regulation of transcription-DNA dependent*, co-occurrences of *regulation of transcription-DNA dependent* and *transcription*, *modification-dependent-protein catabolic process*, *transcription* and *chromatin modification*. As the genesets were derived from independent datasets through independent analyses, the combined *p*-values were calculated (Table 1). They ranged between 4.6E-06 and 5.6E-03, ruling out the possibility that these findings were due to chance. Therefore, even if the DEG were different in MS women and men, many of them played a role in the same biological processes. Furthermore, 4 out of 5 highly enriched biological process categories referred to transcription and chromatin modification, indicating that epigenetic events may be the common basis for MS disease in women and men.

Next, we reasoned that if the DEG were involved in the same processes, then they could exert similar functions. We tested this hypothesis by searching for recurrent molecular functions in the genelists. Among the 41 enriched molecular functional categories in female geneset, the top scores were *protein binding*, *hydrolase activity*, and *transferase activity* (Supplementary file WOMEN, sheet 3). In the male genelist, only 10 significant functions were enriched (Supplementary file MEN, sheet 3), however, similarly to the concordance seen in biological processes and consistent with our hypothesis, 6 out of 10 enriched categories in men were present in women as well. Table 2 illustrates them together with the corrected and the combined *p*-values. The shared molecular functions were *protein binding*, co-occurrences of *protein binding* and *RNA binding*, *DNA binding*, *metal ion binding*, *zinc ion binding* and *RNA binding*. Furthermore, most of the common molecular functions dealt with interactions with nucleic acids, once more affirming the prominence of epigenetic mechanisms in multiple sclerosis.

These high concordances in biological processes and molecular functions might indicate that the diverse genesets share some interacting partners. We tested this hypothesis by a systems biology approach using the VisANT program, a web-enabled tool for data-mining, visualizing, analyzing and modeling biological networks from user given input of genes or proteins. VisANT extrapolates the interacting partners for each gene by querying databases such as Biogrid, MIPS, BIND and HPRD, represents then each gene as a node and connects the interacting nodes by a straight line. We found that there were 1486 and 453 interactors for female and male genesets respectively. Notably, 171 out of 453 male interactors (about 38%) were common to female interactor set. In order to get further insights into disease related mechanisms, we checked the molecular functions and biological processes enriched in the common 171 interactors (Supplementary file INTERACTORS, sheet 1). Among the 134 significant biological processes, *interspecies interaction between organisms*, *positive regulation of I-kappaB kinase/NF-kappaB cascade*

Table 1
Biological processes (GO terms) shared by gender-based MS DEG.

Gene ontology: biological process	Gender	No. of genes	Corrected *p*-value	Combined *p*-value
GO:0006350: transcription	F	23	0.0001283	4.6E-06
	M	10	0.0022488	
GO:0006355: regulation of transcription, DNA dependent & GO:0006350: transcription	F	20	0.0003	1.8E-05
	M	8	0.0042338	
GO:0006355: regulation of transcription, DNA dependent	F	23	0.0010434	7.4E-05
	M	10	0.0054379	
GO:0016568: chromatin modification	F	4	0.0176748	9.0E-04
	M	3	0.004822	
GO:0019941: modification-dependent-protein catabolic process	F	6	0.0218752	5.6E-03
	M	3	0.0310367	

F = Female, M = Male.

Table 2
Molecular functions (GO terms) shared by gender-based MS DEG.

Gene ontology: molecular function	Gender	No. of genes	Corrected *p*-value	Combined *p*-value
GO:0005515: protein binding	F	73	2.142E-13	1.7E-14
	M	21	0.0022441	
GO:0003677: DNA binding	F	16	0.0056011	3.0E-04
	M	8	0.0053933	
GO:0046872: metal ion binding	F	20	0.0464598	9.0E-04
	M	14	0.0019222	
GO:0003723: RNA binding & GO:0005515: protein binding	F	5	0.0206388	9.0E-04
	M	4	0.0039642	
GO:0008270: zinc ion binding	F	13	0.0456205	1.0E-03
	M	9	0.0021955	
GO:0003723: RNA binding	F	9	0.0486291	3.1E-03
	M	5	0.0069794	

F = Female, M = Male.

and *positive regulation of transcription from RNA polymerase II promoter* were the most significant, while *regulation of transcription-DNA dependent* was enriched with 33 genes and *transcription* with 29 genes, and were the GO terms defined by the highest numbers of interactors (Supplementary file INTERACTORS, sheet 2). Furthermore, 49 categories contained ontology terms related to transcription and 3 to chromatin modification. Similarly, 130 categories defined molecular functions for the shared interactors, with *protein binding*, *transcription factor activity* and *transcription factor binding* as the most significantly enriched terms (Supplementary file INTERACTORS, sheet 3). Among the 130 categories, 40 groups were related to transcription and 29 to DNA binding. As highlighted in Fig. 3, a remarkable part of the DEG sharing the interactome (72/89 female and 27/35 male DEG, green nodes in upper and lower left circles) were involved in epigenetic processes and contacted 62/171 common interactors related to transcription and chromatin modification (green nodes in right circles). Additional epigenetic factors appeared in the gender-specific interactomes, as DNMT1 and HGMA2 among the female and male specific interactors respectively. When considering the common interactome, many of the 62 epigenetic interactors (green nodes in Fig. 4A) were connected with more than one female and male DEG (pink and blue nodes respectively). The most connected interactor was HNF4A (with 36 female and 12 male DEG), followed by TP53 (11 female and 2 male DEG), IKBKG (7 female and 2 male DEG) and RELA (5 female and 4 male DEG). Finally, we verified a further network level by reconstructing the interactions among the 62 epigenetic interactors. Interestingly, with two exceptions (KLF5 and TSC22D4), all the other interactors were found to be involved in a complex network, where some nodes, such as histone deacetylase (HDAC)1, HDAC2, RELA, TP53, SP1 and AR, were highly interactive (Fig. 4B, Supplementary file INTERACTORS, sheet 5). Interestingly, most of the epigenetic factors fell in three main groups related to the HDAC, AR or NF-kappaB complexes. These groups could interact each other either directly or through a central group of transcription factors.

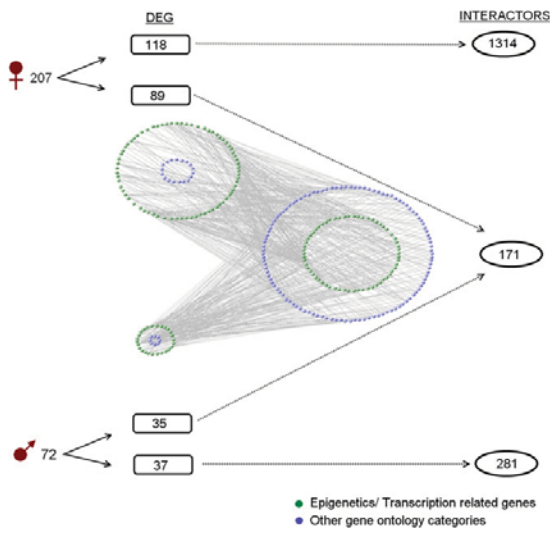


Fig. 3. Gender-related MS interactomes identify shared interactors among male and female MS DEG. Female and male DEG have unique and shared interacting genes. Network reconstruction identified 171 shared interactors contacted by 89/207 female and 35/72 male DEG. Green nodes highlight epigenetic and transcription related genes among the DEG and the interactors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Overall, the distinct MS female and male genesets shared biological and molecular functions as well as interactors. Themes related to epigenetics were predominant in both female and male MS signatures and in the shared interactome, suggesting they might give shape to and/or sustain pathogenic processes in multiple sclerosis.

3.5. Identification of the transcription factor SP1 as a modulator of experimental autoimmune encephalomyelitis

Finally, we checked whether the two MS genesets were potentially regulated by common transcription factors by querying the GeneCodis database. As shown in Table 3, both female and male MS signatures were significantly enriched for genes containing binding sites for four transcription factors (SP1, LEF1, NFY and ELK1) and their combinations. In particular, several DEG contained binding sites for SP1 alone or together with NFY and ELK1 (Table 3). We applied the same analysis tools to two published signatures for the RR-MS population and found that the binding sites for these transcription factors were enriched also among the genes described in other studies (Table 3). As SP1 appeared among the epigenetic interactors connected with both female and male DEG (Fig. 4A) and belonged to the central core of highly connected epigenetic factors (Fig. 4B), we hypothesized that SP1 dependent transcription could play a role in MS pathogenesis.

To validate the involvement of SP1 in immune responses, we initially targeted SP1 dependent gene transcription in peripheral blood mononuclear cells with the specific inhibitor WP631 and induced T cell proliferation with the mitogen concanavalin A (ConA). Interestingly, WP631 concentrations above 0.5 μM strongly blocked proliferation of human T cells (Fig. 5A). Similar results were obtained with ConA-activated mouse splenocytes (Fig. 5A). As WP631 belongs to the family of anthracylines, cytotoxic agents used in cancer therapy, we checked WP631-dependent cytotoxicity in PBMC and splenocyte cultures by flow cytometry. As depicted in Fig. 5B, the drug increased the levels of cell death both in unstimulated and in ConA- stimulated cultures, indicating that part of the reduction in the proliferation was due to cytotoxicity. Finally, WP631 was administered *in vivo* to C57BL6 mice immunized with the encephalitogenic MOG_{35-55} peptide, and T cell responses to the autoantigen were tested ex vivo. Mice receiving the drug (3 mg/kg/day) or its vehicle from day 3 to day 7 post-immunization were sacrificed at day 10 and lymphoid organs were extracted. Interestingly, T cell proliferation to MOG peptide was significantly reduced in WP631-treated animals (Fig. 5C). Moreover, clinical expression of experimental autoimmune encephalomyelitis (EAE), the animal model of MS, was significantly altered by WP631 treatment. In fact, compared to the control group, in the WP631-treated group the incidence of disease was reduced, the onset was delayed and the disease severity in the acute phase was milder (Fig. 5D–E). Similar results were obtained in the EAE model in SJL mice (Fig. 5F).

4. Discussion

4.1. Relevance of a gender-based approach in disease transcriptomics

MS is widely regarded as an erratic disease in terms of clinical symptoms, disease course, gender prevalence and therapeutics [1–3,8]. An optimal transcriptomics study in MS might be the one that attempts to reduce these variable parameters through implementing appropriate inclusion criteria for the recruitment of patients. Even though several past blood gene profiling studies in MS were successful in giving molecular insights [16–23], many of

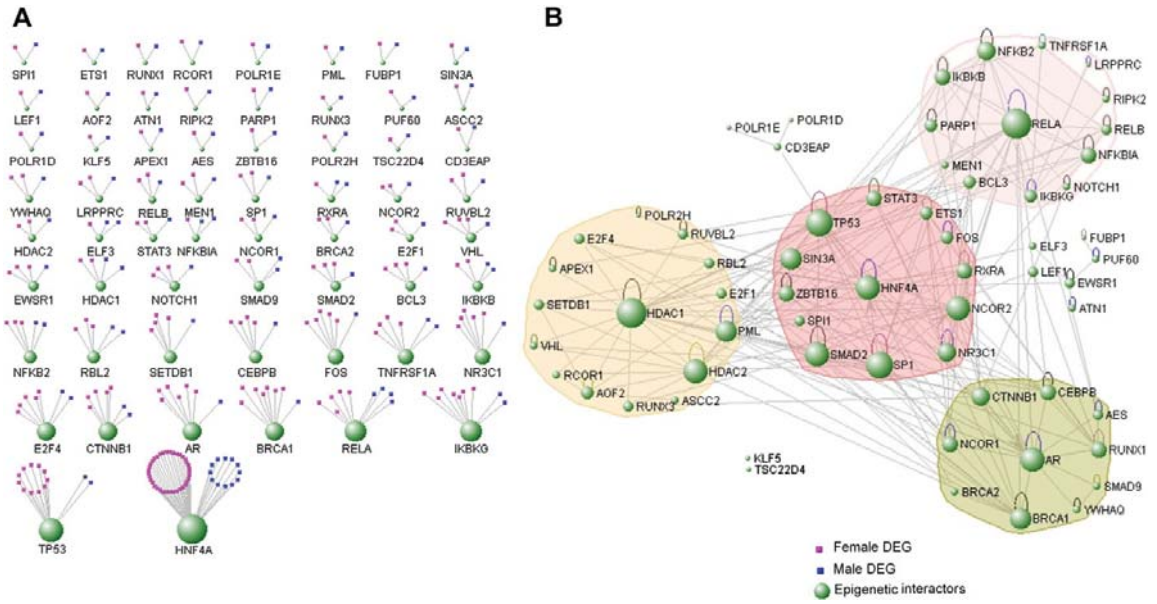


Fig. 4. Epigenetic interactors shared by female and male MS DEG. A. Shared epigenetic interactors (green) of female (pink) and male (blue) MS DEG. B. Interaction network among the shared epigenetic interactors. Most of the epigenetic factors fall in three main groups related to the HDAC, AR or NF-kappaB complexes. These groups can interact each other either directly or through a central group of transcription factors. In A and B the sizes of the interactor nodes are proportional to the number of interactions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

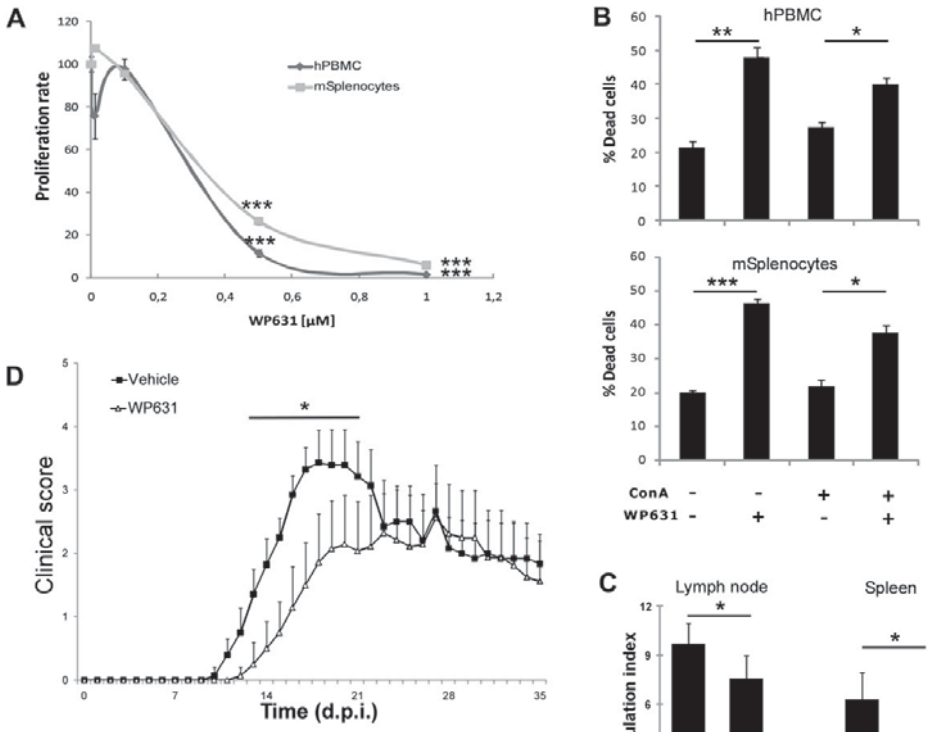
them did not effectively restrict the diverse parameters of the disease. For instance, Bomprezzi and colleagues [17] through PBMC transcriptomics in MS and healthy controls demonstrated that a few candidate genes could accurately classify MS patients and healthy controls. Unfortunately, out of the 24 MS patients used in the study, 6 were in the secondary progressive course of the disease and 18 were RR-MS, and the active or stable phase of disease was not specified. In addition, in this and in other studies [16,17,19,20] the analyses were not unbiased but restricted to a selection of potentially interesting genes. In a high-throughput screening study on PBMC [18], Achiron and coworkers identified molecular signatures of multiple sclerosis that could discriminate relapsing and remission phase of RR-MS. However, a significant group of patients was on immunomodulatory therapy, which may strongly influence global gene expression. Critically considering these issues, we recruited patients with the relapsing-remitting course of the disease, all were clinically stable, free from immunomodulatory therapy and from other inflammatory or autoimmune disorders. Moreover, the standardized Illumina gene expression platform containing more than 22,000 probes was used for all the hybridizations.

The study started with a conventional analysis in which RR-MS samples were compared to healthy controls. In order to identify the DEG in MS, we used three statistical methods with moderate stringencies and subsequently selected the genes which were common to any two of the three tests. Hierarchical clustering of samples based on 167 selected DEG showed satisfactory classification, but still considerable heterogeneity among subjects. Consequently, we hypothesized that the observed heterogeneity might be a consequence of gender differences in the MS and healthy populations. Although factors contributing to gender prevalence of human diseases have been widely addressed in the

Table 3
Transcription factors regulating MS DEG.

Transcription factors-TF		Female		Male			Gandhi et al.		Riveros et al.	
Unique TF	Multiple TF (if any)	DEG	*p*-val[a]	DEG	*p*-val[a]	Combined *p*-val[b]	DEG	*p*-val[a]	DEG	*p*-val[a]
V$SP1_Q6	V$SP1_Q6	33	1.9E-05	10	1.7E-02	5.1E-06	37	4.7E-05	41	1.0E-06
	V$SP1_Q6 & V$NFY_Q6_01	6	2.7E-02	3	4.6E-02	9.6E-03	7	3.0E-02	*n.d.*	*n.d.*
	V$SP1_Q6 & V$NFKB_Q6	3	1.5E-02	3	4.1E-03	6.7E-04	*n.d.*	*n.d.*	*n.d.*	*n.d.*
	V$SP1_Q6 & V$ELK1_02	8	3.4E-03	5	3.3E-03	1.4E-04	*n.d.*	*n.d.*	*n.d.*	*n.d.*
V$LEF1_Q2	V$LEF1_Q2	22	8.3E-03	11	6.9E-03	6.1E-04	*n.d.*	*n.d.*	23	2.7E-02
	V$LEF1_Q2 & V$NFY_Q6_01	5	3.4E-02	4	7.6E-03	2.4E-03	*n.d.*	*n.d.*	*n.d.*	*n.d.*
	V$LEF1_Q2 & V$AR_Q6	3	6.3E-03	3	3.7E-03	2.8E-04	3	3.1E-02	3	2.8E-02
	V$LEF1_Q2 & V$TATA_01	6	3.1E-02	3	5.0E-02	1.1E-02	*n.d.*	*n.d.*	*n.d.*	*n.d.*
V$NFY_Q6_01	V$NFY_Q6_01	11	2.1E-02	8	5.4E-03	1.2E-03	12	4.8E-02	17	1.7E-03
V$ELK1_02	V$ELK1_02	24	2.9E-08	7	6.4E-03	4.2E-09	37	6.9E-16	18	4.9E-04

[a] Corrected *p*-values by Benjamini-Hochberg's method.
[b] Combined *p*-values using Fisher's method.



E

MOG_{35-55} EAE	Vehicle	WP631
Number of animals	14	14
Incidence (%)	100	71
Day of disease onset	12.1 ± 0.7	15.1 ± 1***
Maximal clinical score	3.7 ± 0.4	2.4 ± 0.8*
Score at peak	3.4 ± 0.4	2.1 ± 0.7*
Cumulative disease score (d0-26)	38.1 ± 4.4	23.2 ± 8.1**

F

$PLP_{139-151}$ EAE	Vehicle	WP631 1.5 mg/kg	WP631 3 mg/kg
Number of animals	10	10	10
Incidence (%)	100	100	80
Mortality (%)	30	0	0
Day of disease onset	11.1 ± 0.5	12.9 ± 0.4*	13.3 ± 1
Maximal clinical score	3.8 ± 0.3	2.9 ± 0.3*	2.4 ± 0.5*
Cumulative disease score (d0-37)	53.9 ± 7.8	26.4 ± 5.2**	30.4 ± 7.1*

Fig. 5. Inhibition of SP1 dependent transcription modulates EAE. A. WP631 blocks ConA-induced T cell proliferation. Data are given as percentage of proliferation in absence of WP631 (100%). hPBMC = human PBMC, mSplenocytes = mouse splenocytes. Data are shown as mean ± SEM of 5 independent experiments. B. WP631 induces immune cell death. WP631 concentration was 1 μM. Data are shown as mean ± SEM of at least 4 independent experiments. C. WP631 administration in MOG_{35-55} immunized animals reduces T cell responses to the encephalitogenic peptide. Proliferation to MOG_{35-55} peptide of draining lymph node (left) or spleen (right) cells from WP631 or vehicle-treated EAE mice ($n = 5$ per group). Data are given as average ± SEM. Similar observations were obtained in a second experiment. D–E. WP631 administration *in vivo* attenuates MOG_{35-55} (D–E) and $PLP_{139-151}$ (F) induced EAE. Clinical score chart (D) and clinical parameters (E) are given for MOG_{35-55} induced EAE ($n = 14$ mice per group). WP631 was administered at a dose of 3 mg/kg/day. Mean values ± SEM of 2 pooled EAE experiments are shown. F. Clinical parameters relative to $PLP_{139-151}$ induced EAE in mice treated with two distinct doses of WP631. *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001.

literature, disease transcriptomics with a gender perspective is still in infancy. A first effort in this direction is represented by a recent MS transcriptome study, where gene expression changes in blood were associated with clinical parameters that were diverse among male and female patients [24]. However, this study did not include a healthy reference population, therefore disease specificity of the observations remains questionable.

We assumed that a gender-based analysis of the transcriptome data could considerably reduce the heterogeneity seen in the heatmap. The analysis of the coefficient of variation among selected probes from the couple and gender-based datasets strongly supported our hypothesis as significant reduction in variability both in the healthy and in the MS genelists was observed. Therefore, we subdivided the RR-MS and control samples into male and female

groups. Through the same statistical tests and parameters used for the couple analysis, we identified 207 and 72 genes defining the female or male genesets respectively. When comparing the gender-based to the couple signatures, only about 20% of female or male DEG appeared in the couple geneset. Further, only 43/167 (25%) and 16/167 (~10%) of the couple genes appeared in female and male signatures respectively. This is due to the fact that the conventional couple strategy can single out only those genes that are a compromise among the female and male datasets, while losing gender-specific information. The unsupervised clustering of samples showed that gender-based signatures reduced misclassification. Further, they classified a second independent MS population, demonstrating their reliability.

In conclusion, these results not only affirm the importance of a gender-based transcriptomics approach in MS but also advocate its applicability in other human disease transcriptome studies.

4.2. Sexual dimorphism in MS

The application of the stringent statistics on the female and male populations led to the identification of a lower number of DEG in the male group, suggesting that the pathological processes measured in the periphery are sustained by fewer transcriptional changes in men than in women regardless the same clinical expression of disease. Likewise, brain aging has been associated with sexual dimorphism in terms of different numbers of gene expression changes in the two genders [25]. Moreover, our gender-based analysis illustrated that the DEG in MS women and men were dissimilar not only in quantity but also in contents. In fact, compared to 207 significant genes in female geneset, the male counterpart comprised of a nearly distinct set of 72 genes. Similarly, despite the great number of dysregulated genes both in female and male Parkinson's tissue, only very few changes were common to both [26]. Conversely, about 20% DEG resulted common to female and male datasets in a PBMC gene profiling study in Alzheimer's disease [27]. Anyway, such observations emphasize that gene expression dysregulation in human disorders may be strongly biased by gender. For this reason, we investigated further whether sexual dimorphism in MS was driven by sex-specific genes. Sex-related gene expression differences are normally present in several tissues, including blood [28,29], but they have never been specifically monitored under disease. Using relaxed statistics to maximize the number of differences existing in blood between genders, about 10% of the filtered genes resulted sex-specific in the healthy population, while only 2% in MS subjects. Moreover, only a minor part of the MS sex-specific genes were natural, pre-existing sex-specific genes. These data clearly demonstrate that MS pathology is associated with dysregulation in sex-specific genes. The factors leading to physiological gender differences are complex and include genetic, hormonal and environmental stimuli. Theoretically, if a disease is associated with alterations in these same components, then the expression of the sex-specific genes may result dysregulated. Surely, sex makes a difference in MS susceptibility and clinical course, and sex hormones may modulate disease [30]. However, the contribution of sex-related factors to the same clinical course in men and women remains still elusive. Certainly, our observations cannot be allocated to specific etiological factors. Differently from genetic epidemiology studies, whose task is to isolate the single factors contributing to disease susceptibility, the principal aim of transcriptome studies is to offer a wide picture of the events conserved during disease despite the causes may differ in distinct human subjects. However, our data show for the first time that the genes accountable for the remission state of MS are different in men and women and that part of this diversity is driven by the sex-specific genes.

4.3. Functional annotations, network analysis and the common epigenetics theme

We determined the biological themes involving the DEG in MS women and men by analyzing independently each genelist for gene ontology enrichments. Surprisingly, the dissimilar genesets were wrapped in common biological contexts and this phenomenon was associated with shared molecular functions. More interestingly, the common biological themes were dominated by ontology terms related to transcription, DNA binding and chromatin modification, strongly indicating that epigenetics may be the underlying pathogenic mechanism in multiple sclerosis. Further, we used system biology tools to explore the genes interacting with the DEG. We reconstructed the global interactome relative to the female and male genelists and found 171 distinct genes interacting with both the female and male DEG. Again, the main biological issues related to the common interactome dealt with transcription and chromatin modification. Interestingly, part of these interactors served as convergence points for several distinct gene expression changes as they were contacted by numerous male and female DEG, suggesting that, irrespective of gender, activation of such molecules may be fundamental in sustaining MS in the periphery during the remission state of the disease. The term epigenetics refers to the heritable changes in gene expression that are not due to modifications in DNA sequence [31,32]. Environmental factors may alter physiological epigenetic homeostasis and lead to aberrant gene expression. For example, in multiple sclerosis the low disease concordance in homozygotic twins highlights the importance of environmentally induced epigenetic changes. Among environmental triggers that have been associated with MS, Epstein–Barr virus (EBV) strongly relies on host epigenetic processes to establish infection (reviewed by [33]). Overall, while it is clear that the environment has an influence on disease expression, little is known about how it specifically alters epigenetic control in multiple sclerosis. Our results provide evidence that genes related to epigenetics have altered expression both in the female and male MS populations, although several distinct triggers may have played a role in disease induction in each individual. The observation that a common epigenetic theme exists is of higher significance as it resulted from two distinct sets of genes derived through independent datasets and analyses. Interestingly, the epigenetic interactors were highly connected one with the other and participated to three main complexes, the HDAC, AR or NF-kappaB complexes. It is well known that the transcription factor NF-kappaB has a central role in several biological processes, including inflammation, apoptosis and autoimmunity [34–36]. Histone deacetylases (HDAC) regulate chromatin tightness, as removal of acetyl groups from lysine residues on histone tails promotes chromatin condensation and thereby transcription blockade. Our systems biology model suggests that several DEG may take contact with HDAC proteins or related interactors, however the outcome of all these possible interactions on chromatin accessibility remains unpredictable. Interestingly, the epigenetic drug trichostatin A, a histone deacetylase inhibitor, was effective in reducing neuroinflammation in the animal model of multiple sclerosis [37], indicating that disease is relying on chromatin condensation and that support of transcriptional processes by opening chromatin structure may be beneficial. Notably, also NF-kappaB and AR may be targets of trichostatin A epigenetic action [38,39].

The androgen receptor AR is a cytoplasmic receptor for androgenic hormones, such as testosterone, and moves into the nucleus upon ligand binding where it modulates gene expression. AR has been linked to immune functions, including tolerance induction, as studies in appropriate animal models showed that lack of AR in B cells causes B cell expansion due to apoptosis resistance and

susceptibility to autoimmunity [40]. Interestingly, MS women tend to have less seric testosterone than healthy women, and these levels correlate with brain damage [41]. These observations seem to indicate a link between MS and impairment in AR-mediated functions. Furthermore, treatment of MS men with testosterone improved cognitive functions and slowed brain atrophy [42], demonstrating that AR activation may be beneficial. Moreover, it is intriguing to note on the one hand that several female and male MS DEG may take contact with AR which is fundamental for the development and maintenance of male sexual phenotype, and on the other hand that MS is associated with changes in the pool of sex-specific genes. Unfortunately, no information is available on the molecular networks sustaining gender differences during adulthood, so our data support a role for sex-related factors in MS, whose implications remain at the moment obscure. Clearly, different experimental approaches are required to understand the role of each factor in peripheral events occurring in MS.

4.4. The transcription factor SP1, a novel candidate for immunomodulation

A recently published study proposes that differential expression in MS might be sustained by a network of regulatory transcription factors [43]. They found that using the TRANSFAC database several transcription factor binding motifs were overrepresented in the transcriptional signatures associated with MS. We analyzed this issue by querying the GeneCodis database, in which the annotations for human transcription factors is derived by the systematic catalog described in [44]. This analysis unraveled that several genes in the female and male MS signatures could be regulated by common transcription factors. We focused the attention on SP1 which appeared also among our epigenetic interactors and belonged to the central core of highly connected epigenetic factors. In fact, it is implicated in chromatin remodeling [45] and may form transcriptional complexes with HDAC, NF-kappaB and AR [46,47]. SP1 was not one of the transcription factors described in [43], as TRANSFAC lacks high quality matrices to detect SP1 binding sites [48]. SP1 is the prototype member of a family of related transcription factors that recognize G/C rich tracts in DNA. Importantly, mitoxantrone, one of the approved drugs for the treatment of multiple sclerosis, is a type II topoisomerase inhibitor with intercalating properties at GC-rich sequences including SP1 binding sites [49], raising the hypothesis that SP1 targeting in multiple sclerosis could be beneficial. The drug WP631 is a specific inhibitor of SP1 dependent transcription [50,51]. It belongs to the family of anthracyclines, antineoplastic compounds with potent cytotoxic effects after DNA intercalation. Compared to other anthracyclines, WP631 bisintercalates into DNA with a binding affinity close to that of transcription factors [52–54]. Treatment of Jurkat T lymphocytes with WP631 induces cell cycle arrest and death [51]. Consistently, we observed cell death and blockade of T cell proliferation in cultures of human PBMC and mouse splenocytes exposed *in vitro* to WP631. Moreover, *in vivo* administration of the drug reduced T cell responses to the encephalitogenic MOG peptide and significantly reduced EAE incidence and clinical expression, demonstrating that SP1 dependent transcription modulates autoimmune responses and that its blockade may represent a novel target for MS treatment.

Overall, it is apparent that transcriptomics studies may deliver important information on disease-associated functional processes beyond lists of genes. Here, we provide novel evidences that the "dissection" of MS through gender-based transcriptomics i) leads to more reproducible observations in human populations, ii) uncovers distinct changes in gene expression in diseased women and men but shared biological processes, iii) emphasizes epigenetic events as common basis for relapsing-remitting MS, and iv) may lead to the identification of novel targets and compounds for MS therapy. In light of the results, we propose gender-based systems biology for the analysis of human disease transcriptomics.

Funding

This study was supported by Fondazione Cariplo [Project 2006.0721/10.4898 to C. F.] and by Italian Ministry for Health [Ricerca Finalizzata RFPS-2007-1-637146 to C. F.]. M.D.D. was supported by a training FISM (Fondazione Italiana Sclerosi Multipla) fellowship [Cod. 2009/B/2]. Funding sources had no role in study design, data analysis and manuscript preparation.

Acknowledgments

We thank all the patients and healthy volunteers that donated blood for this study, and Daniela Cantarella at IRCC for the technical support with microarray experiments. We are also grateful to Gianvito Martino and Giancarlo Comi at INSpe for critical discussion and support. The authors declare no conflict of interest.

Appendix. Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jaut.2011.11.004.

References

[1] Zuvich RL, McCauley JL, Pericak-Vance MA, Haines JL. Genetics and pathogenesis of multiple sclerosis. Semin Immunol 2009;21:328–33.
[2] Weiner HL. The challenge of multiple sclerosis: how do we cure a chronic heterogeneous disease? Ann Neurol 2009;65:239–48.
[3] Tremlett H, Zhao Y, Devonshire V. UBC Neurologists. Natural history comparisons of primary and secondary progressive multiple sclerosis reveals differences and similarities. J Neurol 2009;256:374–81.
[4] Bielekova B, Martin R. Development of biomarkers in multiple sclerosis. Brain 2004;127:1463–78.
[5] Comabella M, Martin R. Genomics in multiple sclerosis–current state and future directions. J Neuroimmunol 2007;187:1–8.
[6] Sawcer S, Ban M, Maranian M, Yeo TW, Compston A, Kirby A, et al. A high-density screen for linkage in multiple sclerosis. Am J Hum Genet 2005;77:454–67.
[7] International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med 2007;357:851–62.
[8] Wiendl H, Hohlfeld R. Multiple sclerosis therapeutics: unexpected outcomes clouding undisputed successes. Neurology 2009;72:1008–15.
[9] McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 2001;50:121–7.
[10] Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, et al. TM4 microarray software suite. Methods Enzymol 2006;411:134–93.
[11] Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci USA 2001;98:31–6.
[12] Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001;98:5116–21.
[13] Sebastiani P, Xie H, Ramoni MF. Bayesian analysis of comparative microarray experiments by model averaging. Bayesian Anal 2006;1:707–32.
[14] Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol 2007;8:R3.
[15] Hu Z, Snitkin ES, DeLisi C. VisANT: an integrative framework for networks in systems biology. Brief Bioinform 2008;9:317–25.
[16] Ramanathan M, Weinstock-Guttman B, Nguyen LT, Badgett D, Miller C, Patrick K, et al. In vivo gene expression revealed by cDNA arrays: the pattern in relapsing-remitting multiple sclerosis patients compared with normal subjects. J Neuroimmunol 2001;116:213–9.
[17] Bomprezzi R, Ringner M, Kim S, Bittner ML, Khan J, Chen Y, et al. Gene expression profile in multiple sclerosis patients and healthy controls: identifying pathways relevant to disease. Hum Mol Genet 2003;12:2191–9.
[18] Achiron A, Gurevich M, Friedman N, Kaminski N, Mandel M. Blood transcriptional signatures of multiple sclerosis: unique gene expression of disease activity. Ann Neurol 2004;55:410–7.

[19] Iglesias AH, Camelo S, Hwang D, Villanueva R, Stephanopoulos G, Dangond F. Microarray detection of E2F pathway activation and other targets in multiple sclerosis peripheral blood mononuclear cells. J Neuroimmunol 2004;150:163–77.
[20] Booth DR, Arthur AT, Teutsch SM, Bye C, Rubio J, Armati PJ, et al. Gene expression and genotyping studies implicate the interleukin 7 receptor in the pathogenesis of primary progressive multiple sclerosis. J Mol Med 2005;83:822–30.
[21] Achiron A, Gurevich M, Snir Y, Segal E, Mandel M. Zinc-ion binding and cytokine activity regulation pathways predicts outcome in relapsing-remitting multiple sclerosis. Clin Exp Immunol 2007;149:235–42.
[22] Gurevich M, Tuller T, Rubinstein U, Or-Bach R, Achiron A. Prediction of acute multiple sclerosis relapses by transcription levels of peripheral blood cells. BMC Med Genomics 2009;2:46.
[23] Gandhi KS, McKay FC, Cox M, Riveros C, Armstrong N, Heard RN, et al. The multiple sclerosis whole blood mRNA transcriptome and genetic associations indicate dysregulation of specific T cell pathways in pathogenesis. Hum Mol Genet 2010;19:2134–43.
[24] Achiron AGurevich M. Gender effects in relapsing-remitting multiple sclerosis: correlation between clinical variables and gene expression molecular pathways. J Neurol Sci 2009;286:47–53.
[25] Berchtold NC, Cribbs DH, Coleman PD, Rogers J, Head E, Kim R, et al. Gene expression changes in the course of normal brain aging are sexually dimorphic. Proc Natl Acad Sci USA 2008;105:15605–10.
[26] Simunovic F, Yi M, Wang Y, Stephens R, Sonntag KC. Evidence for gender-specific transcriptional profiles of nigral dopamine neurons in Parkinson disease. PLoS One 2010;5:e8856.
[27] Maes OC, Xu S, Yu B, Chertkow HM, Wang E, Schipper HM. Transcriptional profiling of Alzheimer blood mononuclear cells by microarray. Neurobiol Aging 2007;28:1795–809.
[28] Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci USA 2003;100:1896–901.
[29] Karlovich C, Duchateau-Nguyen G, Johnson A, McLoughlin P, Navarro M, Fleurbaey C, et al. A longitudinal study of gene expression in healthy individuals. BMC Med Genomics 2009;2:33.
[30] Tomassini V, Pozzilli C. Sex hormones, brain damage and clinical course of multiple sclerosis. J Neurol Sci 2009;286:35–9.
[31] Hewagama A, Richardson B. The genetics and epigenetics of autoimmune diseases. J Autoimmun 2009;33:3–11.
[32] Urdinguio RG, Sanchez-Mut JV, Esteller M. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. Lancet Neurol 2009;8:1056–72.
[33] Tao Q, Robertson KD. Stealth technology: how Epstein-Barr virus utilizes DNA methylation to cloak itself from immune detection. Clin Immunol 2003;109:53–63.
[34] Li Q, Verma IM. NF-kappaB regulation in the immune system. Nat Rev Immunol 2002;2:725–34.
[35] Ghosh S, Hayden MS. New regulators of NF-kappaB in inflammation. Nat Rev Immunol 2008;8:837–48.
[36] Yan J, Greer JM. NF-kappa B, a potential therapeutic target for the treatment of multiple sclerosis. CNS Neurol Disord Drug Targets 2008;7:536–57.
[37] Camelo S, Iglesias AH, Hwang D, Due B, Ryu H, Smith K, et al. Transcriptional therapy with the histone deacetylase inhibitor trichostatin A ameliorates experimental autoimmune encephalomyelitis. J Neuroimmunol 2005;164:10–21.
[38] Chen L, Fischle W, Verdin E, Greene WC. Duration of nuclear NF-kappaB action regulated by reversible acetylation. Science 2001;293:1653–7.
[39] Rokhlin OW, Glover RB, Guseva NV, Taghiyev AF, Kohlgraf KG, Cohen MB. Mechanisms of cell death induced by histone deacetylase inhibitors in androgen receptor-positive prostate cancer cells. Mol Cancer Res 2006;4:113–23.
[40] Altuwaijri S, Chuang KH, Lai KP, Lai JJ, Lin HY, Young FM, et al. Susceptibility to autoimmunity and B cell resistance to apoptosis in mice lacking androgen receptor in B cells. Mol Endocrinol 2009;23:444–53.
[41] Tomassini V, Onesti E, Mainero C, Giugni E, Paolillo A, Salvetti M, et al. Sex hormones modulate brain damage in multiple sclerosis: MRI evidence. J Neurol Neurosurg Psychiatry 2005;76:272–5.
[42] Sicotte NL, Giesser BS, Tandon V, Klutch R, Steiner B, Drain AE, et al. Testosterone treatment in multiple sclerosis: a pilot study. Arch Neurol 2007;64:683–8.
[43] Riveros C, Mellor D, Gandhi KS, McKay FC, Cox MB, Berretta R, et al. A transcription factor map as revealed by a genome-wide gene expression analysis of whole-blood mRNA transcriptome in multiple sclerosis. PLoS One 2010;5:e14176.
[44] Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 2005;434:338–45.
[45] Cakouros D, Cockerill PN, Bert AG, Mital R, Roberts DC, Shannon MF. A NF-kappa B/Sp1 region is essential for chromatin remodeling and correct transcription of a human granulocyte-macrophage colony-stimulating factor transgene. J Immunol 2001;167:302–10.
[46] Lu S, Jenster G, Epner DE. Androgen induction of cyclin-dependent kinase inhibitor p21 gene: role of androgen receptor and transcription factor Sp1 complex. Mol Endocrinol 2000;14:753–60.
[47] Liu S, Wu LC, Pang J, Santhanam R, Schwind S, Wu YZ, et al. Sp1/NFkappaB/HDAC/miR-29b regulatory network in KIT-driven myeloid leukemia. Cancer Cell 2010;17:333–47.
[48] Chang JT, Nevins JR. GATHER: a systems approach to interpreting genomic signatures. Bioinformatics 2006;22:2926–33.
[49] Chiang SY, Azizkhan JC, Beerman TA. A comparison of DNA-binding drugs as inhibitors of E2F1- and Sp1-DNA complexes and associated gene expression. Biochemistry 1998;37:3109–15.
[50] Martin B, Vaquero A, Priebe W, Portugal J. Bisanthracycline WP631 inhibits basal and Sp1-activated transcription initiation in vitro. Nucleic Acids Res 1999;27:3402–9.
[51] Mansilla S, Portugal J. Sp1 transcription factor as a target for anthracyclines: effects on gene transcription. Biochimie 2008;90:976–87.
[52] Chaires JB, Leng F, Przewloka T, Fokt I, Ling YH, Perez-Soler R, et al. Structure-based design of a new bisintercalating anthracycline antibiotic. J Med Chem 1997;40:261–6.
[53] Leng F, Priebe W, Chaires JB. Ultratight DNA binding of a new bisintercalating anthracycline antibiotic. Biochemistry 1998;37:1743–53.
[54] Hu GG, Shui X, Leng F, Priebe W, Chaires JB, Williams LD. Structure of a DNA-bisdaunomycin complex. Biochemistry 1997;36:5940–6.

PART III

LIST OF PUBLICATIONS

1. *Gender-based blood transcriptomes and interactomes in multiple sclerosis: involvement of SP1 dependent gene transcription*. **Menon R**, Di Dario M, Cordiglieri C, Musio S, La Mantia L, Milanese C, Di Stefano AL, Crabbio M, Franciotta D, Bergamaschi R, Pedotti R, Medico E, Farina C. J Autoimmun. 2012 May;38(2-3):J144-55.

2. *Shared molecular and functional frameworks among five complex human disorders: a comparative study on interactomes linked to susceptibility genes*. **Menon R**, Farina C. PLoS One. 2011 Apr 21;6(4):e18660.

3. *Developmentally coordinated extrinsic signals drive human pluripotent stem cell differentiation toward authentic DARPP-32+ medium-sized spiny neurons*. Carri AD, Onorati M, Lelos MJ, Castiglioni V, Faedo A, **Menon R**, Camnasio S, Vuono R, Spaiardi P, Talpo F, Toselli M, Martino G, Barker RA, Dunnett SB, Biella G, Cattaneo E.Development. 2013 Jan 15;140(2):301-12.

4. *MiR-30e and miR-181d control radial glia cell proliferation via HtrA1 modulation*. Nigro A, **Menon R**, Bergamaschi A, Clovis YM, Baldi A, Ehrmann M, Comi G, De Pietri Tonelli D, Farina C, Martino G, Muzio L. Cell Death Dis. 2012 Aug 2;3:e360.

5. *Human neurotrophin receptor p75NTR defines differentiation-oriented skeletal muscle precursor cells: implications for muscle regeneration*. Colombo E, Romaggi S, Medico E, **Menon R**, Mora M, Falcone C, Lochmüller H, Confalonieri P, Mantegazza R, Morandi L, Farina C. J Neuropathol Exp Neurol. 2011 Feb;70(2):133-42.

6. *Activated macrophages release shedding vesicles containing polarized M1 or M2 mRNAs*. Garzetti L, **Menon R**, Finardi A, Bergami A, Martino G, Comi G, Farina C, Verderio C, Furlan R. *Under review* - PLoS ONE.

//continued//

7. *iPSC-derived neural precursors exert a neuroprotective role in immune-mediated demyelination by preserving viability of endogenous oligodendrocytes via the secretion of leukemia inhibitory factor.* Laterza C, Merlini A, De Feo D, Ruffini F, Onorati M, **Menon R**, Fredrickx E, Muzio L, Lombardo A, Quattrini A, Taveggia C, Farina C, Cattaneo E, Martino G. *Under review* - Cell Stem Cell.