Identification of Genetic Susceptibility Factors for Fibromyalgia

Identificació de factors de susceptibilitat genètica per a fibromiàlgia

Elisa Docampo Martínez

UNIVERSITAT DE BARCELONA

FACULTAT DE BIOLOGIA

DEPARTAMENT DE GENÈTICA

GENÈTICA MOLECULAR HUMANA

BIENNI 2008-2010

IDENTIFICATION OF GENETIC SUSCEPTIBILITY FACTORS FOR FIBROMYALGIA

Identificació de factors de susceptibilitat genètica per a fibromiàlgia

Memòria presentada per **Elisa Docampo Martínez** per a optar al títol de Doctora per la Universitat de Barcelona

Xavier Estivill i Pallejà

(Director)

Raquel Rabionet Janssen

(Directora)

Daniel-Raúl Grinberg Vaisman

(Tutor)

Tesi realitzada al programa Gens i Malaltia del Centre de Regulació Genòmica (CRG)

ABSTRACT

Fibromyalgia (FM) is a highly disabling syndrome defined by a low pain threshold and a permanent state of pain. Widespread pain is accompanied by a constellation of symptoms such as fatigue, sleep disturbances and cognitive impairment, among others. The mechanisms explaining this chronic pain remain unclear. Nowadays, the most established hypothesis underlying FM ethiopathogenesis is the existence of a dysfunction in pain processing, as supported by alterations in neuroimaging and neurotransmitters levels. The etiology of FM involves the interaction of environmental and genetic susceptibility factors. The genetic contribution to FM has been proven by the presence of a higher concordance in monozygotic than dizygotic twins as well as family aggregation. However, the individual genetic and environmental factors involved have not been identified. The aim of this thesis was to elucidate genetic susceptibility factors for fibromyalgia. We assessed this objective through three main approaches: the identification of FM clinically homogeneous subgroups with a two step cluster analysis, a genome-wide association study in order to evaluate the possible contribution of single nucleotide polymorphisms with Illumina 1 million duo array, and array comparative genomic hybridization experiments to identify regions varying in copy number that could be involved in FM susceptibility. In the cluster study, 48 variables were evaluated in 1,446 Spanish FM cases fulfilling 1990 ACR FM criteria. A partitioning analysis was performed to find groups of variables similar to each other. Variables clustered into three independent dimensions: “symptomatology”, “comorbidities” and “clinical scales”. Only the two first dimensions were considered for the construction of FM subgroups, classifying FM samples into three subgroups: low symptomatology and comorbidities (Cluster 1), high symptomatology and comorbidities (Cluster 2), and high symptomatology but low comorbidities (Cluster 3). These subgroups showed differences in measures of disease severity and were further implemented in genetic analysis. Genome-wide association study was performed in 300 FM cases and 203 controls. No SNP reached GWAS association threshold, but 21 of the most associated SNPs were chosen for replication in over 900 cases and 900 pain free-controls. Four of the strongest associated SNPs selected for replication showed a nominal association in the joint analysis. In particular, rs11127292 (*MYT1L)* was found to be associated to FM with low comorbidities. Array comparative genomic hybridization detected 5 differentially hybridized regions, which were followed up by direct genotyping. Follow up experiments validated association for one of these regions. An intronic deletion in *NRXN3* was associated to female cases of FM and in particular those with low levels of comorbidities. This enhances the importance of gender in FM ethiopathogenesis and could be pointing to the existence of a different genetic background for FM in males and females. It also highlights the importance of identifying FM homogeneous subgroups for the detection of FM genetic susceptibility factors. If the proposed FM candidate genes are further validated in replication studies, this would constitute a change in the FM etiological concept, as several of our proposed candidates are known neuropsychiatric disease associated genes (autism, addiction, mental disability). This would highlight a novel neurocognitive involvement in this disorder, currently considered as a musculoskeletal and affective disease.

CONTENTS

INTRODUCTION

"Pain, like love, is all consuming: when you have it, not much else matters, and there is nothing you can do about it".
Clifford J. Woolf

PAIN: A HOMEOSTATIC EMOTION ESSENTIAL FOR THE SURVIVAL OF SPECIES (CHARLES DARWIN)

Pain definition and classification

Pain is the ability to detect noxious stimuli. It is a physiological protective system essential to an organism's survival: individuals who are incapable to detect painful stimuli do not engage in protective behaviours against life threatening conditions (1).

Based on its function and on the elapsed time from the painful stimulus, three kinds of pain can be considered: nociceptive, inflammatory and pathological pain (2). Nociceptive pain is the early-warning physiological protective system, essential to detect and minimize contact with damaging stimuli. Acute nociceptive pain occurs when powerful or noxious stimuli activate nociceptors, specialized sensory neurons characterized by high activation thresholds (3). This type of pain is not a clinical problem, except in the specific context of surgery and other clinical procedures that necessarily involve noxious stimuli, where it must be surpassed by local and general anaesthetics (4). After tissue damage, inflammatory pain, which is caused by the activation of the immune system, assists in the healing of the injured body part. By creating hypersensitivity, or tenderness, it reduces further risk of damage and promotes recovery. Inflammatory pain is adaptative but it still has to be reduced in patients with severe and/or extensive injury. Finally, there is the pain that is not protective but maladaptative, resulting from an abnormal functioning of the nervous system. Pathological pain is a disease state of the nervous system. It can occur after nervous damage (neuropathic pain), but also in conditions in which there is no such damage (dysfunctional pain). It is the consequence of amplified sensory signals in the central nervous system (CNS). Conditions that evoke dysfunctional pain include fibromyalgia (FM) and irritable bowel syndrome among many other conditions.

Pain circuitry and structures: how it works

Noxious stimuli are detected by nociceptors, whose cell bodies are located in the dorsal root ganglia (body sensibility) and the trigeminal ganglion (face sensibility). They have both a peripheral and central axonal branch that innervate their target organ and the spinal cord, respectively (5). This specialized set of nerve fibers includes unmyelinated C fibers and thinly myelinated Aδ fibers, which are distinct from myelinated tactile sensors (Aβ fibers) and proprioceptors[*1]. The peripheral terminal of the nociceptor will only respond to environmental stimuli (painful hot, cold or mechanical stimulation), when the stimulus intensities reach the noxious range. The physicochemical properties of noxious stimuli, such as heat, extreme cold, pressure and chemicals, are converted to electrical activity by transient receptor potential-generating channels, and

[1]Terms marked with an asterisk are defined in the Glossary section (Annexes).

this electrical activity is amplified by sodium channels to produce action potentials. Nociceptive afferents carrying these peripheral inputs form glutamatergic (excitatory) synapses onto second-order neurons, mostly in the superficial dorsal horn. Sensory inputs are integrated and processed, and the net output from the spinal networks is carried by several pathways to distinct projections in the brain. These include the lateral spinothalamic tract, which projects to the lateral thalamus and has been implicated in sensory-discriminative aspects of the pain experience (such as discriminating where the stimulus is? and, how intense is it?), and the medial aspect of the spinothalamic tract and the spinoparabrachial tract, which project to the medial thalamus and limbic structures and are believed to mediate emotional and aversive components of pain (6). Finally, from these brainstem and thalamic loci, the experience of pain is perceived in the cortex, where several structures, some of them being more associated with the sensory discriminative properties (such as somatosensory cortex) and others with emotional aspects (such as insular cortex) are activated, and information is sent to the spinal cord to enable withdrawal from the noxious stimuli (5).

The pain ascending network is regulated by mechanisms of descending inhibition and descending facilitation (7). These descending pathways originate in the hypothalamus, the cortex, the rostroventral medulla and other brainstem nuclei, and can interact with several neuronal elements in the dorsal horn: the terminals of the primary afferent fibers, projection neurons, intrinsic inhibitory and excitatory interneurons and terminals of other descending pathways (Figure 1). Multiple neurotransmitters are involved, and their co-localization, as well as the presence of multiple receptors per neurotransmitter, allow the regulation of the nociception signal through descending facilitatory and inhibitory pathways (7). Therefore, precise controls must exist to maintain a balance between activation and inhibition (8). Several theories point out that psychological factors are known to modulate pain perception. Variables such as attentional state, emotional context, hypnotic suggestions, attitudes, expectations or anaesthesia-induced changes in consciousness now have been shown to alter both pain perception and forebrain pain transmission in humans (9). This psychological pain modulation is exerted by descending projections from frontal brain areas. It has been proposed that in a negative emotional state, activity in the frontal cortex is increased, enhancing the nociceptive signal transmitted from the spinal cord to the brain by engaging descending facilitatory pathways (10).

As a summary, in order to ensure a very precise regulation of the message, nociceptive transmission is very complex and involves different pathways that connect different CNS structures.

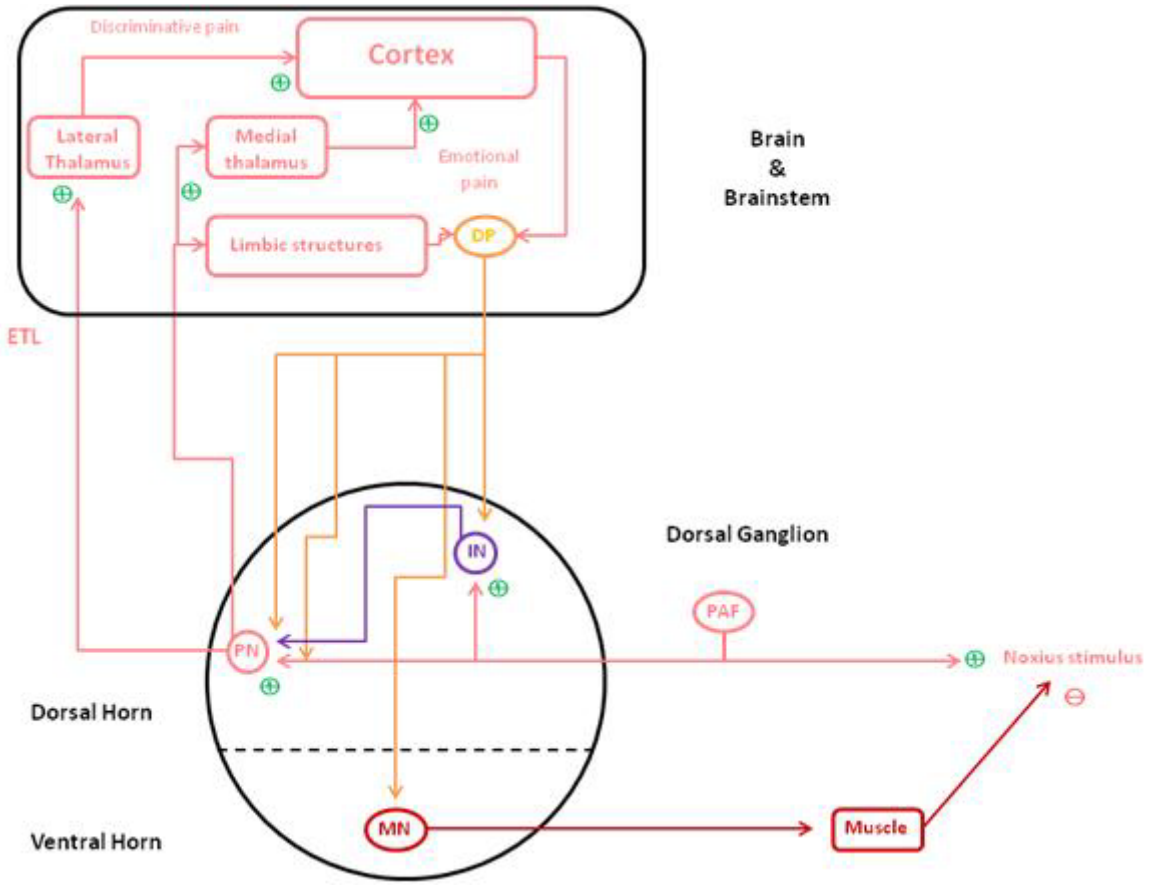


Figure 1: Pain transmission and structures involved (based on Millan *et al* (7). DP, Terminals of descending pain; PAF, afferent fibres; IN, interneurons; MN, motoneurons; PN, projection neurons; and ETL, spinothalamic lateral tract. Synapses that are not marked with a plus or a minus symbols can be both activatory and inhibitory.

Visualizing pain

For many years, the study of nociceptive pain was restricted to the analysis of sensory neurons and circuits in the spinal cord. One main reason was the difficulty to examine how the brain processed pain signals in anaesthetized animals, when the standard definition of adequate anaesthesia is the loss of pain related behaviour. However, functional neuroimaging in human volunteers and patients allowed for definition of those brain areas activated by nociceptive inputs as reviewed in Shweinhardt *et al* (11). The primary modes of functional imaging that have been used in FM include functional magnetic resonance imaging (fMRI), single-photon emission computed tomography (SPECT), and positron emission tomography (PET). More details on these techniques are included in Box 1.

Box 1: Imaging and functional neuroimaging

Neuroimaging methods infer brain activity from regional cerebral blood flow, glucose metabolism and neurochemical concentrations, and changes in brain function from structural measures of water motility and brain volume.

MRI produces high-resolution structural images by placing the patient in a strong magnetic field that aligns all the body atomic nuclei. This alignment is altered by radio frequency fields, which produce a rotating magnetic field in the nuclei. These are detectable by a scanner that constructs an image of the scanned area of the body. Macroscopic changes in brain structure can be evaluated by voxel-based morphometry, which uses MRI images to assess the volume of specific brain regions.

FMRI measures changes in neural activity by measuring vascularization changes. Increased neural activity causes an increased demand for oxygen, which is ensured by the vascular system by increasing the amount of oxygenated haemoglobin. Because deoxygenated haemoglobin attenuates the MR signal, the vascular response leads to a signal increase that is related to the neural activity.

Magnetic resonance spectroscopy (MRS) is used to measure the levels of different metabolites in body tissues. Magnetic resonance (MR) signal varies among the different molecules, and MRS evaluates the signal of a given source and expresses its magnitude in relation to a control standard molecule (often creatinine). Thus, the relative concentration of molecules such as neurotransmitters can be inferred.

SPECT evaluates changes in cerebral blood flow with the infusion of a radioactive tracer. This tracer produces gamma radiation that is detected by a camera (gamma camera) to acquire multiple 2-D images, from multiple angles. A computer is then used to apply a tomographic reconstruction algorithm to create a three-dimension dataset.

In PET, when the radioactive tracer decays, a positron is delivered and, in combination with a tissue electron, it forms a *positronium,* which is unstable and produces two gamma ray photons after an average lifetime of 125 picoseconds. The photons are detected by a camera, creating three-dimensional images. Image resolution of PET is better than that of SPECT, as photons produce more radiation.

Mechanisms of pathological pain

Although physiological pain has an important protective function, pain can take on a disease character in pathological chronic states such as inflammation, neuropathy, cancer, viral infections chemotherapy and diabetes. This state is manifested as hyperalgesia (increased sensitivity to painful stimuli). Furthermore, individuals with chronic pain often show disease-induced, therapy-resistant deviations from normal tactile sensations, such as paresthesias and dysestesias. Finally, the most common complaint from individuals with chronic pain is spontaneous pain (6).

There are evidences, both at the CNS and at the peripheral nervous system level, indicating that these pathological pain states can be explained by an alteration in pain processing. In physiological conditions, there is an enhanced net descending inhibition after inflammation in sites of primary hyperalgesia. However, in models of neuropathic pain, the tactile allodynia (when a non-painful stimulus becomes painful) after nerve injury is dependent upon a tonic activation of net descending facilitation. Thus, descending modulation of persistent pain involves both inhibition and facilitation, in order to improve nociception at the spinal level (12). This involves prolonged functional changes in the nervous system, evidenced by the development of dorsal horn hyperexcitability. An impairment of this endogeneous pain modulation due to an imbalance of the excitatory and inhibitory tracts, known as **pain disinhibition,** can explain the alteration in pain processing occurring In individuals with chronic pain (13).

Permanent stimulation of the nociceptive pathways that are present in chronic pain states can also cause an increase in the magnitude of responses to a defined sensory stimulus, an increase in the level of spontaneous activity, or after discharges, which represent continued activity after the termination of a nociceptive stimulus. This leads to a central amplification of pain also known as **central sensitization** (higher excitability and spontaneous activity) (6). Fibres responsible of the transmission of non-painful stimuli (Aβ fibers) get C fiber (nociceptive) qualities. As a result, touch becomes painful (Figure 2). This pathological enhancement of pain transmission is mediated by functional and structural changes. Functional plasticity includes molecular changes; for example, persistent inflammation with higher and continuous amounts of inflammation mediators can cause a Ca^{2+} ion channel sensitization, leading to a synaptic potentiation that finally causes central sensitization. Also, at the structural level, persistent nociceptive activity leads to an increase in the number and size of synaptic spines, causing changes in neuronal connectivity and cell proliferation (6).

At the peripheral nervous system, continuous stimulation increases the excitability of nociceptors, in a process known as **peripheral sensitization**. This leads to a reduction in the pain threshold and, therefore, to a state of hyperalgesia.

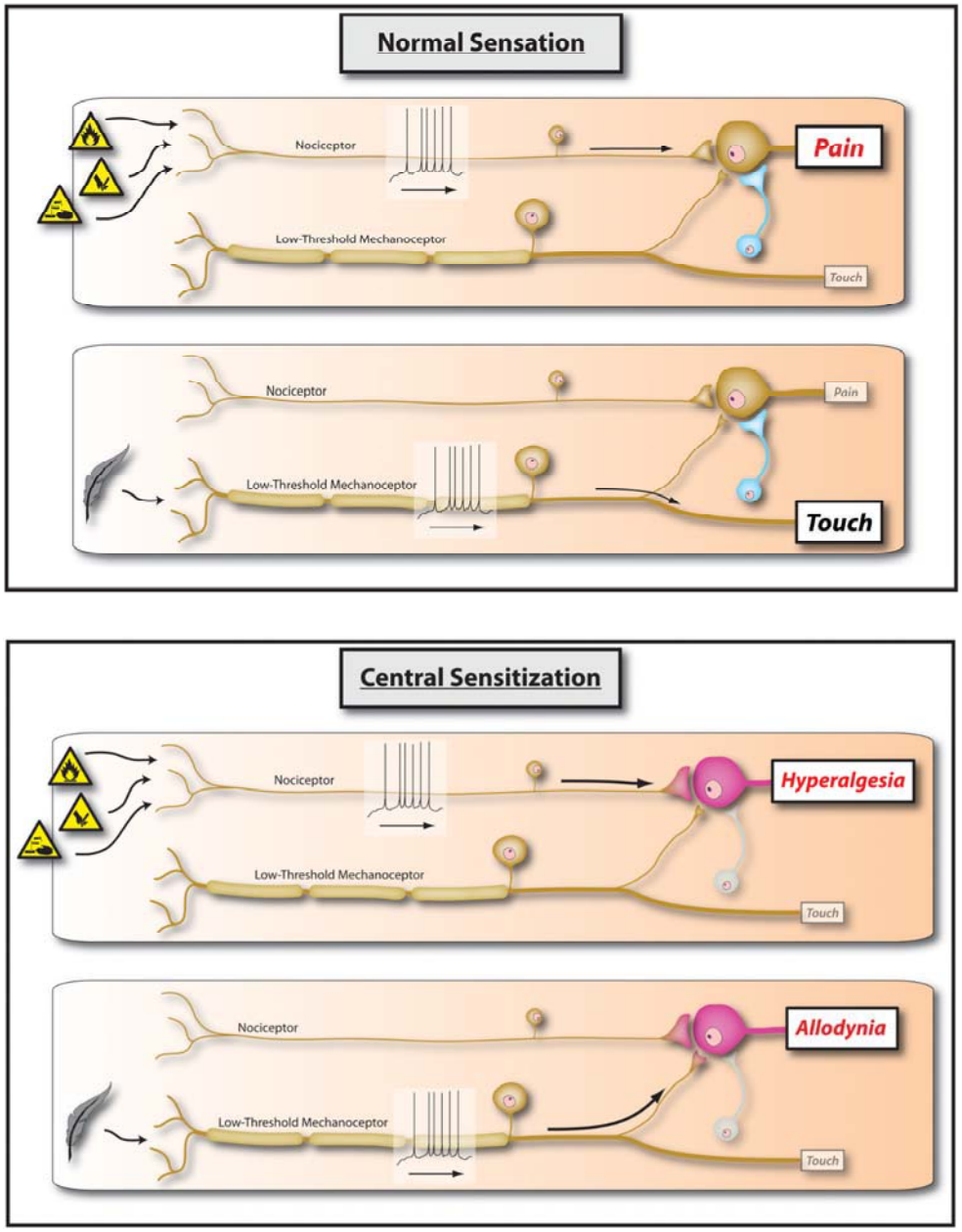


Figure 2: Normal and central sensitization. Normal sensitization (top images): primary sensory neurons that encode low intensity stimuli only activate central pathways leading to innocuous sensations, while high intensity stimuli that activate nociceptors only activate the central pathways that lead to pain. The induction of central sensitization in somatosensory pathways (bottom images) is accompanied by increases in synaptic efficacy and reductions in inhibition, leading to a central amplification. Pain response to noxious stimuli is enhanced, in amplitude, duration and spatial extent (hyperalgesia), while the strengthening of normally ineffective synapses recruits subliminal low threshold sensory inputs that activate the pain circuit (allodynia). The two sensory pathways converge (From Wolf *et al* (14)).

THE SYNAPSE

Definition and structure

Neural information is propagated, along a neuron, as an electrical signal. This transmission is not continuous. The connection from one neuron to the other takes place through the synapses, whose existence was a matter of discussion during the end of the XIXth century (Box 2). Synapses are specialized intercellular junctions whose function is the transfer of information from a neuron to a target cell, usually another neuron. Synapses can be electrical or chemical, depending upon whether transmission occurs via direct propagation of the electrical stimulus in the presynaptic cell or via chemical intermediates. Electrical synapses are gap junctions between neurons; they can be either excitatory or inhibitory (15), allowing bidirectional propagation of the signal and playing a role in synchronizing neuronal activity. At the chemical synapses, the presynaptic electrical signal is converted into neurotransmitters that bind to specific postsynaptic receptors, leading to the production of a new electrical signal in the postsynaptic neuron. We will focus on chemical synapses, and we will refer to them henceforward as synapses (16, 17).

Synapses share many properties with other intercellular junctions, but they present differential features: they are asymmetrical, transmit information by an extremely fast and tightly regulated mechanism, and are highly plastic (16, 17). In spite of their great morphological variability, all synapses share some common structural features. On the presynaptic cell, neurotransmitter-containing synaptic vesicles accumulate in the active zone, where the neurotransmitter release occurs. These vesicles associate with the presynaptic plasma membrane through a network of scaffolding proteins known as the cytomatrix of active zones. The postsynaptic membrane includes an accumulation of neurotransmitter receptors, a thickening of the membrane, the postsynaptic density and a submembranous electron-dense scaffold (16, 18). For synapses to function properly, all these components must be recruited and precisely aligned across the synaptic cleft, a 20 nm wide extracellular space that separates two neurons at synaptic junctions (18). The proper release of neurotransmitters into the synaptic cleft and their binding to the receptors is facilitated by the existence of a complex matrix of adhesion molecules located in both the presynaptic and postsynaptic neurons.

Many families of neural adhesion molecules have been described. These transmembrane molecules bind to each other extracellularly to promote adhesion between the presynaptic and postsynaptic neurons. During development they enable the formation and specification of functional synapses, as they ensure stable contacts between neurons (8, 17, 18). Diversification of adhesion molecules through alternative splicing, among other mechanisms, provides an important repertoire of molecules to enable synapses function in the complexity and specificity of the CNS (19).

Box 2: Continuity versus contiguity: the XIX-XXth debate

The nature of the contact between neural cells was a controversial issue that dominated the thinking of neuroanatomists and physiologists between 1870 and 1920, with two main theories:

- the neuron theory: neurons as independent cellular units
- the reticular theory: central nervous system as a complex syncitium: a network of fibers that are in direct cytoplasmic continuity

To settle the question of continuity versus contiguity it was necessary to demonstrate the final ramifications of the nerve fibers. This was accomplished by Santiago Ramón y Cajal. His work derived in large part from his application of the chrome-silver impregnation method or "reazione nera" that had been introduced by Golgi in 1873. This method offered two advantages. First, the method stained, in an apparently random manner, only about 1% of the cells in any particular region of the brain or spinal cord. This made it possible to study the morphology of individual nerve cells in isolation from their neighbours. The second advantage was that the neurons that were stained were often impregnated throughout their entire extent, so that one could clearly visualize cell bodies, axons, axon collaterals, the full dendritic arbour and, in developing brains, axonal and dendritic growth cones. By examining in detail nerve cells and their contacts in histological sections of almost every brain region, Cajal was able to describe not only differences between various types of nerve cells but also the great variety of axonal endings found in the central nervous system. This led him inexorably to conclude that, at the sites of interaction, they are not continuous with their cellular targets, and therefore not part of a diffuse network. Cajal's work, published in Spanish, was not widely known or appreciated until 1889, when he attended a meeting of the German Anatomical Society in Berlin and attracted the attention of Kölliker, who encouraged him to have his work translated into French or German. But it was a physiologist, Charles Sherrington, who coined the term synapse in 1897, from the Greek "syn-" ("together") and "haptein" ("to clasp").

> Once the morphological issue of how the nerve cells interact had been resolved, attention naturally turned toward understanding the mechanism of synaptic transmission: was it electrical or was it chemical …

Extracted from *Synapses, Chapter 1 A brief history of synapses and synaptic transmission. 2003 Cowan W.M., Südhof T.C. and Stevens C.F.*

Neurexins and neuroligins

Neurexins are among the most widely studied adhesion and scaffolding molecules involved in synapse stability and function, and were discovered thanks to spider venom (Box 3). They are transmembrane proteins located in the presynaptic neuron that have three extracellular binding partners: neuroligins, dytroglycan and neurexophilins (20). In particular, the binding with neuroligins has shown to be essential for the development and function of GABAergic and glutamatergic synapses (17, 21). They are mainly expressed in the CNS, and their distribution across the different brain regions is heterogeneous (22).

In humans, there are three neurexin proteins encoded by three genes. Each gene has two independent promoters and generates two classes of transcripts, which give rise to two classes of proteins with different length and domain composition: α and β-neurexins. α and β-neurexins share the transmembrane and

intracellular regions. The extracellular sequences of the longer α isoforms contain six laminin-like domains (LNS) with three intercalated epidermal growth factor (EGF)-like domains, whereas β isoforms only contain the last LNS following a short N-terminal β-specific sequence (18, 23). Neurexins are essential for survival, but redundant: knockout (KO) mice for the three genes were alive at birth but died the first day, double KO mice died within the first week, and single KO mice exhibited impaired survival (24). α-neurexins are required for normal neurotransmitter release: their deletion impairs the function of Ca^{2+} channels, in excitatory and inhibitory synapses (24). α-neurexins also affect transmission at the neuromuscular junction (25). Synapse-inducing activity has also been demonstrated for β-neurexins.

Neurexins bind to a family of postsynaptic transmembrane proteins, the neuroligins. The Neurexin/neuroligin cell-adhesion complex can promote the formation of *de novo* synapses and the differentiation of postsynaptic receptors, at least *in vitro*. The binding of neurexins to neuroligins is controlled by alternative splicing (18). There are five neuroligins (1-4 and Y) in humans. Neuroligin-1 was identified as a result of its ability to bind certain isoforms of all three β-neurexins (26) (Figure 3). α-Neurexins have five canonical sites of alternative splicing, and there is evidence suggesting that the different splice sites are used independently, potentially generating more than 1000 different isoforms. The six neurexin (α and β of Neurexin 1, 2, and 3) isoforms are expressed in all brain regions, but the relative abundance of each isoform is in some extent tissue specific (22). Two different neurexins can be expressed in the same cell (27).

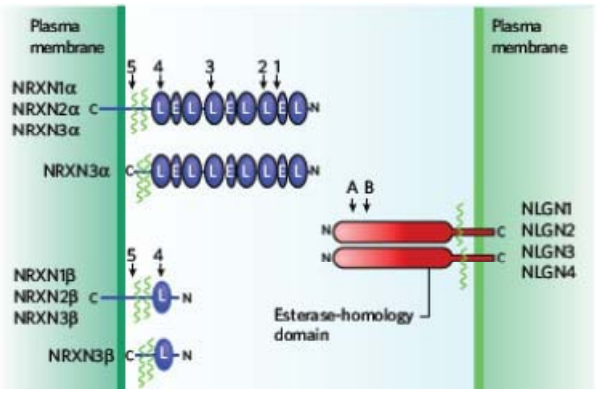


Figure 3: Neurexins and neuroligins in the synaptic cleft (taken from Sudhof *et al* (17)). The left part of the figure represents the presynaptic neuron, where α and β neurexins (NRXN) bind to postsynaptic neuroligins (NLGN). Their binding is regulated by alternative splicing (AS): NRXNs have five (2 for shorter β forms) (1-5) and NLGN two (A and B) canonical sites of AS. (L, LNs domains; E, EGF-like domains; and CHO, carbohydrate attachment sequence).

Recent studies have identified mutations in the genes encoding NRXNs and neuroligins (NLGN) as a cause of schizophrenia (28), autism (29), and Tourette's syndrome (30). These synaptic molecules participate in key circuits influencing addictive behaviors (22, 31). Two members of the neurexin gene family have been associated with nicotine dependence, Neurexin 1 and Neurexin 3 (5, 32). Moreover, Neurexin 3 has also

been associated with smoking behaviour in schizophrenic patients (33), impulsivity and substance abuse (34), illegal substance abuse (35), alcohol dependence (36) and obesity (37). In addition to single nucleotide variants (SNPs), large and rare structural variants involving neurexin genes have been described in autism spectrum disorders (38-42) and schizophrenia patients (43).

Box 3: Neurexins and spiders (27)

α-Latrotoxin, a component of black widow spider venom, triggers massive neurotransmitter release from the presynaptic nerve terminals of vertebrates, binding to specific receptors in the presynaptic plasma membrane by a mechanism that was unknown. Extensive peptide sequences from the purified bovine α-latrotoxin receptor were used to synthesize degenerate oligonucleotides, which were then used in polymerase chain reaction (PCR) experiments to clone the cDNA sequences encoding these peptides. The PCR products and oligonucleotides were then used as probes to screen rat brain cDNA libraries. Two sets of overlapping cDNAs that encoded homologous but distinct proteins were isolated, only one of which contained the peptide sequences obtained from the isolated α-latrotoxin receptor. Each set of cDNA clones was highly polymorphic. The proteins encoded by these transcripts were referred as neurexins I and II because they constituted neuron-specific cell surface proteins.

FIBROMYALGIA

Definition and clinical features

Fibromyalgia (FM) is a highly disabling syndrome defined by a low pain threshold and a permanent state of pain. Widespread pain is accompanied by a constellation of symptoms such as fatigue; sleep disturbances and cognitive impairment, among others. The mechanisms explaining this chronic pain remain unclear. Onset of symptoms is usually gradual, but in some cases a sudden onset following an identifiable event has been described (44). FM is often debilitating and frustrating for physicians and patients: the chronic nature of the symptoms, the multiplicity of possible aetiologies and the lack of effective treatments, make difficult FM management (45).

Pain throughout the body is the pivotal symptom of FM (46). It is a cause of considerable suffering and functional impairment. Pain is continuous, although it is worse in the morning, improves during the day and worsens again by night. Pain is worsened by static positions, weather changes and physical burdens (47). It localizes in muscles and joints, mainly shoulders and hips, mimicking inflammatory joint diseases. It presents also neuropathic and visceral pain characteristics (48). Some patients also present pain with a superficial burning quality, increased sensitivity to painful stimuli (hyperalgesia) and features of allodynia (pain following an innocuous stimulus) (44).

The clinical understanding of FM has evolved over the last two decades, emphasizing the importance of symptoms other than pain that contribute to global suffering. More than 70% of the patients present fatigue. This can be continuous or appear as exhaustion periods of 1-2 days. Sleep disturbances is the third most frequent symptom and it is correlated with FM severity. Many components of sleep have been measured as abnormal in FM patients, including sleep quality, latency, duration and disturbances, and impaired daytime functioning (49). In particular, FM patients present alterations in the deepest stage of sleep, the delta-wave sleep (50). Furthermore, poor quality and duration of sleep has been shown to have a negative impact upon fatigue and affect (51). This classical FM triad (pain, fatigue and sleep disturbances) is accompanied by a wide spectrum of symptoms of different systems, since FM is a very heterogeneous disorder. These additional symptoms include sensitive symptoms, motor symptoms, vegetative symptoms, cognitive symptoms, and affective symptoms (Figure 4).

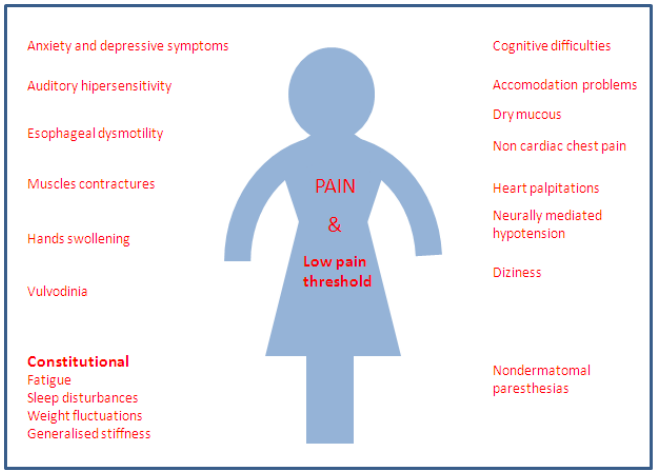


Figure 4: FM is a heterogeneous disorder.

FM patients frequently complain of cognitive symptoms, popularly referred as “fibrofog” (52). These include difficulty with reading and calculation, memory impairment, forgetfulness, and even anomic aphasia (recalling names) episodes. Cognitive deficits are more prevalent in FM patients than in pain free individuals but they are reported by other chronic pain patients (osteoarthritis and rheumatoid arthritis patients). A study showed that, regardless of disease status, chronic pain patients demonstrated cognitive impairment when performing everyday attentional tasks in comparison with matched pain-free controls (53). However, another study showed that these symptoms are more prevalent in FM than in other rheumatic diseases, suggesting the relevance to FM (52). Whether they are more prevalent in FM than in other chronic pain states may be unclear, but what is sure is that they constitute a major concern for FM patients. As a proof of concept, cognitive symptoms have been included in the new FM classification criteria (14).

Affective symptoms, mainly anxiety and depressive symptoms are also common FM features, as well as neurovegetative symptoms, including dizziness, sweating or palpitations.

FM patients, who are characterized by a low pain threshold, present tender points, which are points that are painful under pressure. Physical examination findings also include other musculoskeletal unspecific alterations, such as postural alterations, muscular hypertonia, muscular contractures, and painful stretching of affected regions.

Epidemiology

FM is estimated to affect more than 5 million Americans (2%-5% of the adult population) (54). The overall prevalence of FM in the general population of five European countries (France, Germany, Italy, Portugal and Spain) is 2.9% (55). In particular, according to the EPISER study, FM affects 2.4% of the Spanish population over 20 years old (56). FM patients are mainly women in the fourth decade of life, with a female:male ratio of 21:1 (48).

Only a small number of FM epidemiological studies have been performed in non-developed countries, showing a frequency similar to that of developed countries: 2.5% in Brazil and 3.2% in Bangladesh (57). This suggests that FM is not a disorder specific of developed countries. In fact, a study performed in the Amish population, which is an isolated society where modern life influence is absent, showed an increased prevalence of FM (7.3%) (58).

Diagnostic criteria

In the absence of suitable diagnostic tests, FM diagnosis is established by a history of symptoms and the exclusion of somatic diseases explaining these symptoms (14, 59). In the beginning of the nineties, with the development of the American College of Rheumatology fibromyalgia criteria, FM was defined as the presence of widespread pain with duration of at least 3 months. This pain should be bilateral, both above and below the waist, and it should include axial skeletal pain in combination with tenderness at 11 or more of 18 specific point sites on the body, called tender points (60) (Figure 5).

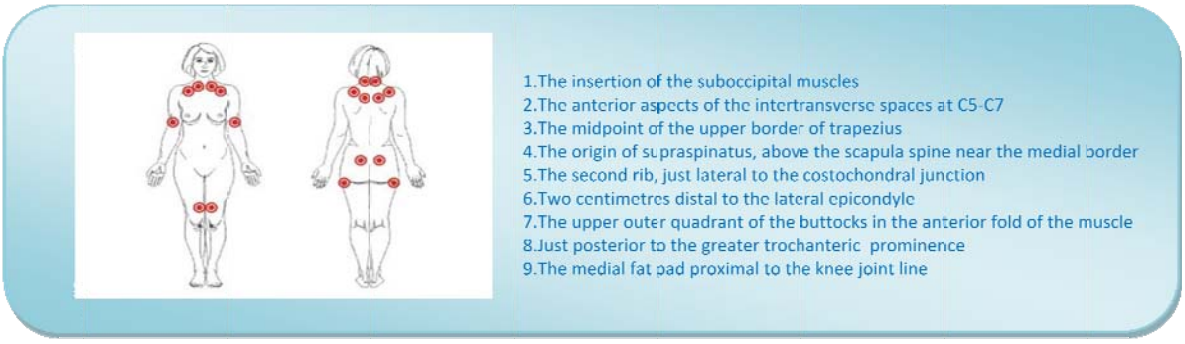


Figure 5: 1990 American College of Rheumatology (ACR) FM classification criteria based on the presence of 11/18 tender points. Each tender point is located on both sides of the body and painful under palpation with approximately 4 kg of pressure.

A series of objections to these classification criteria developed over time (14). First, it became increasingly clear that the tender point count was rarely performed in primary care, where most FM are diagnosed, and when performed, was performed incorrectly. Consequently, in practice, FM diagnosis has often been a symptom-based diagnosis. Furthermore, the specificity of these tender points became a matter of discussion because it was known as a hallmark of FM and was used by malingerers to simulate the disease. Second, the importance of symptoms that had not been considered by the American College of Rheumatology (ACR) became increasingly known as key FM features, as for example, fatigue, cognitive symptoms and somatic symptoms. Finally, there was still another important problem with FM diagnosis: patients who improved failed to satisfy the ACR 1990 classification definition. For all these reasons, a need to develop a broad-based severity scale able to detect clinical changes emerged. Based on a multicenter study of patients with diagnosis of FM and a control group of rheumatic disease patients with non-inflammatory disorders, new FM diagnostic criteria were developed (14).

These 2010 FM criteria where not meant to replace 1990 ACR criteria. They proposed a new definition of FM with increasing recognition of the importance of cognitive problems and somatic symptoms. They were based on a Widespread Pain Index (WPI), with scores between 0 and 19, for the number of areas in which the patient had had pain over the previous week; a Symptom Severity (SS) score, where fatigue, sleep disturbances and cognitive symptoms severity were scored from 0 to 3, plus the extent of somatic symptoms in general. FM 2010 classification criteria were simplified a few months after their publication, by substituting somatic symptoms in general by headaches, pain in abdomen, and depression (Figure 6). A 0-31 FM (FS) scale was created as the sum of the WPI and the SS score, enabling the use of these 2010 modified FM classification criteria in studies without the need of an examiner.

APPENDIX. Fibromyalgia criteria modified from American College of Rheumatology diagnostic criteria.

Criteria

A patient satisfies modified ACR 2010 fibromyalgia diagnostic criteria if the following 3 conditions are met: (1) Widespread Pain Index ≥ 7 and Symptom Severity Score ≥ 5 or Widespread Pain Index between 3–6 and Symptom Severity Score ≥ 9. (2) Symptoms have been present at a similar level for at least 3 months. (3) The patient does not have a disorder that would otherwise sufficiently explain the pain.

Ascertainment

1). Widespread Pain Index (WPI): Note the number of areas in which the patient has had pain over the last week. In how many areas has the patient had pain? Score will be between 0 and 19.

Shoulder girdle, Lt.	Hip (buttock, trochanter), Lt.	Jaw, Lt.	Upper Back
Shoulder girdle, Rt.	Hip (buttock, trochanter), Rt.	Jaw, Rt.	Lower Back
Upper Arm, Lt.	Upper Leg, Lt.	Chest	Neck
Upper Arm, Rt.	Upper Leg, Rt.	Abdomen	
Lower Arm, Lt.	Lower Leg, Lt.		
Lower Arm, Rt.	Lower Leg, Rt.		

2). Symptom Severity Score: Fatigue; Waking unrefreshed; Cognitive symptoms.

For the each of these 3 symptoms, indicate the level of severity over the past week using the following scale: 0 = No problem; 1 = Slight or mild problems; generally mild or intermittent; 2 = Moderate; considerable problems; often present and/or at a moderate level; 3 = Severe: pervasive, continuous, life-disturbing problems.

The Symptom Severity Score is the sum of the severity of the 3 symptoms (fatigue, waking unrefreshed, and cognitive symptoms) plus the sum of the number of the following symptoms occurring during the previous 6 months: headaches, pain or cramps in lower abdomen, and depression (0–3). The final score is between 0 and 12.

Figure 6: FM 2010 modified classification criteria (from Wolfe *et al* 2011) (61).

Disease evolution and prognosis: Fibromyalgia comorbidities

FM is a chronic disease. Usually pain remains at the same level, with episodes of enhanced clinical activity. Better economical and educational levels, as well as the absence of psychiatric comorbidity, are linked with a better prognosis (62). A Danish study showed, however, that despite no overall increase in mortality, FM patients have an elevated risk of suicide (10-fold increase in mortality risk), liver cirrhosis/biliary tract disease (six-fold increase in mortality risk) and cerebrovascular disease (threefold increase in risk) (63). It is controversial whether FM patients have a higher risk of cancer (64, 65).

In addition to the multiple symptoms accompanying chronic pain, comorbid conditions are often present in FM; in particular somatoform disorders* (Figure 7). FM has even theoretically been defined as an overlap of syndromes and symptoms rather than as a discrete entity (66). In fact, some physicians even claim that FM is more a rag bag than a disease by itself. It is usual to find in many scientific publications the term fibromyalgia syndrome (FMS) instead of FM. The Chronic fatigue syndrome (CFS), present in 21-70% of FM cases, is the most common of these overlapping disorders (Box 4) (45). Psychiatric comorbidities are also frequent, but the role they play in FM onset and development has not been well defined yet. Major depression, panic attacks; posttraumatic stress disorder and personality disorders are the most commonly associated with FM.

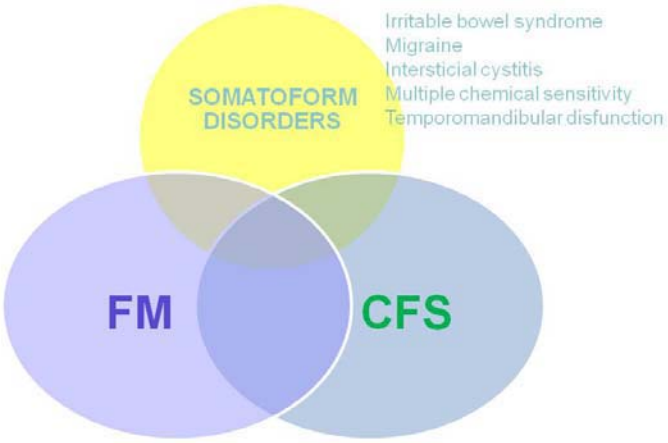


Figure 7: Fibromyalgia overlaps with other somatoforms disorders.

Box 4: Fibromyalgia or chronic fatigue syndrome?

Numerous physicians underscore the difficulty in the diagnosis and classification of patients with somatic complaints of diffuse pain and constant fatigue, by giving them two diagnoses: FM and CFS. CFS is defined by unexplained, persistent, or relapsing fatigue with a definite onset. It is quite controversial whether these are two distinguishable entities or different manifestations of the same disorder. In fact, 20-70% of FM fulfills CFS diagnostic criteria. In the Spanish health care system, one curious think is that FM patients are controlled by rheumatologists and CFS by internists. In fact, some theories claim that there is a continuum in FM and CFS symptomatology and that those patients fulfilling both FM and CFS criteria are the ones with a more severe disease.

Differential diagnoses

For a diagnosis of FM, first all disorders where pain and fatigue are major symptoms have to be discarded (Box 5). Most of these disorders can be discarded through a blood test including a complete blood count, kidney and liver function tests, creatine kinase, fasting serum calcium and phosphorus, C-reactive protein/ erythrocyte sedimentation rate, thyroid function tests, and rheumatologic serologies. Any abnormalities in any of these should be followed up with more focused and specific tests (e.g.: the presence of anaemia and high erythrocyte sedimentation is consistent with multiple myeloma; it should be discarded through a protein electrophoresis in blood serum). Furthermore, endocrine disorders and infectious diseases could be mimicking FM because they are causing muscle pain and/or fatigue. Psychiatric disorders constitute a differential diagnosis for FM, as FM patients show very often affective symptoms. When present along with symptoms of FM, the physician must come to a decision about which causes what (45). The presence of pain and fatigue as major symptoms concentrating the patient concern will point out to an FM diagnosis. More

details about the other disorders listed in Box 5 can be found in the glossary section (Annexes). It must be taken into consideration that, in most of these chronic diseases, the diagnosis of FM is not excluded by the presence of other diseases, as it may be secondary to these.

Box 5: Fibromyalgia differential diagnosis

- Autoimmune disorders
  - Celiac sprue
  - Polymialgia rheumatica*
  - Polymiositis*
  - Seronegative spondyloarthropathy*
  - Systemic lupus erythematosus*
- Endocrine disorders
  - Hyperpartiroidism
  - Hypophosphatemia
  - Hypothyroidism
- Infectious diseases
  - Hepatitis C
  - HIV acute infection
  - Nonviral meningoencepahlitis
  - Postviral encephalitis and meningitis
- Malignancies
  - Neoplasia: primary tumor (multiple myeloma) and bone metastasis
  - Paraneoplastic disorders*
- Musculoskeletal disorders
  - Costochondritis
  - Lumber nerve root compression
  - Reflex sympathetic dystrophy*
  - Spinal stenosis
- Neuromuscular disorders
  - Miastenia gravis
  - Multiple sclerosis
  - Neuropathies
- Psychiatric disorders
  - Dysthimia
  - Seasonal affective disorder
  - Melancholic major depression
- Sleep disorders
  - Obstructive sleep apnea
- Malingerers

Finally, since FM is diagnosed based on symptoms and physical examination, without an objective diagnostic test, it has become an appealing disease for malingerers. The high rates of disability claims for FM patients, despite very little objective data, explain this phenomenon. This complicates even more the diagnosis of FM.

Clinical impact of fibromyalgia

FM is one of the rheumatic diseases with a higher impact on life. Patients refer big consequences on their lives in terms of physical activity, intellectual disability, emotional state, mental health and employment. Patients that have a paid employment show a better prognosis (67).

Despite the lack of organic pathology, patients with FM often suffer disruption of their social structure and this potentiates their symptoms (Figure 8). For example, an executive in a corporation may, secondary to severe stress, have an emotional breakdown that evolves into a chronic pain syndrome. This may, in turn, lead to a loss of employment, disability, separation or divorce, and maladaptive illness behaviour (45).

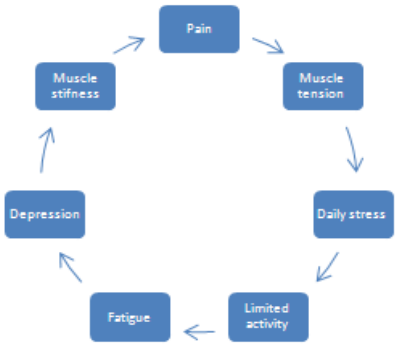


Figure 8: The fibromyalgia cycle. FM symptoms have a big impact in patients' life, and this aggravates symptoms.

Not only is the disease by itself highly disabling. There are still many physicians and people in patient's social environment that are reluctant to consider FM as a disease, and think that FM patients are hypochondriacs or malingerers (Box 6). This lack of comprehension wears out the patient.

FM can have a devastating effect on people's lives but also their relationship with family, friends and work colleagues (68). Parents, siblings and partners of FM patients have a worse life quality than family members of non affected subjects (67).

Furthermore, FM constitutes a significant economic burden for society. FM subjects are found to have substantial costs, over 75% of which are driven by indirect costs from lost productivity, and these costs increase as FM severity increases (59). In addition to that, FM patients use more health resources than general population and incur similar costs as other rheumatic diseases such as rheumatoid arthritis (69). Given the prevalence of the disorder it has become a major concern as an economical issue. In order to optimize FM management and minimize its costs by reducing the number of consultations to general practitioners, in 2010, the Catalan Institute of Health (ICS) created FM-CFS units. These units consist of multidisciplinary teams, including a rheumatologist, a psychiatrist, a psychologist and a nurse, to ensure the

best management of FM patients. It remains to be seen whether this initiative leads to a reduction of FM costs. Furthermore, evaluation of work disability in FM patients is a controversial issue (47). Disability claims are more and more frequent. For the physicians it is difficult to evaluate a patient based on non objective arguments. In Catalonia, the final decision is taken by ICAMS (Institut Català d'Avaluacions Mèdiques i Sanitàries). Last year, the first patient was dismissed of work because of FM. (http://www10.gencat.cat/sac/AppJava/organisme_fitxa.jsp?codi=13404).

Finally, we should not forget that the economical impact is also at the personal level, since non-pharmacological treatments are not covered by the health service system. FM patient has to assume the cost of psychotherapies, and alternative medicine treatments.

Disease activity measurement: clinical scales

FM diagnosis is based on the presence of clinical symptoms. There is no test to confirm and to monitorize disease evolution or response to treatment. For this reason, in everyday clinics, clinical scales are widely used for FM management. Some of these scales have been specifically designed for FM, such as Fibromyalgia Impact Questionnaire (FIQ), while most of them are used to measure FM symptoms (mainly psychiatric comorbidities) and impact on everyday life (life quality scales such as SF-36). FIQ is easy to use and is sensitive to clinical evolution. These scales have been translated and adapted to different languages. (See Scales annex).

Symptoms treatment

As stated before, FM is characterized by pain plus a large constellation of symptoms. As there is no cause for these symptoms, treatment is only symptomatic, with no impact on disease prognosis or evolution. In fact, there is no effective treatment. Response is poor, and the modest benefits of FM treatment are a cause of frustration for both patients and physicians (45). Practitioners have to emphasize that it is a chronic disease and to fight against the stigma attached to this 'invisible illness' (Box 6). They also have to take into consideration that most patients experience years of suffering and many doctors' visits. Patients need validation of their symptoms and need to know that their disease is real: physician's empathy is crucial (70).

The major goals of FM treatment are decreasing pain, improving sleep and establishing a regular exercise program. Moreover, as each patient has a different constellation of symptoms, treatments have to be tailored (70), treatment programs have to be defined with care. Sometimes, due to the difficulty of the patient in coping with symptoms that don't improve in spite of the medication, there is a tendency to consult different physicians at the same time (public system, private system, and general practitioners).

This can lead to a dangerous pharmacological cocktail than, far from be beneficial to the patient, may result in the accumulation of side effects rather than in a therapeutic outcome. In addition, many FM patients are unusually sensitive to the adverse effects of medication (70). For this reason, treatments should be started at low doses and increased gradually.

Evidence-based guidelines suggest that FM has to be managed with multidisciplinary therapies involving medication and non-pharmacological interventions. This multidisciplinary treatment has been successfully implemented in FM multidisciplinary units (71), which have also helped in patients coping with symptoms and avoiding multiconsultation.

The relative role of role of each of the treatments varies among the different guidelines. Whereas the American Pain Society strongly recommends a pharmacological intervention (amytriptiline) plus non pharmacological treatments (62), the European League against Rheumatism (EULAR) strongly recommends multiple pharmacological treatments only (72). A recent systematic review of randomised trials has tried to clarify these contradictory guidelines (59). From this study, a combination of pregabalin or serotonin-noradrenaline reuptake inhibitors (SNRIs), as pharmacological interventions, and aerobic exercise and cognitive behavioural therapy (CBT), as non-pharmacological interventions, appear as the best options for FM management.

As mentioned before, the treatment is individualized, mainly according to treatment response and medication side effects. A summary of the most common treatments, both pharmacological and non-pharmacological FM interventions, is presented below.

Pharmacological treatments

Paracetamol and mild opioids (tramadol)

Paracetamol can be considered for FM treatment, although there is little evidence of response. According to EULAR, mild opioids such as tramadol are also recommended for the management of FM (72).

Non-steroidal anti-inflammatory drugs (NSAIDs) and glucocorticoids

According to different trials, NSAIDs and glucocorticoids are ineffective and are therefore not recommended in the treatment of FM (72). However, most FM patients are current users of NSAIDs. Their availability and common use by the general population explain their use in FM treatment.

Antidepressants

Tricyclic antidepressants, amytriptiline in particular, are among the best-studied and most effective pharmaceutical interventions in FM. Serotonin reuptake inhibitors (SSRIs) have mostly been less effective than tricyclic medications, but newer studies demonstrate some efficacy of SSRIs. A randomized, double-blind crossover trial of fluoxetine (and SSRIs), amytriptiline and placebo in FM patients showed that both were effective in decreasing FM symptoms, and that the two drugs given simultaneously were more effective than either drug alone. Finally, a meta-analysis of FM treatment with antidepressants showed that antidepressants tended to improve sleep, fatigue, pain, and well-being, but not trigger points (45).

Anticonvulsants

Gabapentin and pregabalin, drugs that were initially synthesized to treat epilepsy, have shown efficacy in the management of painful diabetic neuropathy and postherpetic neuralgia. Their mechanism of action likely consists of binding to the voltage-dependent calcium channel in the CNS, blocking the influx of calcium into the neuron and thereby reducing excitatory neurotransmitter release. Gabapentin showed significant improvement in pain, FIQ score and sleep, and it was generally well tolerated. Pregabalin significantly decreased pain, and, for higher doses, lead to significant improvement in sleep quality.

Dopamine agonists

FM patients may benefit of dopamine agonists if they also have restless leg syndrome. Pramipexole improved pain, fatigue, function and global status.

Corticosteroids infiltrations

For common tendinitis such as epicondilitis or trocanteritis, that do not improve with oral treatments, infiltrations with corticosteroids will improve symptoms.

Sympathetic blockade

Regional sympathetic blockade through stellate ganglion blockade, with local bupivacaine injection, has shown to decrease trigger points and resting pain, but it does not offer a practical therapeutic option for FM (45).

Non-Pharmacological treatments

Cognitive behavioural therapy

Cognitive behavioural therapy is a process that examines a patient's way of reacting to experiences and attempts to restructure maladaptive coping habits into effective coping skills. CBT has been shown to be effective in reducing disability in most of FM studies (45). Emotional and cognitive factors (thoughts, memories, beliefs, expectations and fears) interact with how a patient handles sensory input, including pain perception (70).

Balneotherapy

Heated pool treatment or balneotherapy was reported to be effective in improving pain and function (72).

Aerobic exercise

Individually tailored exercise programmes, including aerobic exercise and strength training can be beneficial to some patients with FM (72). Aquagymn and pilates are the most recommended exercise programmes for FM patients, as they are mild and don't imply an overcharge that could worsen fatigue.

Relaxation and Tai chi

Tai chi has shown to reduce disease activity as assessed by FIQ and quality of life (SF-36) in a randomized controlled trial (73).

Lifestyle changes

Improving sleep hygiene and eating a healthy diet have also shown to be beneficial.

Alternative medicine

In spite of the many treatments offered in the Internet, there is little research on alternative medicine and FM. Acupuncture effects are not clear (70). Desperate FM patients may suffer the abuse of cheaters with bizarre treatments not based on medical evidence such as anal ozone.

Evidences for a biological basis

In the eighties and nineties there were several hypotheses pointing at almost any possible system alteration/dysregulation, as being responsible for FM symptomatology: from muscle, to the adrenal system, the immune system, a sleep disorder and the nervous system.

Nowadays, the most established hypothesis underlying FM etiopathogenesis is the existence of a dysfunction in pain processing. Since pain and tenderness are the defining features of FM, it is currently attributed to an increase in central pain processing. However, there are still other hypotheses that claim that FM could be an endocrine or an immune mediated disorder, or even due to the dysfunction of the autonomic nervous system. We will focus on the evidences supporting central and peripheral nervous system alterations, as they now appear as the more plausible and more evidence based ones, but we will also summarize the existing evidences supporting the other hypotheses.

Box 6: Fibromyalgia: a little bit of history and controversy

Although it may seem that FM is a modern disorder, it was in fact already described in the beginning of the twentieth century. The first references to patients who have chronic pain, an "aching, stiffness, a readiness to feel muscular fatigue, interference with free muscular movement, and very often a want of energy and vigour," was made by Ralph Stockman, an Edinburgh pathologist in 1904. Just a few years earlier, William Osler, in the third edition of his text *The Principles and Practice of Medicine,* defined *myalgia* in similar terms. Because of wide variability in symptoms in patients with pain with no apparent reason, there was little enthusiasm among clinical investigators to examine these patients more closely until 1939, when Lewis and Kellgren enlisted volunteers and subjected them to injections of muscles and ligaments with hypertonic saline (which is very painful in the injection region). These deep injections caused discomfort in a myotomal, or referred pattern, as opposed to a dermatomal or radicular pattern. A resurgence of interest in FM was introduced in the 1980s when several groups of investigators published data showing that patients with fibrositis or fibromyalgia had particular tenderness in at least 11 of 18 defined "tender points" not shared by asymptomatic control patients. During the 1990s however, it has become apparent that healthy people have tender points, that FM patients are often tender all over their bodies and could best be characterized as having a low pain threshold, and that FM, the chronic fatigue syndrome, exposure syndromes and somatoform disorders have considerable overlap in their clinical manifestations.

Central nervous system pain-processing dysfunction

Extensive research suggests that FM's chronic widespread pain has a neurogenic origin. First of all, FM patients present structural differences in the brain. Two studies have detected a decrease in grey matter volume, which was greater than in controls and the equivalent to 9.5 times the normal loss with age (74). This has been postulated to be related to stress, deficits in cognitive function and impaired endogenous pain inhibition, as many of the grey matter loss occur in regions related to stress and pain processing. Other studies have demonstrated changes in volume of different brain structures such as the amygdala, cingulated cortex and hippocampus, as reviewed in Gracely *et al* (75).

Furthermore, there are several evidences of central sensitization at various levels in the nervous system (44). Neuroimaging studies support this, showing that FM is associated with aberrant processing of painful stimuli

in the CNS. fMRI studies of the brain demonstrate that in patients with FM a pain response can be produced using a much lower pain stimulus than in controls (68). Under the same stimulus intensity, several areas of the brain (secondary somatosensory cortex, insula, and the anterior cingulated cortex) consistently show greater activation in FM patients than in control individuals with fMRI. SPECT imaging has shown reduced blood flow in the right thalamus of FM patients (75, 76). FM patients present a widespread reduction in thermal and mechanical pain thresholds and greater cerebral laser evoked potentials after mechanical stimuli (4). Heat pain thresholds and cerebral event-related potentials following painful CO_2 laser stimulation are also altered in subjects with fibromyalgia syndrome (77). Differential central pain processing has also been observed following repetitive intramuscular proton/prostaglandin E(2) injections in female FM patients and healthy controls (78).

The pain seems to result from neurochemical imbalances in the central nervous system leading to a “central amplification” of pain perception. PET has shown a reduction of levo-dopamine uptake within the brainstem, thalamus, and multiple elements of the limbic cortex (79), and a reduction of opioid μ receptors binding potential in structures playing a role in nociception, such as amygdala, cingulated and nucleus accumbens, (80). This could explain the lack of response to opioids in FM patients. In addition, MRS has demonstrated differences in concentrations of glutamate and combined glutamate/glutamine within the insula and posterior gyrus (81) in FM patients as compared to controls. Furthermore, these changes in glutamate levels have been associated to changes in pain perception (80).

Moreover, widespread pain in fibromyalgia is related to a deficit of endogenous pain inhibition, due to an imbalance in descending pathways as proven by altered levels of neurotransmitters in the CNS (82). Noradrenaline (NE) and 5-hydroxytryptamine (serotonin) (5-HT) are key neurotransmitters in descending inhibitory pain pathways and they have a significant modulatory effect on peripheral and central pain processing. Levels of primary metabolites of NE, 5-HT and dopamine (DA) are reduced in patients with FM (83, 84). Another study found correlation between the levels of glutamine and aspargine (glutamate and aspartate metabolites respectively) and the number of tender points (85). Four different studies have also found an elevation of substance P (a neuropeptide released in spinal fluid when axons are stimulated) in cerebrospinal Fluid (CSF) of FM patients as reviewed in (76). Nerve growth factor also has shown to be elevated in the CSF of FM patients (86).

Finally, enhanced central pain processing of FM patients is maintained by muscle afferent input: a randomized, double-blind, placebo-controlled study showed the important role of peripheral impulse input in maintaining central sensitization in FM (87).

Disordered stress response and endocrine or hormonal factors

The hypothalamic-pituitary- adrenal (HPA) axis along with the locus coeruleus-sympathetic nervous system is the principal system of stress response in the body. The HPA axis has been found to be dysregulated in FM patients. Several studies using different physiological stress generators showed enhanced or normal response of adrenocorticotrophic hormone (ACTH) with elevated levels of cortisol in the evening as reviewed in Di Franco (84). Because the 5-HT system influences the HPA axis, some of these findings could be caused by the reduced levels of 5-HT in plasma. These alterations are commonly detected in patients with a defined cause of chronic pain, suggesting that these changes could be a consequence of FM symptoms, rather than their cause (45). In any case, FM studies suggest an impaired stress ability to activate the system, rather than an overall increased function of the HPA system.

Autonomic nervous system dysfunction

The balance between sympathetic and parasympathetic systems is essential to preserve homeostasis. Studies show that FM patients have increased sympathetic and decreased parasympathetic tones, as assessed by 24 hours electrocardiogram registration (reviewed in Martínez-Lavin (88)). This dysautonomia is characterized by a persistently hyperactive sympathetic nervous system that is hyporeactive to stress, as proven by different studies using orthostatism and tilt-table test among other stressors. These alterations explain many of the FM symptoms, such as intestinal dysfunction, or heart palpitations, and suggest that FM could be a sympathically maintained pain syndrome. However, as in the case of the HPA axis dysfunction, the vast majority of the scientific community considers these alterations a consequence of FM, rather than a cause.

Sleep disorder

The delta sleep alterations found in FM have shown to interfere with sleep function, causing non restorative sleep, fatigue and musculoskeletal pain (89). In fact, experimental studies with normal control patients undergoing delta sleep disruption showed FM like symptomatology (90). The question is whether these alterations are cause or contributing factors to FM. In this sense, it has been proposed that the impaired slow-wave sleep could be causing the decrease in growth hormone secretion that is present in FM patients (91). Also, sleep disturbances could be affecting alterations in 5-HT, NA, DA and melatonin levels in FM.

Muscle

Microscopical examination of FM muscle showed muscle fibres were connected by a network of reticular or elastic fibres, which are not present in muscle from healthy individuals (92). Also, other studies have detected changes in intramuscular microcirculation and in muscle energy metabolism as reviewed in (13). As

mentioned before, some hypotheses point out that peripheral activation, which could take place through these muscular changes, is necessary for the development of central sensitization and pain disinhibition mechanisms.

Inflammatory/autoimmune disorder

Chronic conditions, such as cancer or autoimmune disorders, often have associated mood disorders. It has been shown that pro-inflammatory cytokines, such as Interleukine (IL)1 and tumoral necrosis factor (TNF) α, could be involved in the development of sickness behaviour, including fatigue and increased pain sensitivity (reviewed in Dantzer *et al* (93)). This, and the fact that synergistic neuroimmune interactions promote sensitization to pain and the development of chronic pain (3), point out at a possible role of an immune dysregulation in FM pathogenesis. Antinuclear antibodies (ANA) have been shown in 11% to 30% of all FM patients, but the risk to develop connective tissue diseases does not seem to increase in FM, and detection of ANA does not have a predictive value. Various studies have evaluated Interleukin levels in FM patients, with contradictory results. Some of them showed increased levels, notably IL-8, while others show decreased levels (reviewed in Dadabhoy *et al* and Di Franco *et al* (76, 84)). A recent meta-analysis (94) concluded that only IL-6 plasma levels were higher in FM patients compared to controls. The different methodologies used to measure different cytokines in different tissues materials explain in part this lack of reproducibility across studies. A posterior work has shown increased CSF and serum levels of IL-8, proving that the relationship between IL and FM is still a controversial non-resolved issue.

Finally, the role of the immune system in FM pathogenesis is also supported by some genetic studies that link human leukocyte antigen (HLA) locus with FM familial cases (46).

Environmental factors

Several trigger factors for the development and onset of FM, such as surgery, stressful life situations (Box 7), infections, traumatisms, and blood transfusion have been described. Up to 40% of patients report that the onset of symptoms was preceded by some triggering event, which might be either psychological or physical. FM development would be favoured by a vulnerable psychosocial setting. Experiences of physical and sexual assault in adulthood showed a strong association with FM (95). In a retrospective case-control study, FM was significantly associated to physical trauma (a fracture, surgery, miscarriage or childbirth, and a traffic or other type of accident) six months prior to FM onset (96). Other factors described in FM risk are low levels of vitamin D in women (97), certain infections (98), lower levels of education, unemployment, divorce and obesity.

Finally, several infectious agents have been linked to the development of FM. Due to the similarity between the viral infections and FM symptomatology, several viral agents, in particular hepatitis C and B and human

immunodeficiency virus (HIV), have been explored in FM epidemiological studies with inconclusive results, as reviewed in (99). The most plausible explanation for these findings is that after these chronic infections, patients may develop secondary FM.

Box 7: Pain and war

During times of war, chronic pain and fatigue with normal physical examinations has approached epidemic proportions. Comrow, in his text *Arthritis and Allied Conditions*, noted in 1944 that "a large percentage of our soldiers entering the medical service of an Army General Hospital with symptoms simulating fibrositis have developed these on a psychogenic basis and that these symptoms cannot be relieved by heat, massage and exercise, but are abolished by discharge from the army". The Persian Gulf syndrome may be another example of this, because 45 % of deployed veterans, compared to 15 % of those not deployed to the first Persian Gulf War, developed a constellation of symptoms including muscle and joint pain, fatigue, memory problems, headaches, and gastrointestinal complaints (45). A specific mechanism that may link vaccination against biological warfare agents and later FM symptoms has been suggested (100). Hotopf *et al* concluded that among veterans of Gulf war there was a specific relation between multiple vaccinations during deployment and later ill health (101).

Fibromyalgia genetics

The response to painful stimuli has a genetic component, as heritability is estimated between 22% and 55% (102). However, the exploration of genetic contribution to pain response and chronic pain states is so far scarce. Copy number variants (CNV) have not been explored in FM or in other chronic pain conditions (103). Regarding single nucleotide polymorphisms (SNP), only one pain genome-wide association study has been performed, evaluating pain levels after third molar extraction. This only found association at one SNP, which was in linkage disequilibrium with a gene encoding a zinc finger protein, but this was not evaluated in a replication cohort (104). Regarding chronic pain, a study evaluating 3295 SNPs related to pain research in 348 cases of chronic temporo-mandibular disorders and 1612 controls failed to detect a statistically significant association after correcting for multiple testing (105). A similar situation occurs in regard with FM: little is known about genetic factors underlying this chronic pain state.

FM has a genetic component: family and twin studies

FM has a genetic component, as there are evidences of family aggregation. Several studies have tried to assess the occurrence of FM among family members of FM patients. First, a study in 50 parents and siblings of FM patients showed that 52% of the individuals presented FM symptoms (106). Another showed a higher prevalence of FM among offspring of FM mothers (28%) (107). Furthermore, this increased prevalence among family members of FM patients was shown to be higher in blood family members than in husbands of FM patients (108), supporting that genetic factors must prevail over environmental factors in determining FM susceptibility. Finally, genetic contribution to FM was also supported in a study, showing that FM first degree relatives had a relative risk of 8.5 to develop FM, as compared to rheumatoid arthritis first degree relatives (109).

Several twin studies have also been performed in order to dissect out FM genetic factors. First, a study in 11-year-old Finnish twins with widespread pain showed that genetic factors did not account for twin similarity for widespread pain (110). However, pain was not chronic in these individuals, and therefore FM criteria were not fulfilled. Later, a twin study in the Swedish population showed that chronic widespread pain concordance was higher in monozygotic than in dizygotic twins (0.29 versus 0.16) and estimated that genetic factors accounted for 48% to 54% of the total variance (111). Finally, a twin study in idiopathic chronic fatigue showed also a higher concordance in monozygotic (55%) than in dizygotic twins (19%). In conclusion, studies performed so far support the importance of genetic factors in susceptibility to FM and chronic pain states.

Fibromyalgia genetic studies

Most genetic studies performed so far in FM have been candidate gene studies. These studies have not been able to establish a clear genetic association, as they present several limitations: most of them have been performed in small cohorts; associations are not FM specific as they are shared with other disorders, in particular FM psychiatric comorbidities; they often present borderline significance without correction for multiple testing; and attempts to replicate them have shown contradictory results. So far, studies have focused on genes related to HLA and neurotransmitters. In a study, where 18 FM patients and 23 controls were typed for HLA alleles, 67% of FM patients presented HLA-DR4 versus 30% of controls. Given the reduced number of samples included in the study, FM HLA-DR4 frequency was also compared to a larger cohort of controls (n=1676) and the difference remained statistical significant (112). However, not unexpectedly, a similar study, performed in 60 FM cases and 159 controls, failed to replicate these findings (113). Another study, investigating the HLA region, showed genetic linkage to this region in an analysis of forty Caucasian families with FM, in which HLA was typed for A, B and DRB1 alleles (46). Nevertheless, none

of these HLA associations have been replicated in larger FM cohorts. On the other hand, various studies have examined the possible role of neurotransmitters in the serotoninergic, dopaminergic and catecholamiregic systems in FM. The analyses of serotonin transporter and catechol-O-methyltransferase variants in FM are good examples that illustrate the limitations of FM genetic studies performed so far.

A possible association of FM with a polymorphism in the regulatory region of the serotonin transporter gene has been suggested. There is a common variant leading to a shorter form of the promoter, which has been associated to several FM psychiatric comorbidities, including affective and anxiety disorders. This variant showed a higher frequency in FM individuals in respect with controls (62 FM versus 110 controls, 31% of short homozygous in FM, and 16% in controls p=0.046), although it did not reach statistical significance at the allelic level (114). It was a borderline association that was replicated in another study confirming the association (115), but not in a third one when analysing FM patients without psychiatric comorbidities (116) (Table 1). Other serotonin transporter variants have been further explored in FM, finding association of a SNP (rs6313) to FM both at the allelic and genotypic levels, but again in a reduced sample set of 168 FM versus 115 controls and without replication. This variant had also been previously associated with schizophrenia and migraine (117). Its association to FM was not replicated in a second study (118) (Table 2).

A common variant affecting catecholamine's metabolism on the modulation of responses to sustained pain in humans has also been explored in FM cases. Individuals homozygous for the Met158 allele of the catechol-O-methyltransferase (COMT) polymorphism (Val158Met) show diminished regional μ opioid system responses to pain compared with heterozygotes (119). Several studies have evaluated this variant in FM cases with contradictory results, the one including more samples (115), only being able to detect an association between FM and the number of tender points (Table 3). A meta-analysis showed no association when considering three (Vargas *et al* (120), Tander *et al* (121) & Cohen *et al*(115)) FM studies performed on this variant until March 2010 (122). Although an ulterior study showed that the frequency of COMT variants associated with low enzyme activity was significantly higher in 113 FM patients versus 65 controls (123), the reduced number of samples included and the multiple contradictory results prove that this association still has to be established.

In addition, genetic variants in the dopaminergic and adrenergic systems have also been explored in FM with the same lack of consistency in the results, as reviewed in Lee *et al* (122). One work performed in Mexican cases and controls was able to find an association to a β2 adrenergic receptor variant, and replicated it in an independent Spanish dataset (124). However, the association did not pass correction for multiple testing (as several variants were tested), and the size of the study cohorts was small for the exploration of a common variant in a complex disorder.

Finally, so far only one study has attempted to explore the genetic contribution to FM in a genome-wide manner. Over 3200 SNPs in 350 genes implicated in pain transmission, inflammatory responses, and in influencing mood and affective states associated with chronic pain conditions, were genotyped in 496 FM cases and 348 controls. However, the strongest associations (*GABRB3 (*gamma-aminobutyric acid A receptor, beta 3) and *TAAR1* (trace amine-associated receptor 1) SNPs) did not replicate in independent cohorts (105).

In conclusion, the many genetic studies performed so far in FM have not been able to identify its genetic component. A summary including most of the genetic variants that have been explored in FM is presented in Table 4.

Table 1: Evaluation of a polymorphism in the promoter region of the serotonin transporter (5-HTTPRLR) gene in FM cases show contradictory results in different studies.

5-HTTPLR Short/long allele	**FM**	**CONTROLS**	**p-value**
(Offenbaecher (114))	61	110	0.046
(Cohen 2002(115))	48	54	0.001
	51	497	0.024
	99*	551*	0.0019*
(Gursoy2002 (116))	53	60	NS
(Potvin 2010(125))	58	60	NS

NS, non significant.*joint analysis

Table 2: Evaluation of rs6313 5-HT2A (serotonin 2A) receptor gene in FM cases versus controls.

5-HT2Areceptor rs6313	**FM**	**CONTROLS**	**p-value**
(Bondy 1999(117))	168	115	0.008
(Gursoy2001(118))	58	58	NS
(Tander 2008 (121))	80	91	NS

NS, non significant.

Table 3: Evaluation of catechol-O-methyltransferase (COMT) Val158Met polymorphism in several FM cohorts.

COMT rs4680 (Val158Met)	**FM**	**CONTROLS**	**p-value**
(Gursoy2003 (126))	61	61	0.024
(Vargas2007 (120))	57	33	NS
	78	80	0.023
(Tander2008 (121))	80	91	NS
(Cohen 2002(115))	209	152	<0.05 tender points
(Lee 2012 (122))	424	356	0.9
(Martinez2012(123))	113	65	<0.05

NS, non significant

Table 4: Summary of genetic variants tested in FM cohorts

Gene	Variant	Number Individuals FM	Controls	p-value	Ref
AAT	E342K	87	200	NS	(127)
ADRA1A	(rs574584, rs138914, rs573542)	78	48	0.02	(124)
ADRB2	(rs1042713, rs1042714)	78	48	0.04	(124)
		78	71 (replicate)	0.05	
ADRB3	rs4994	78	48	NS	(124)
COMT	(rs6269, rs4633, rs4818, rs4680 (Val158Met)	57	33	NS	(120)
		78	80 (replicate)	0.006	(120)
	rs4818	57	33	NS	(123)
		78	80 (replicate)	0.001	
DAT	VNTR	87	200	NS	(127)
DRD3	Ser9Gly	37	36	NS	(125)
DRD4	Exon 3 VNTR	81	458	0.034	(128)
eNOS	Glue298Asp	96	79	NS	(129)
GABR3	rs4906902	496	348	0.0000036	(105)
GBP1	rs7911	496	348	0.000106	(105)
HLA	HLA-DR4	18	23	0.08	(112)
		16	76	0.016	
HTR2A	rs6311	80	91	NS	(121)
HTR3A	Exon 1 42 C/T	96	312	NS	(130)
	Exon2 97 G/A	96	312	NS	(130)
	Exon3 IVS3	96	312	NS	(130)
	Exon 6 576 G/A	96	312	NS	(130)
	Exon9 1377 G/A	96	312	NS	(130)
HTR3B	Exon 1 -102 -100delAAG Exon 4 IVS4	96	312	NS	(130)
	Exon 5 386A/C	96	312	NS	(130)
	Exon 6 IVS6 + 72A/G	96	312	NS	(130)
		96	312	NS	(130)
IL-4	Intron 3 VNTR	62	101	NS	(131)
MAO-A	Promoter VNTR	107	90	0.055	(132)
	941 G/T	62	101	NS	(131)
MAO-B	Intron 13 G/A	107	90	NS	(132)
SNC9	14 SNPs tested (rs6754031)	73	48	0.036	(133)
TAAR1	rs8192619	496	348	0.000011	(105)
TACR1	1354G/C	87	200	NS	(127)

AAT, alpha-1 antitrypsin; *ADRA1A,* adrenergic receptor alpha-1-A; *ADRB2,* adrenergic receptor beta-2; *ADRB3,* adrenergic receptor beta-3; *COMT,* catechol-O-methyltransferase; *DAT,* dopamine transporter; *DRD3,* Dopamine-D_3-receptor; *DRD4,* Dopamine-D_4-receptor; *eNOS,* endothelial nitric oxide synthase; *GABRB3,* gamma-aminobutyric acid A receptor, beta 3; *HTR2A,* serotonin 2A receptor; *HTR3A,* serotonin 3A receptor; *HTR3B,* serotonin 3B receptor; *IL-4,* interleukin 4; *MAO-A,* monoamine oxidase A; *MAO-B,* monoamine oxidase B; *TAAR1,* trace amine-associated receptor 1); *TACR1,* substance P receptor; VNTR, variable number tandem repeat. NS, non significant.

GENOMIC SOURCES OF HUMAN VARIATION

The human genome sequence was completed in 2001 (134, 135). This was the start of a new era in human genetics, as the availability of a reference sequence constituted an incredible tool for researchers in order to evaluate genetic variants and their contribution to disease susceptibility. It was claimed that genetic differences between individuals accounted for less than 0.1% of the 3 million nucleotides in the human DNA sequence, and 1.4 million single nucleotide polymorphisms (SNPs) were detected. In the last years many efforts have been undertaken in order to identify and interpret the consequences of the variants present in the human genome. These variants refer to several sources of variation, including SNPs, copy number variants, small polymorphic indels (one or two nucleotides), VNTRs, among others. Nowadays, the most commonly studied variants in relation with susceptibility to complex diseases are SNPs, and, to a lesser extent, copy number variants (CNV), which are described below.

Single nucleotide polymorphisms

A SNP is a variation in the human genome affecting a single base pair that is present in at least 5% of a given population. SNPs can be classified in exonic, splicing and intronic, and exonic SNPs are further divided into non-synonymous and synonymous depending on whether they modify or not the amino acid sequence. Most SNPs are bi-allelic. Nowadays, new sequencing technologies have allowed the identification of over 15 million SNPs in the human genome (136).

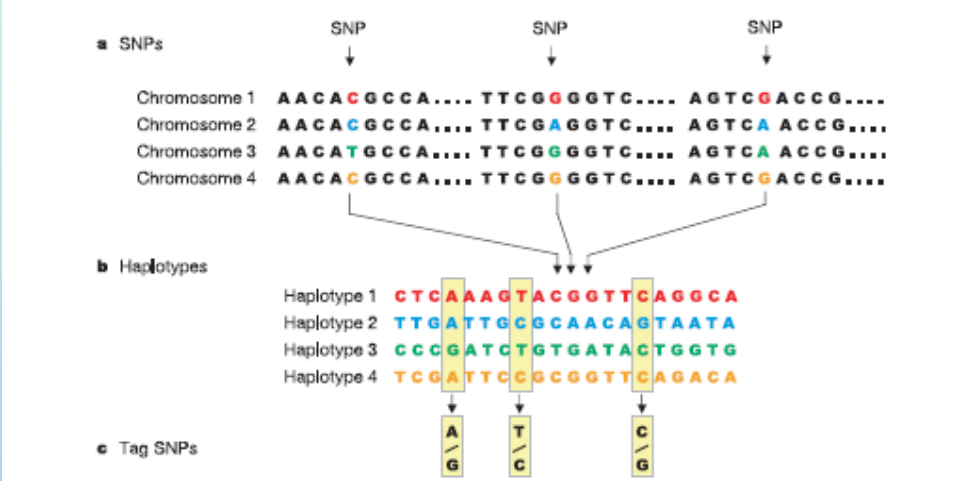
The HapMap project was pioneer into the detection of this kind of variants (Box 8). It was an essential tool to develop many association studies, since maps of linkage disequilibrium helped in the selection of the variants to genotype. Before the development of next generation sequencing technologies, it constituted a key resource for researchers to find genetic variants affecting health, disease and responses to drugs and environmental factors.

SNPs genotyping

In the last decade, several techniques have been developed in order to evaluate SNPs at different scales: from a single variant (Kaspar-Taqman Technology) to dozens and hundreds of variants (SNPlex, Massarray, and Veracode technology) and even millions of variants with array based platforms for genome wide association (GWA) studies (Table 5). Recently, some arrays are including variants with a frequency below the "polymorphic" 5% level: various companies have developed the exome arrays, including thousands of rare coding variants identified through next generation sequencing approaches. Also specific arrays such as the immunochip, which includes all the associations detected in genes related to the immune system (137) or Axiom® Genome-Wide Human Origins 1 Array which is used in population genetics. Finally it is also possible to design custom arrays evaluating thousands of SNPs.

Box 8: The HapMap Project

The international HapMap Project (http://hapmap.ncbi.nlm.nih.gov/) that started in 2002, was developed in order to build a haplotype map (HapMap) of the human genome through the study of SNPs. It was focused originally in 270 individuals coming from three populations: 90 samples (30 trios of two parents and a child) from a US Utah population with Northern and Western European ancestry (samples collected in 1980 by the Centre d'Etude du Polymorphisme Humain (CEPH)), 90 samples (30 trios of two parents and a child) collected from Yoruba people in Ibadan, Nigeria, 45 unrelated individuals from Tokyo, Japan and 45 unrelated individuals from Beijing, China. Currently, three phases have been completed including the genotyping of more populations and SNPs.



Four versions of a given chromosome are mostly identical and they only differ at a few bases, the SNPs, which are biallelic (a). Haplotypes (b) are defined by the particular combination of alleles at adjacent locations (loci) on the chromosome that are transmitted together. The genotyping of TagSNPs (c), which are in high linkage desequilibrium with the nearby SNPs, allows the identification of each of the four haplotypes without having to genotype all the variants in the region. (Taken from the International HapMap project (138)).

Table 5: Summary of some of the most commonly used technologies to evaluate SNPs.

Name	Variants explored	Amount of DNA	Technology
Taqman	1	10 to 20 ng	SNP Genotyping Assays include two allele-specific probes and a PCR primer pair to detect specific SNP targets. Probes contain distinct fluorescent dyes and include a non-fluorescent quencher that eliminates background fluorescence. Detection is achieved with proven 5' exonuclease chemistry by means of exonuclease cleavage of the 5' allele specific dye which generates the permanent assay signal (Figure 9). Following PCR amplification an endpoint read is performed on a thermocycler.
SNPlex	Up to 48	37 ng	Allele-specific oligonucleotide probes (ASO) and locus specific probes hybridize to the genomic target sequence. Linkers with universal PCR primer–binding sequences are ligated to the probes. A unique ZipCode sequence is attached at the 5' end of the genomic equivalent sequence within each ASO. This is followed by universal PCR reaction to amplify ligation products. Biotin-labeled PCR products are later on captured in streptavidin coated microtiter plates and finally detected by capillary electrophoresis (139).
Massarray	Up to 15 Up to 36 with SBE iPLEX Gold	150 ng (30 µl at 5 ng/µl) 300 ng(30 µl at 10 ng/µl)	Different sizes products depending on the allele are generated and subsequently detected with Mass spectrometry (MaldiTof). http://www.sequenom.com/sites/genetic-analysis/applications/snp-genotyping
Veracode **(array based)**	48 to 384	750 ng (15 µl at 50 ng/µl)	Glass microbeads (240 microns in length by 28 microns in diameter), with universal oligonucleotides are used. When a laser passes through the bead, it is defracted by the holographic elements creating a code image detected by a CCD camera. (http://www.illumina.com/technology/veracode_technology.ilmn)
GWA array **(array based)**	100000 to 2.5 millions	500 ng to 1 µg	Affymetrix technology DNA is digested, amplified by a PCR, fragmented, end labelled and hybridized against the array containing the probes (one for each allele) to test SNPs. Illumina technology (Infinium technology). An unlabeled DNA fragment is hybridized and subsequent enzymatic single base extension with a labelled nucleotide is performed. Dual-color fluorescent staining allows the labeled nucleotide to be detected by Illumina's iScan imaging system, which identifies both color and signal intensity. http://www.illumina.com/technology/infinium_hd_assay.ilmn

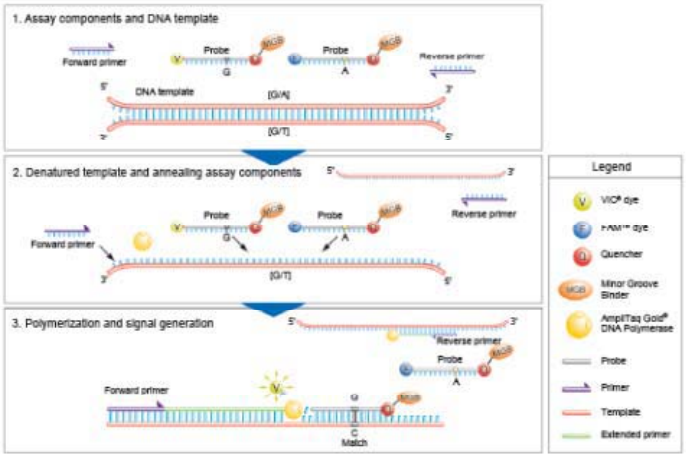


Figure 9: Taqman SNP genotyping system
(from http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/)

Genome-wide association studies

A genome-wide association study (GWAS) is an examination of many common genetic variants in different individuals to test if any variant is associated with a given trait. GWAS typically focus on associations between SNPs and traits or diseases. In contrast to methods, which specifically test one or a few genetic regions, GWAS investigate the entire genome. The approach is therefore said to be non-candidate-driven, in contrast to gene-specific, candidate-driven studies. GWAS identify SNPs that are associated with a disease, but cannot, on their own, specify which genes are causal. SNPs showing the strongest associations are genotyped in a replication cohort in order to confirm the association. This is performed with the different genotyping platforms existing.

The first strikingly successful GWAS was published in 2005 and investigated age-related macular degeneration. It found two SNPs which had a significantly altered when comparing with healthy controls (140). As of 2013, GWAS include hundreds or thousands of individuals; more than 1,400 human GWAS have examined over 200 diseases and traits; and over 8,000 SNP associations have been found (Figure 10). Several GWAS have received criticism for omitting important quality control steps, rendering the findings invalid, but modern publications address these issues. Since hundreds and thousands of samples are used, ensuring the homogeneity and the same populational origin for both cases and controls is a matter of concern. And this crucial aspect of these kinds of studies is becoming more and more important since it is necessary to merge different studies through meta-analyses in order to get enough power to detect variants with small effect. Several methods have been proposed to minimize the costs of genome wide association studies: using pools of cases and controls, sharing controls across populations. On the other hand an accurate consideration of

epidemiological characteristics of the study population and the selection of the most homogenous samples possible can increase the power or reduce costs under a variety of conditions (141). The paradigm of this is the success in the macular degeneration GWAS mentioned above (140), in which association to the complement factor H gene was detected with the inclusion of only 96 cases and 50 controls. This was due to the accuracy in the diagnosis, the homogeneity of the cases, and the fact that the association was very strong with a large effect (odds ratio of 7.4).

The Welcome trust Case Control Consortium (WTCCC) was the first to release large scale GWAS where thousands of SNPs were genotyped in complex disorders (Box 9). It was the start of the Genome Wide Association (GWA) studies era, with the use of very large cohorts of cases and controls. It constituted a great tool for researchers around the world, not only in terms of genotype data produced but also in the development of algorithms and statistical analysis methods for GWA studies.

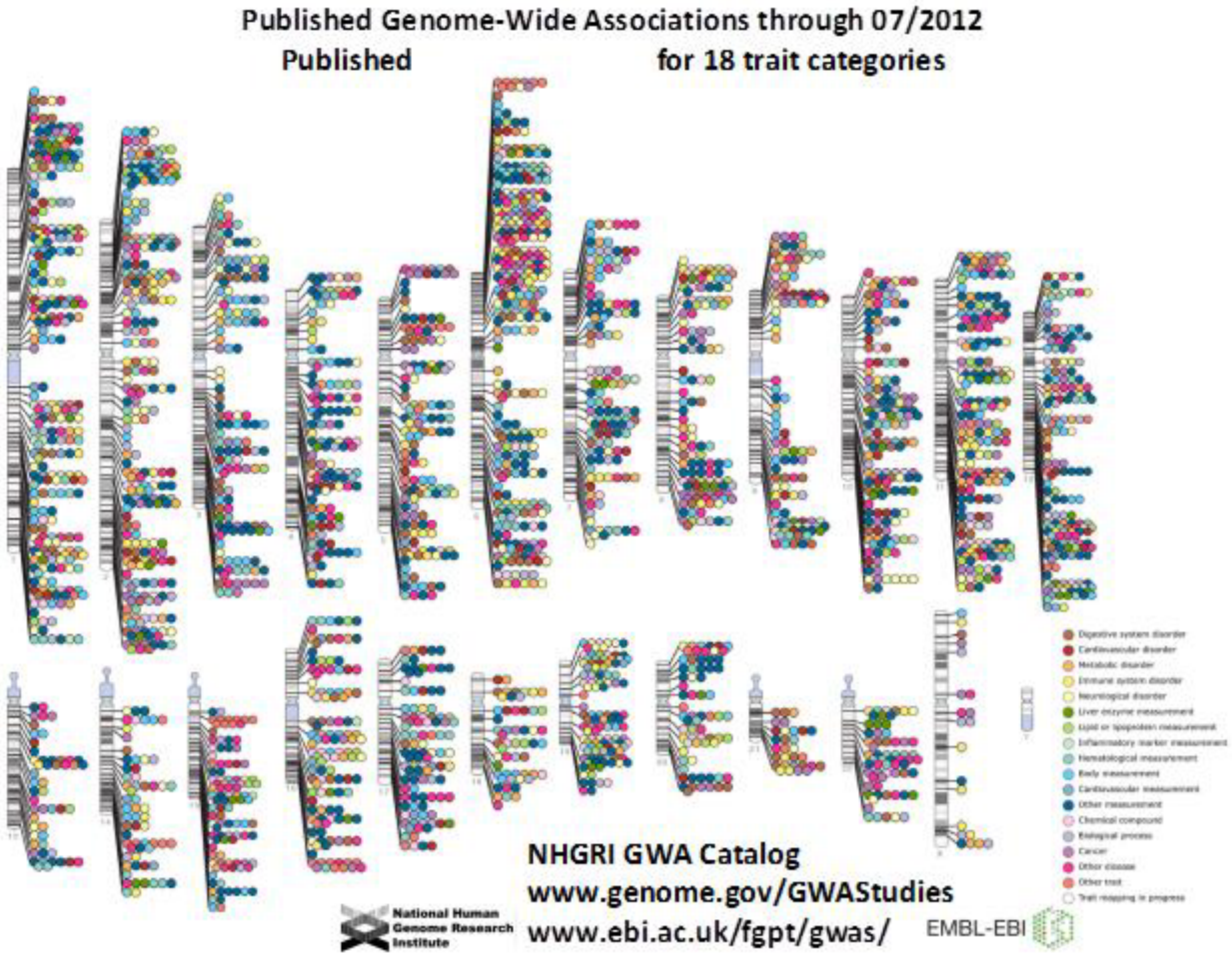


Figure 10: Chromosomal distribution of the hits of the 1617 GWAS published through July 2012.

Imputation

Genotype imputation is the process of predicting genotypes at positions that are not directly typed in a dataset (142). This is done based on linkage disequilibrium (LD) values between the genotyped SNPs and the SNPs identified in a reference panel. The most commonly used reference panels are HapMap and 1000genomes. The genotypes are imputed with uncertainty and a probability distribution over all three possible genotypes is produced (Figure 11). This uncertainty is taken into account in association analysis of the imputed data. Imputation can be used for fine mapping of signals detected from direct genotyping or to capture new associations. The development of imputation methods has enabled the capture of more variants without directly genotyping them. Around 10 million SNPs can be imputed by using reference panels, so that, the number of markers considered for association is considerably increased. If this should modify statistical significance threshold is still a matter of concern. Finally, many factors such as sample size should be considered as potential interfering factors in the imputation process, based on probabilistic genotypes.

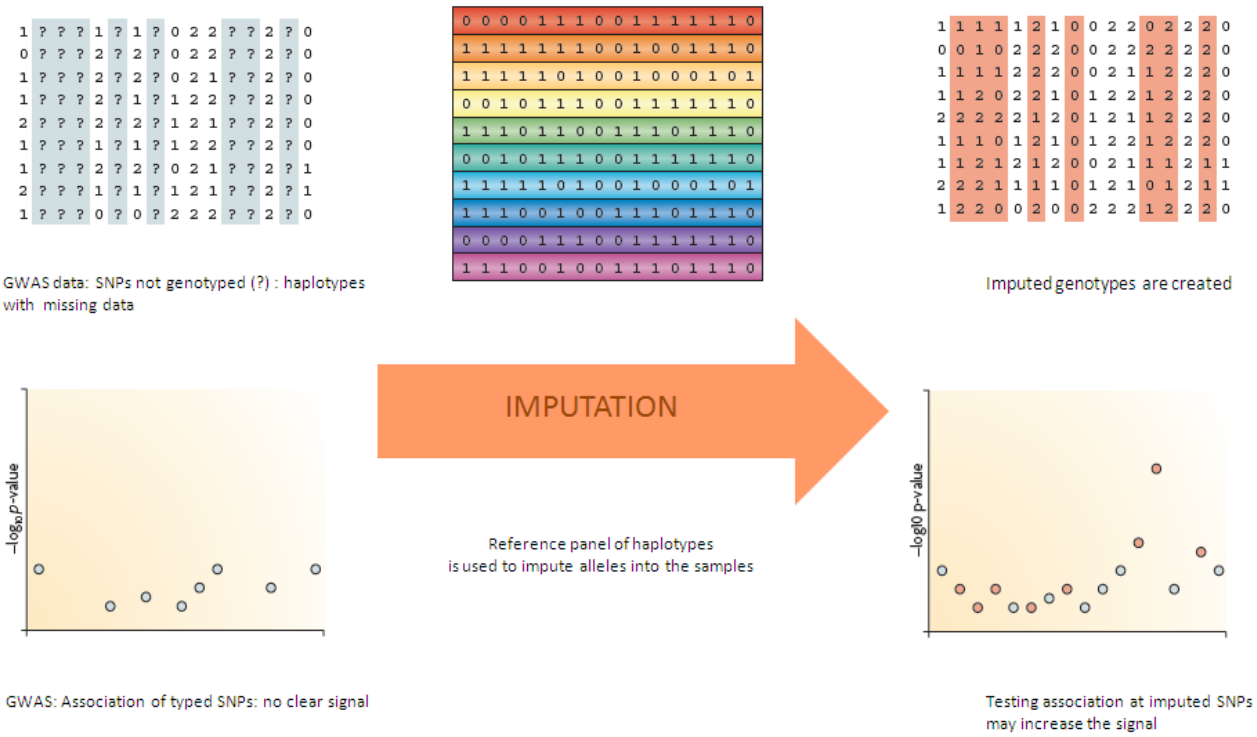


Figure 11: Overview if the imputation process (figure modified from Marchini *et al*) (142).

Box 9: The Welcome Trust Case Control Consortium

The Wellcome Trust Case-Control Consortium (WTCCC) (http://www.wtccc.org.uk/) is a group of 50 research groups across the UK which was established in 2005. The WTCCC aims were to exploit progress in understanding of patterns of human genome sequence variation along with advances in high-throughput genotyping technologies, and to explore the utility, design and analyses of genome-wide association (GWA) studies. The WTCCC has substantially increased the number of genes known to play a role in the development of some of our most common diseases and has, to date, identified approximately 90 new variants across all of the diseases analysed. In 2007, the WTCCC presented a GWAS of 14000 cases of seven common diseases and 3000 controls (143). WTCCC2 performed genome-wide association studies in 13 disease conditions analysing over 60,000 samples: Ankylosing spondylitis, Barrett's oesophagus and oesophageal adenocarcinoma, glaucoma, ischaemic stroke, multiple sclerosis, Parkinson's disease, psychosis endophenotypes, psoriasis, schizophrenia, ulcerative colitis and visceral leishmaniasis.

WTCCC3 is performing genome-wide association studies in the following 4 disease conditions: primary biliary cirrhosis, anorexia nervosa, pre-eclampsia in UK subjects, and the interactions between donor and recipient DNA related to early and late renal transplant dysfunction. The WTCCC3 will also carry out a pilot in a study of the genetics of host control of HIV-1 infection. Over 40,000 samples are being analysed using the Illumina 660K chip. The WTCCC3 uses the 6,000 control genotypes generated by the WTCCC2.

From SNPs genotyping to sequencing

In the recent years, next generation sequencing technologies have determined the beginning of a new era in human genetics. The increase in sequencing efficiency and proportional decrease in costs is enabling the use of massive sequencing in research projects, and therefore changing the scenario in the evaluation of genomic variants (mostly single base SNPs or rare variants) (Figure 12). Sequencing is replacing genotyping as the first step of the evaluation of variants linked to disease, since it allows not only the detection of common known SNPs but also the identification of novel mutations. Especially since the development of exome sequencing, many genetic studies are now being performed based on these techniques, mostly for the study of Mendelian disorders, as the identification of causative genetic variation in these types of disorders is much easier. These technologies are also being used for complex diseases, but their application into complex disorders will need the development of statistical tests applicable to rare variants. The 1000 genomes project (Box 10) has constituted the largest project so far to apply next generation sequencing. It can be defined as the equivalent of the HapMap project with the use of next generation sequencing. It combines both whole genome sequencing at different depths of coverage and exome sequencing in individuals from different populations. All the data produced by this project can know be used as a reference for researchers around the world. For instance, data from the 1000 genomes project is being used as a

reference panel for GWAS imputation, and also to take into account variant frequencies in performing data analysis for exome studies in several disorders. In a decade, the reference genome has moved into 1000 references. This perfectly illustrates the revolution that is living human genetics.

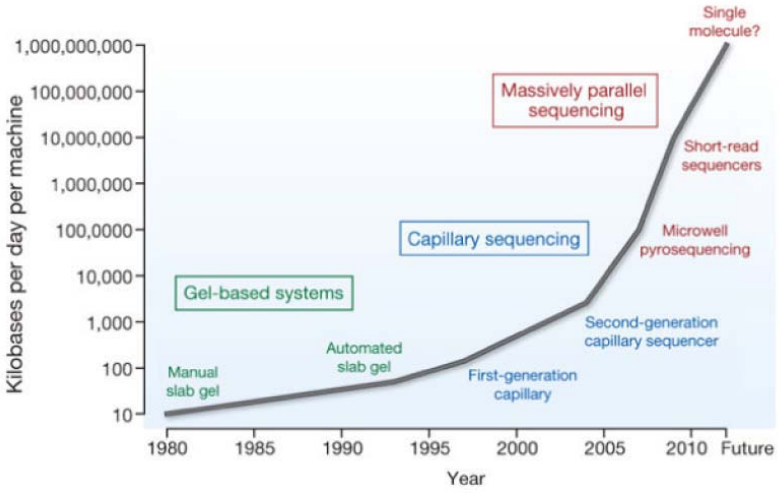


Figure 12: Evolution of the DNA sequencing technologies over the past 30 years and subsequent improvement in the rate of sequencing (taken from Stratton *et al* (144)).

Box 10: The 1000 genomes project

The goal of the 1000 Genomes Project (http://www.1000genomes.org/) is to find genetic variants that have frequencies of at least 1% in the populations studied. To this purpose, low coverage whole genome sequencing of 179 individuals from four populations, high-coverage sequencing of two trios and exome sequencing of 697 individuals of seven populations have been performed in the pilot phase (The 1000 Genomes Project Consortium). These three experimental approaches are complementary as they provide different information. The trio design allows an accurate discovery of multiple variants across the genome with Mendelian transmission helping in genotype estimation, inference of haplotypes and quality control. The low-coverage project identifies shared variants on common haplotypes. Finally, the exome design enables accurate discovery of common and rare variants but only in the exome. At the time of writing, 2500 samples had been sequenced.

Copy Number Variants

In the last years, several publications have analyzed genomic structural variation, including insertions, deletions, translocations and deletions of genetic material, as a “new” major form of polymorphism (145). Although the existence of structural variation has been known for a long time, this type of variation was considered to consist mostly of rare individual events, and their real importance remained unknown until

the last few years. One type of structural variation are copy number variants (duplications, insertions, deletions): these were initially defined as DNA segments larger than 1 kb and up to several megabases (Mb) that can be found in a variable number of copies in the genome. This size restriction is no longer considered in CNV definition, as the higher resolution of the techniques has allowed the detection of smaller events (146). CNVs can be simple or complex, where the fragments duplicated vary in size between individuals, involving several subregions (Figure 13). Over 15% of the human genome may be affected by copy number variation (147, 148). According to the database of genomic variants (http://projects.tcag.ca/variation/), in November 2012, 291,801 CNVs had been described in the human genome (Figure 14). Several mechanisms are involved in the generation of CNVs. In most cases, copy number variants are generated by non-allelic homologous recombination (NAHR), but other mechanisms involve non-holomogous end-joining, fork stalling and template switiching (FoSTeS)/ microhomology-mediated rearrangements (MMBIR) (149) and Line1- mediated retrotransposition as reviewed in Malhotra *et al* (150) (Figures 15 and 16).

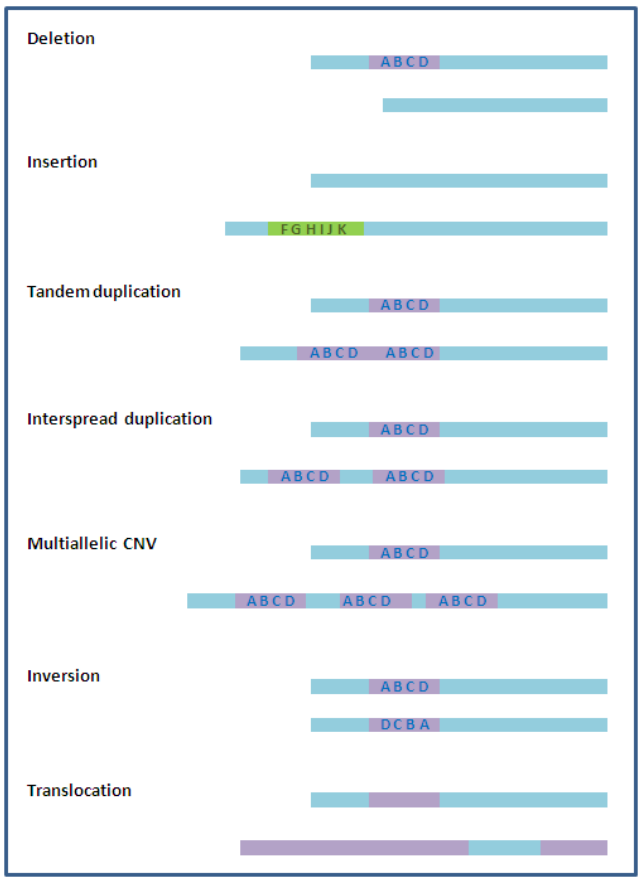


Figure 13: Types of structural variants. CNVs include deletions, insertions, tandem duplications and unbalanced translocations (with loss or gain of genetic material).

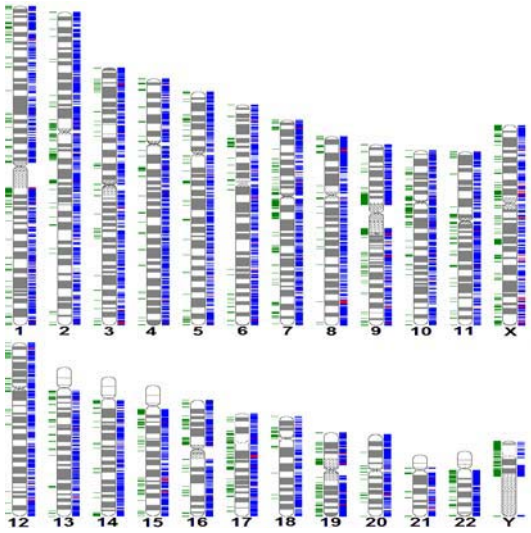


Figure 14: Chromosomal distribution of CNVs identified by November 2010. Blue bars indicate reported CNVs; Red bars indicate reported inversion breakpoints; Green bars to the left indicate segmental duplications (taken from http://projects.tcag.ca/variation/).

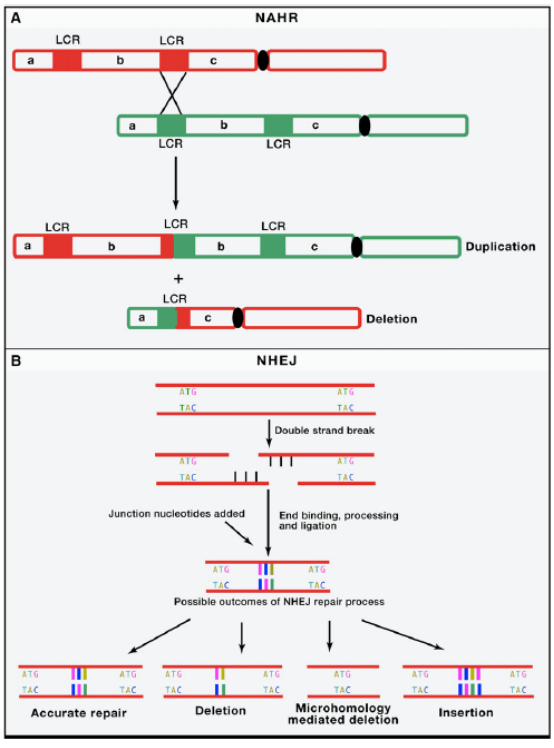


Figure 15: Mechanisms underlying CNV formation 1 (figure taken from Malhotra *et al* (150)). A) Non-allelic homologuous recombination (NAHR) occurs over flanking segmental duplications resulting in deletion and duplication. B) In Non-homologous end-joining (NEHJ) double strand break are repaired though several rounds of enzymatic activity that can lead to an accurate repair, deletions of different sizes and, to a less extent, insertions.

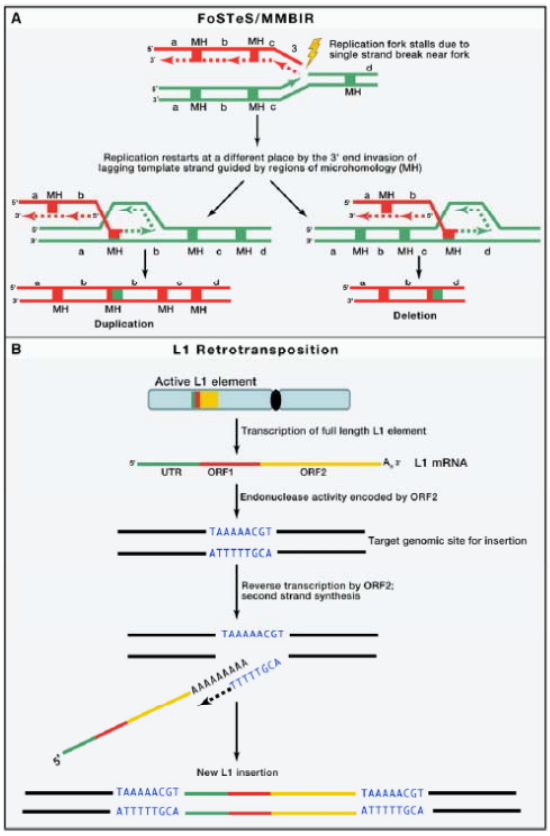


Figure 16: Mechanisms underlying CNV formation 2 (Figure taken from Malhotra *et al* (150)). A) FoSTeS/MMBIR: when a replication fork encounters a nick (shadow arrow), one arm of the fork breaks. The 3' single-strand end of lagging-strand invades the sister leading-strand guided by regions of microhomology (MH) forming a new replication fork. Whether the template switch occurs in front of or behind the position of the original collapse determines a deletion or duplication. B) LINE-1 retrotransposition. L1 is transcribed and translated leading to a L1 ribonucleoprotein (RNP). L1 RNP is transported to the nucleus and retroranspostion occurs by target site primed reverse transcription (TPRT). During TPRT L1 endonuclease activity nicks genomic DNA exposing a free 3' that serves as a primer for reverse transcription of the L1 RNA. This results in the insertion of a 5' truncated L1 copy flanked by target site duplications. ORF, open reading frame.

Despite most CNVs are thought to be benign polymorphisms, several studies have already proven CNVs may have a pathogenic role in the aetiology of disorders such as Charcot Marie Tooth neuropathy type I (151) and they have also been shown to contribute to complex diseases' susceptibility such as HIV, Crohn's Disease, Systemic Lupus Erythematosus or psoriasis (152-155). However, we have to take into consideration that the largest attempt to evaluate CNV role in complex disorders, in which aCGH from 16,000 cases of eight common disorders and 3,000 controls were analyzed, failed to detect any association (156). The most obvious explanation for the effect of this kind of variants can be the alteration of gene dosage, but they also may disrupt genes, uncover deleterious alleles, interrupt transvection effects or have positional effects on the internal or surrounding genes, such as distancing a gene from its regulatory region (157).

Several approaches are used in order to study this kind of polymorphisms at a genome-wide scale. The most direct analysis consists of array comparative genomic hybridization (aCGH), which allows for detection of gains or losses of genetic material in comparison to a reference sample. Also, given the amount of information generated by SNP arrays for whole genome association scans, several algorithms have been developed to identify CNVs based on intensity evaluation. Finally, new methods, such as paired-end mapping (158) combine high-throughput sequencing and computational analysis in order to detect structural variation. In any case, the results obtained by these methods have to be validated with quantitative methods at individual or multiplex loci, such as quantitative real-time PCR, multiplex ligation-dependent probe amplification or FISH.

Whole genome methods to detect CNVs

aCGH

In aCGH, DNA from a test sample and normal reference sample are labelled with different fluorophores, and hybridized to an array containing several thousand probes across the genome. The fluorescence intensity of the test and of the reference DNA is then measured, to calculate the ratio between them and subsequently the copy number changes for a particular location in the genome (Figure 17). A dye-swap experiment, in which the fluorophores are exchanged between the test and the control sample, is used in order to eliminate the fluorophore bias.

Initially, aCGH was performed with bacterial artificial chromosomes (BAC) that had a resolution of about 46 kb (159). Currently, the arrays used are more dense: for example, Agilent 2X400K microarray has about 420,000 60 bp-long probes with a median probe spacing of 1,1 KB and an average probe spatial resolution of 6.6 kb, covering the entire genome, with 84.8% of CNVs <10 kb being covered by at least one probe. This has allowed the detection of smaller events.

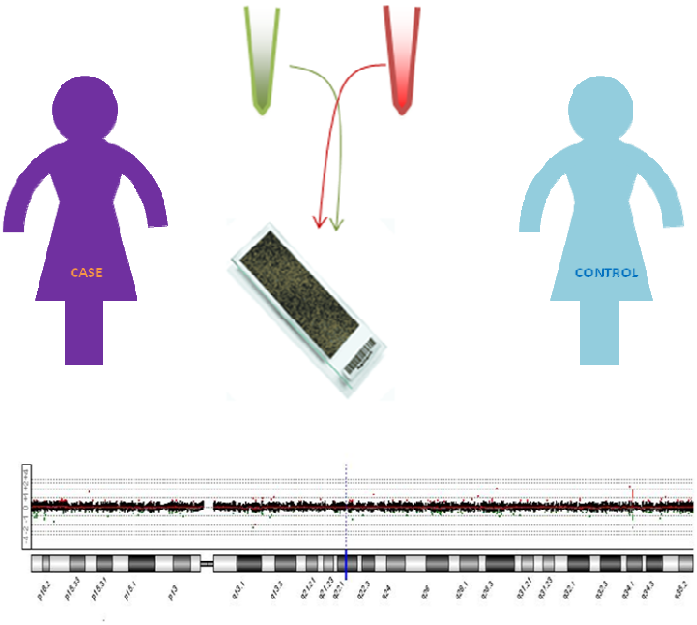


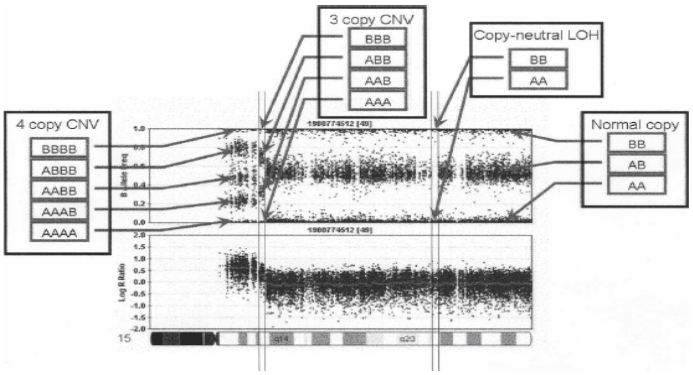
Figure 17: aCGH experiment. The same amount of case DNA and control DNA are used, labelled with different fluorfores and hybridized against an array. Regions showing different ratios of the intensities are regions varying in CNV.

As the results obtained are relative to the control sample, the selection of the control sample is important. In most studies of human CNVs, DNA from a male of European ancestry, NA10851, has been used as a reference for aCGH (160). For this reason, Ju *et al* performed whole-genome sequencing and aCGH of NA10851 in order to characterize its CNVs (160). A strategy used to minimize reference bias, is to pool samples. By hybridizing a pool of cases against a pool of controls, interindividual differences may be diluted and differences in copy number due to disease may be enhanced. This approach was successful in detecting a new CNV associated with psoriasis susceptibility (155). Although pooled DNA may resolve reference biases it may also decrease the power of CNV detection in highly polymorphic regions (160) and lead to false positive results.

SNP arrays algorithms: PennCNV

CNVs can also be inferred from SNP based arrays. Several algorithms have been developed, using Bi allele frequency (BAF) and log R ratio (LRR) to detect, in a given sample, regions that show an increase or decrease in genomic material in respect to the others. Log R ratio is the ratio of intensity of the two alleles of each SNP compared to the background total fluorescent intensity, and the BAF is the relative ratio of the fluorescent signals between the two probes/alleles at each SNP position. PennCNV (161) is possibly the widest used algorithm for CNV identification from SNParray data. It uses a Hidden Markov Model algorithm that integrates both LRR and BAF to infer CNVs leading to the identification of six different copy number

states (Figure 18). This offers the advantage that information is produced from already generated data, but its principal limitation is that the resolution and reliability of the CNV genotype is less than those offered by aCGH.



Copy no. state	Total copy no.	Description (for autosome)	CNV genotypes
1	0	Deletion of two copies	Null
2	1	Deletion of one copy	A, B
3	2	Normal state	AA, AB, BB
4	2	Copy-neutral with LOH	AA, BB
5	3	Single copy duplication	AAA, AAB, ABB, BBB
6	4	Double copy duplication	AAAA, AAAB, AABB, ABBB, BBBB

Figure 18: PennCNV allows the identification of six copy number states based on LRR and BAF (figures taken from Wang *et al* (161)).

The overlap between the different existing software to detect CNVs from SNP array data is poor (Table 6) (162). For this reason, in order to identify possible CNV events related to disease, many authors propose to implement several algorithms and only consider for validation the overlapping CNVs (163).

Table 6: Comparison of SNP array based software for CNV prediction (taken form Winchester *et al* (162)

		Published results (Redon)	Birdsuite Affymetrix	CNAT Affymetrix	CNV Partition Illumina	GADA Affymetrix	GADA Illumina	Nexus Affymetrix	Nexus Illumina	PennCNV Affymetrix	PennCNV Illumina	QuantiSNP Affymetrix	QuantiSNP Illumina
Published data (Redon)			17 (4%)	4 (40%)	3 (19%)	32 (5%)	2 (2%)	11(10%)	2 (25%)	12 (18%)	7 (16%)	18 (9%)	8 (13%)
Birdsuite	Affymetrix	17 (44%)		9 (90%)	13 (81%)	135 (22%)	21 (24%)	62 (56%)	6 (75%)	43 (64%)	20 (47%)	97 (50%)	20 (33%)
CNAT	Affymetrix	4 (10%)	15 (4%)		4 (25%)	34 (6%)	0	23 (21%)	1 (13%)	13 (19%)	2 (5%)	17 (9%)	5 (8%)
CNV Partition	Illumina	3 (8%)	16 (4%)	4 (40%)		37 (6%)	7 (8%)	20 (18%)	7 (88%)	9 (13%)	11 (26%)	16 (8%)	16 (27%)
GADA	Affymetrix	17 (44%)	106 (28%)	9 (90%)	13 (81%)		32 (37%)	91 (82%)	7 (88%)	58 (87%)	23 (53%)	153 (79%)	27 (45%)
GADA	Illumina	2 (5%)	96 (25%)	0	13 (81%)	208 (34%)		25 (23%)	2 (25%)	26 (30%)	17 (40%)	67 (35%)	23 (38%)
Nexus	Affymetrix	7 (18%)	57 (15%)	10 (100%)	7 (44%)	116 (19%)	8 (9%)		4 (50%)	45 (67%)	15 (35%)	78 (40%)	17 (28%)
Nexus	Illumina	2 (5%)	6 (2%)	1 (10%)	7 (44%)	22 (4%)	2 (2%)	4 (4%)		6 (9%)	7 (16%)	10 (5%)	9 (15%)
PennCNV	Affymetrix	11 (28%)	51 (13%)	10 (100%)	9 (56%)	105 (17%)	10 (11%)	65 (59%)	6 (75%)		19 (44%)	71 (37%)	21 (35%)
PennCNV	Illumina	6 (15%)	25 (7%)	2 (20%)	11 (69%)	44 (7%)	9 (10%)	23 (21%)	6 (75%)	18 (27%)		26 (13%)	28 (47%)
QuantiSNP	Affymetrix	14 (36%)	97 (25%)	10 (100%)	10 (63%)	199 (32%)	18 (21%)	86 (77%)	7 (88%)	65 (97%)	21 (49%)		24 (40%)
QuantiSNP	Illumina	6 (15%)	14 (4%)	5 (50%)	15 (94%)	55 (9%)	10 (11%)	30 (27%)	8 (100%)	23 (34%)	32 (74%)	31 (16%)	

High-throughput sequencing: paired-end and exome algorithms

The development of Next Generation Sequencing (NGS) technologies, has lead to the implementation of different algorithms in order to evaluate CNVs (164) from NGS data. In particular, pair-end sequencing is based on the sequencing of adapters ligated to both ends of inserts of different sizes. Adapters sequences are afterwards mapped to the reference genome, allowing the detection, with high resolution of insertions, deletions and inversions (158). Also algorithms that combine the use of depth of coverage and distance between adapters sequence data are used to define regions varying in copy number, both in exome and whole genome sequencing data. When completely developed and validated, these methods will offer the highest resolution in terms of breakpoints definition and copy number state.

Validation of whole genome methods

Various validation strategies have been applied to subsets of putative variants in each of the discovery reports. These included FISH of metaphase, interphase or fiber chromosomes using various clones or PCR-amplified molecules; PCR or quantitative PCR (qPCR) for allele loss or quantitative variation; multiple ascertainment, whereby considerable weight was given to whether or not a putative variant was seen in more than one individual or had been reported in previous studies; array CGH to validate computational screening results, or for finest resolution of BAC-screening results by oligonucleotide arrays; sequence analysis of fosmid inserts to confirm calls and to assess some discordant ones; allele-specific fluorescence intensities and familial clustering (165).

Multiplex PCR

This validation approach for aCGH results can only implemented in the cases of insertion-deletion polymorphisms with only three possible states: 0 (homozygous deleted), 1 (heterozygous) or 2 (homozygous non deleted) copies (Figure 19). First, in order to identify the breakpoints of a CNV region detected by a screening method, a long range PCR can be performed. The development of new technologies with higher resolution has facilitated this analysis, by performing a computational analysis using the coordinates resulting from the aCGH or sequencing results and checking for known CNVs in the available catalogues of structural variants (Database of genomic variants among others). PCR primers are designed outside the deleted region, which will only amplify in a deleted sample when the size of the CNV is larger than 3-4 kb, or for smaller insertion/deletions will give a smaller than expected fragment for the deleted allele. Sequencing this PCR product and blasting it to the reference genome will detect the breakpoints. In particular, in the case of the 2X400K array, since the design of this array is based on the common CNVs detected by Conrad *et al* (146), the PCR primer design and breakpoint detection is even easier, by using the supplementary data from the paper, especially if the CNV has been experimentally validated. Once the breakpoints are defined, it

is possible to design a multiplex assay, using two sets of primers. A pair of primers located inside the deleted region, which will only amplify in the non-deleted allele, and another set of primers located outside the deleted region, which will only amplify in the deleted allele. This will allow the genotyping of cases and controls in a replication cohort in order to confirm if the difference in CNV frequencies detected with the aCGH is associated to the studied phenotype.

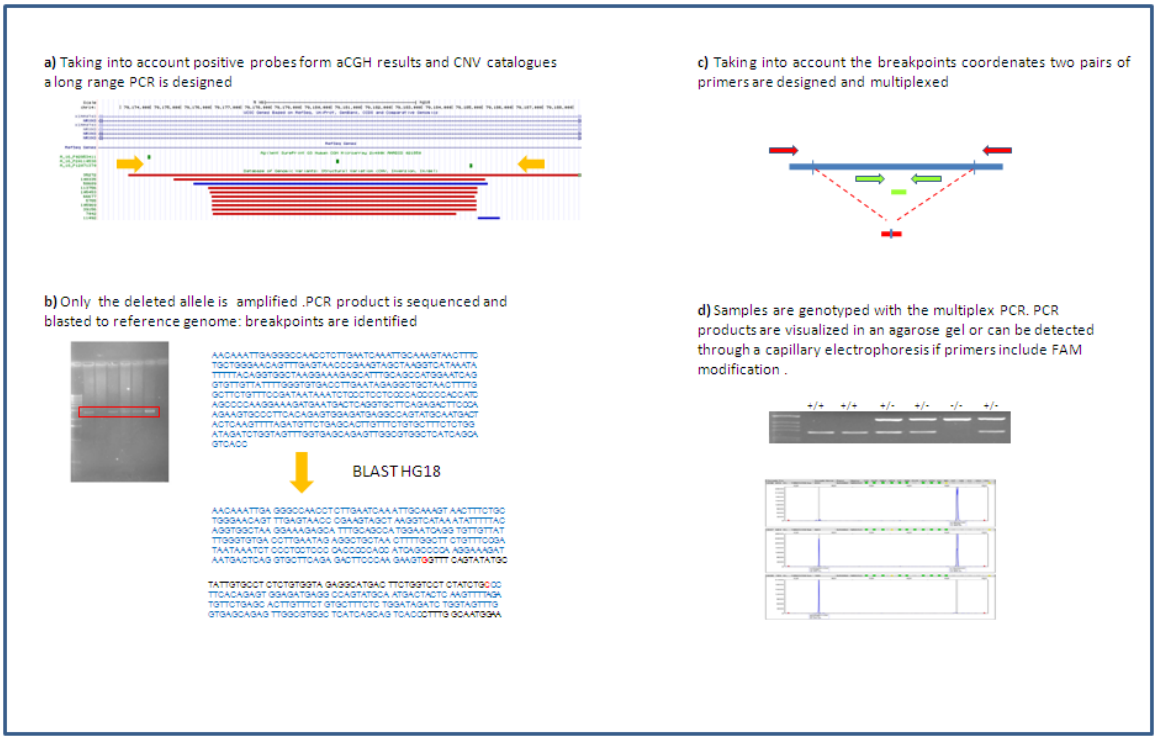


Figure 19: From aCGH results to genotyping of a replication cohort with a multiplex PCR.

Real time quantitative PCR

In the cases of complex CNVs, in order to quantify the number of copies both in cases and controls, it is necessary to perform quantitative PCR. This type of validation refers the number of copies in any individual to a reference sample. The most commonly used technique to quantify is Taqman® (Applied Biosystems). In this a pair of specific primers plus a specific probe with a fluorophore are used. Fluorescence is only produced when the probe is cleaved with the PCR extension and the quencher is detached (Figure 20). In addition to the region that we want to test, a region that does not vary in copy number is used in order to normalize the DNA amount. Roche® Universal probe library uses the same technology.

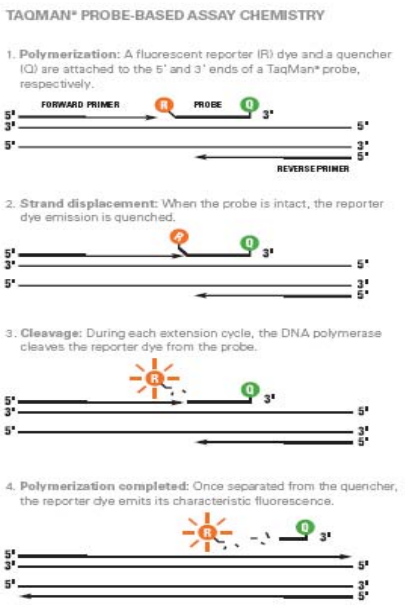


Figure 20: Taqman® technology (http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/real-time-pcr/taqman-and-sybr-green-chemistries.html)

Other validation techniques

When the CNV events are large enough, visualization through fluorescence in situ hybridization (FISH) can be used in order to validate aCGH results. It is based on the use of fluorescent probes that bind specifically to complementary regions in DNA of metaphase cells and are targeted by antibodies for their detection. The resolution of the techniques is >3 Mb. For a higher resolution, FiberFISH allows the visualization of smaller events, 1-400 kb (166), by stretching out interphase chromosomes in a straight line, by applying mechanical shear along the length of the slide. However, these visualization techniques of CNV events have the limitation that they cannot be performed in a high throughput manner and that cells are needed in order to perform the experiments. Their use will be more centered in family studies or merely as a secondary validation of the CNV.

MLPA (multiple ligation-dependent probe amplification) is also a widely used technique. It allows the detection of insertions and deletions in a multiplex way (50 loci can be tested in the same experiment) and using a low amount of DNA (20 ng) (167). However, the reproducibility of the technique, when using custom assays, is not high and reduces the cost benefit ratio of the technique.

GENETICS UNDERLYING COMPLEX DISORDERS SUSCEPTIBLITY

Where are we?

The genetic model underlying susceptibility to complex disorders is still a matter of discussion. Different hypothesis try to explain how genetic factors play a role in the development of these disorders. The most accepted explanation these days includes both rare variants with a large effect and common variants with a smaller effect. The exploration of these kinds of variants presents different difficulties. In the case of common variants with a small effect, it is essential to work with very large cohorts in order to have enough power to detect differences among cases and controls. This is somehow a greater deal in heterogeneous disorders such as FM. The selection of the study subject's study is crucial, and another critical issue is phenotype accuracy. On the other hand, rare variants will probably be a new focus of analysis, as next generation sequencing technologies, such as exome sequencing, have facilitated their detection. One of the limitation in the analysis of these variants is determined by private variants: how to identify variants linked to disease from private variants that can be somehow common by chance to a group of individuals?

The approaches used in human genetics for the study of genetic susceptibility to diseases are several and have incredibly evolved over the past years. And this evolution has been in concordance with the development of new technologies allowing a more efficient evaluation of the genetic variants. Linkage studies in families have been widely used. First using microsatellites, then single nucleotide polymorphisms genotyped on arrays and, in the last years, by analyzing exome data. In complex diseases, the most common approach, above family based transmission disequilibrium, has been case control association studies. Initially it was focused on the study of variants in candidate genes, then on GWAS and, more recently, by performing exome sequencing, although the results of the latter have not been published yet, as the analysis is complex. And this is going to get more and more complex as these new techniques and the algorithms that are being developed in order to analyse and interpret the results are now focusing to the analysis of other variants such as copy number variants. What is more, the possibility to explore epigenetic factors such as DNA methylation and histone modifications opens a new field of research adding a new complexity level to the study of complex disorders genetic susceptibility. Finally, the Encode project has enhanced the importance of other non-coding sequences in human genome.

Large cohorts and population admixture

In most GWAS one of the principal limitations is population heterogeneity. By increasing the number of individuals included in a study, the population of study get more heterogeneous and then ancestry may become a confounder. This situation gets dramatic in populations such as that of the United States in which

there is extensive population admixture. It is essential that the confounding effect caused by this population admixture be considered when performing the association analysis. STRUCTURE analysis using Ancestry Informative Markers (AIMs) scores would be most preferable for this kind of study compared to self-reported ethnicity. AIMs scores should be used as covariates in association analyses.

Population admixture is a major issue in genetic association studies. However, the Iberian Peninsula shows a substantial genetic homogeneity (168), only showing questioned differences among the Basque (169) and Canary island inhabitants (170). A recent paper by Julia *et al* (171), in which a whole genome scan was performed in Spanish controls and rheumatoid arthritis samples, only showed a west to east trend, but correction for ancestry informative markers in that study did not show a significant difference from the uncorrected data, not supporting an effect of population stratification.

Power calculation and correction for multiple testing

In GWAS one of the principal limitations is power. For complex disorders, genetic factors are expected to have a limited effect. In fact, not considering the effect of the major histocompatibility complex in autoimmune disorders, most GWAS results show associations with odd ratios lower than 1.5. Moreover, in order to detect these mild associations thousands of samples are needed. Several software can be used to perform a statistical power calculation, as for example Quanto (http://hydra.usc.edu/gxe/), and predict the sample size necessary to reach statistical significance for low effects variants.

An important issue in the human genetic studies nowadays is the correction for multiple testing. When different outputs are maesures/evaluated, finding an association with a p-value 0.05 can be obtained by chance. For this reason and especially when performing whole genome approaches a correction for multiple testing is performed. The most astringent one is the Bonferroni in which the level of significance is established as 0.05 is divided by the number of tests performs. As in some case some of the markers evaluated can be linked, a less astringent correction may be performed. For whole genome association scans the level of significance has been established at 10E-8.

Dealing with clinical data: Cluster analysis

Cluster analysis is a multivariate statistical technique, which evaluates the degree of similarity among heterogeneous variables to try to identify related groups of variables based on their similarities. This procedure attempts to reduce data dimensionality to fewer latent variables. It has been applied to other disorders in an effort to understand the relationships among clinical features and outcome variables (172). It was first applied in the oncology field to group the secondary effects of chemotherapy that apparently were not related at all since they were manifestations of different organs (for example: vomiting, aenemia and

fever). It has been later implemented in order to link related symptoms or to identify groups of homogenous patients of a particular disease.

Many investigators have used cluster analysis to group FM symptoms, in order to define FM subgroups. These previous studies had different results, depending on sample sizes, variables studied and methods used**.** In most cases, cluster analysis was used to categorize FM patients, based on somatic or psychological symptoms (173), quantitative sensory testing (174) or pressure-pain thresholds and psychological factors (175). A recent study tried to discern clinically relevant subgroups across psychological and biomedical domains (176), while another attempted to identify clusters of clinical features meaningful to FM patients that corresponded to their treatment priority goals in the context of desired improvement (177) . Finally, a more recent work (178) indicated the existence of two latent dimensions underlying FM symptomatology: FM core symptoms and distress. Most of these studies have been performed in small cohorts taking into account few clinical variables, and using clustering to group either variables or patients. Only two studies were performed in large cohorts (179, 180). In these studies, patients were assessed through web based method and their purpose was, in one case, to evaluate patients perception of symptoms management (179) and, in the other case, to examine differences among FM subgroups in healthcare utilization (180). So far, in spite in the different implementation of clustering procedures into FM patients, no clear FM subgroups have been identified.

OBJECTIVES

The aim of this thesis was to elucidate genetic susceptibility factors for fibromyalgia, a common, heterogeneous disorder whose genetic contribution had not been explored in a comprehensive manner. We assessed this objective through three main approaches:

1. The identification of FM clinically homogeneous subgroups with a two step cluster analyses:
    a) Construction of dimensions from clinical variables, and
    b) Identification of FM subphenotypes.

2. To perform a genome-wide association study of FM in order to evaluate the possible contribution of single nucleotide polymorphisms and also to infer regions varying in copy number and their potential presence in mosaicism. This includes the following subobjectives:
    a) Association analysis of GWAS data,
    b) Identification of CNVs,
    c) Replication of strongest associated SNPs, and
    d) Implementation of cluster analysis into the GWAS study.

3. To perform array comparative genomic hybridization experiments to identify regions varying in copy number that could be involved in FM susceptibility:
    a) aCGH experiments analysis and validation,
    b) Implementation of cluster analysis into aCGH analysis, and
    c) Evaluation of possible functional consequences of associated CNVs.

To perform these analyses we had access to a large and very well characterized cohort of FM patients.

MATERIALS AND METHODS

SAMPLES: THE FIBROMYALGIA BIOBANK

We studied unrelated FM samples of a multicentric study, whose first purpose was the creation of the Spanish Genetic and Clinical Data Bank of Fibromyalgia and Chronic Fatigue Syndrome, (FFSGCDB). Five FM Units of five Spanish Hospitals (Hospital del Mar, Barcelona, Jordi Carbonell; Hospital Clínic i Provincial, Barcelona, Antonio Collado; Hospital de la Vall d'Hebrón, Barcelona, Jose Alegre; Instituto General de Rehabilitación de Madrid, Madrid, Javier Rivera; and Hospital General de Guadalajara, Guadalajara, Javier Vidal) participated in the collection of samples. A total of 1,510 FM patients, fulfilling 1990 ACR criteria, were selected by rheumatologists. They were then evaluated by another set of physicians trained in the evaluation of FM patients. All samples were Spanish of Caucasian origin and had signed informed consent before enrolment. The ethics committee at all recruitment centers approved the project. They all passed the same questionnaires and physical examination.

Data collection followed a standard protocol of questionnaires and physical examination that were recorded by principal investigators of the FFSGCDB. It included collection of demographic variables (age, marital status, educational level, and occupational status), family and personal history of diseases (in particular, history of chronic fatigue, chronic pain, connective disorders, spine degenerative disease and psychopathology), time of disease evolution, presence of representative symptoms covering the wide spectrum of symptomatology in FM (musculoskeletal, neurological, autonomic and somatic), and treatments. Core measures of FM severity were assessed by different Spanish validated scales: the intensity of pain and fatigue with a 11 points visual analogue scale (VAS) (where 0 represented no pain or fatigue, and 10 the maximum pain or fatigue); the number of tender points with standard manual examination (181); the level of anxiety or depression with the Hospital Anxiety and Depression Scale (HAD) (182, 183); the sleep quality by the Pittsburgh Sleep Quality Index (PSQI) (184); and the general Health Status by the Fibromyalgia Impact Questionnaire (FIQ) (185), the Fatigue Impact Severity scale (FIS) (186, 187) and the Quality of Life survey (SF-36) (188, 189). Finally, patients were evaluated for Fukuda's CFS criteria (190).

1,000 control samples came from the National DNA Bank of Salamanca). They had low levels of pain and fatigue as assessed by a questionnaire. We had access to 5 μg of DNA of FM cases and 1 μg of DNA of the control samples.

FM CLUSTERS: IDENTIFYING FIBROMYALGIA SUBGROUPS

A two-step clustering procedure was performed in 1,446 FM patients in order to identify FM subgroups (Figure 21). After exclusion of treatment and socio-demographic variables, 48 variables were selected for the clustering analysis. Given the mixed nature of the variables, these were transformed into binary types (0 = mild; 1 = severe). For symptoms (dichotomic variables), the presence of the symptom was codified as 1 and

the absence as 0. For continuous variables, the median was considered as a cut-off value. For scales, values below the median were codified as 0 and values above the median as 1, while for age of onset, as a younger age of onset (≤38 years) is considered more severe, the codification was reversed. Variables included in the cluster analysis, as well as their medians (interquartile range), are summarized in Table 7.

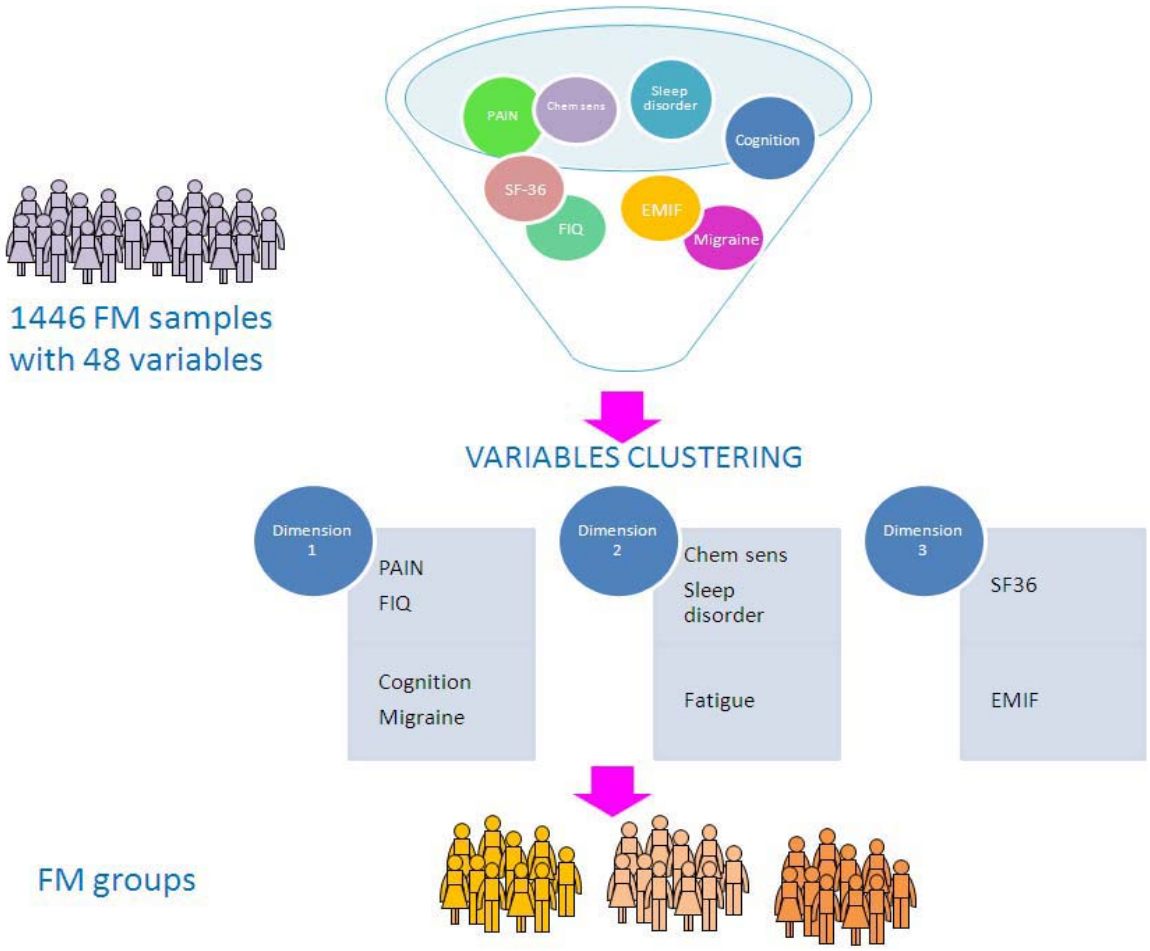


Figure 21: Identification of FM subgroups through a two-steps clustering procedure. First, variables were classified into different dimensions, based on variables' similarities. In a second step, variables' dimensions were used to identify FM subgroups.

Statistical analysis

Building variable's dimensions

The underlying dimensions of FM were evaluated using *partitioning cluster analysis*. This is a method to partition data (clinical features) into meaningful subgroups when the number of subgroups and other information about their composition may be unknown (191). In this sense, the procedure attempts to reduce data dimensionality to fewer latent variables, which, in our study, are termed FM dimensions.

Table 7: Variables included in the cluster analysis. They are listed in alphabetical order.

VARIABLES
Adjustment disorder
Age of onset (38;p_{25}: 30; p_{75}: 45)
Blackouts
Concentration problems
Connective disorder
Dizziness
Excessive Perspiration
Facial oedema
Family history chronic pain
Family history of autoimmune disorders
Family history of chronic fatigue syndrome
Family history of fibromyalgia
Fatigue Impact Scale (FIS) (66; p_{25}: 56.50; p_{75}: 75.00)
Fatigue level (VAS 1-10 cm) (8; p_{25}: 6.4; p_{75}: 9)
Fibromyalgia Impact Questionnaire (FIQ) (74.66; p_{25}: 63.05; p_{75}: 84.25)
Forgetfulness
HAD anxiety subscale (12; p_{25}: 8; p_{75}: 15)
HAD depression subscale (10; p_{25}: 7; p_{75}: 14)
Headache
Impaired urination
Intestinal dysfunction
Life quality SF36 mental subscale (35; p_{25}: 25; p_{75}: 48)
Life quality SF-36 physical subscale (27; p_{25}: 22; p_{75}: 32)
Major depression
Memory complaints
Migratory joint pain
Months of pain (96; p_{25}: 48; p_{75}: 156)
Morning stiffness
Muscle weakness
Muscular contractures
Onset type
Pain level (VAS 1-10 cm) (7.5; p_{25}: 6.5; p_{75}: 8.5)
Pain subtle movements impairment
Palpitations
Panic attacks
Personal history of chronic pain
Personality disorders
Pittsburgh Sleep Quality Index (PSQI) (14; p_{25}: 11; p_{75}: 17)
Post exercise fatigue
Posttraumatic stress disorder
Previous Personal history psychopathology
Sleep Disturbances
Spine osteoarthritis
Trembling
Trigger Presence
Visual accommodation impairment
Widespread pain

The basic procedure behind partitioning cluster analysis is to construct subgroups with homogeneous objects, which in our study are variables corresponding to clinical features, based on a well-defined proximity measure. Given the non-continuous nature of our variables, we used the Gower's general similarity measure (192). The most common partitioning clustering method is the K-means algorithm (193); however, we used a more robust approach called *position around medoids* (PAM), which is similar to K-means, but group membership depends on proximity to an actual observation (medoid) instead of proximity to an average object (centroid). The number of clusters or subgroups was determined using silhouette plots. The names that were eventually given to each of the subgroups were determined post-hoc in an attempt to characterize the nature of each FM dimension. The clustering was constructed in an initial sample of 559 FM patients, and a cross-validation of the tool was performed in a second sample of 887 patients.

Construction of samples subgroups

The clustering tool to define the underlying FM dimensions was used to further construct patient synthetic indexes based on the clinical features' composition of each FM dimension. The values of the synthetic indexes were calculated from linear functions (one per dimension) that weight the dichotomous variables constituting the specific dimension. The weighting factor is based on the silhouette value of the specific variable in the cluster to which it is assigned, which can be understood as a measure of the contribution of that variable to the dimension. The silhouette values take meaningful values from 1 (well assigned) to -1 (closer to variables assigned to neighbouring clusters), where 0 indicates no clear assignment. The resulting values of the synthetic indexes result into continuous measures for each FM dimension. The synthetic indexed values calculated per each sample and dimension where then used to find patient subgroups. Given now the continuous nature of the measures we used the K-means clustering procedure to group patients with similar FM profiles. All analyses were performed in the R environment using the *cluster* package.

GENOME-WIDE ASSOCIATION STUDY

321 FM cases, selected by FFSGCDB principal investigators for having low levels of psychiatric comorbidities and being the best fitting the FM diagnosis, were genotyped with Illumina 1M-Duo chip, which interrogates nearly 1.2 million loci per sample, containing tag SNPs, SNPs in genes, and SNPs and non-polymorphic markers in known and novel copy number variation (CNV) regions. The median spacing between markers is 1.5 kb (a mean of 2.4 kb), with the 90^{th} percentile for the gap size between SNPs being 6 kb. The array captures 95% of CEU HapMap population variation (HapMap 1) and also includes 60000 CNV targeted markers. Genotyping was performed in CeGen (Barcelona Node), following the manufacturer's protocol. 200 ng of DNA per sample was used. A BeadArray Reader extracted images and read fluorescence intensities, and all data was uploaded into BeadStudio software for quality control and processing.

Data from 425 Spanish controls aged 20-44 years old (general Spanish population from GABRIEL consortium http://www.cng.fr/gabriel/index.html) genotyped with Illumina 610-quad chip was available. Genotyping had been performed in Centre National de Génotypage (http://www.cng.fr/) (Évry, France). Since both cases and controls had been genotyped separately with different platforms, in order to minimize platform bias, quality control (QC) procedures were performed separately and QCed cases and controls were subsequently merged together for association analysis (Figure 22).

Quality control

Quality control was performed with PLINK software (http://pngu.mgh.harvard.edu/~purcell/plink/) (194). These QC steps were executed separately in the FM and control datasets, only taking into account the SNPs that overlapped between the two datasets (582,892 SNPs). The QC was first performed at sample level.

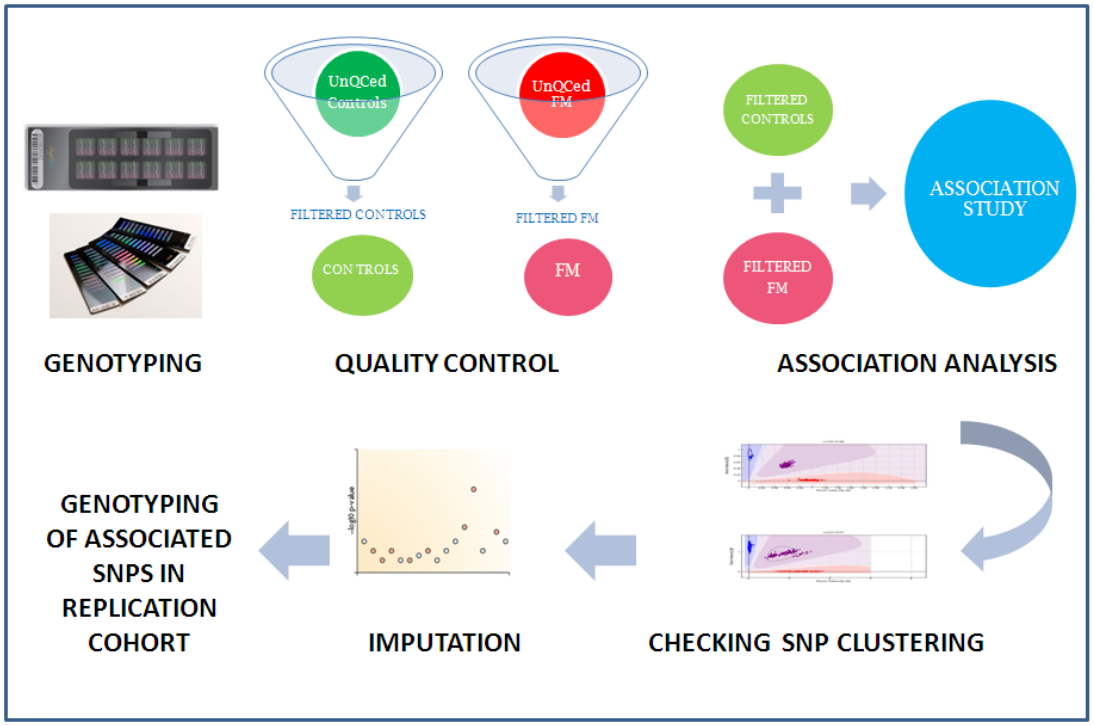


Figure 22: GWAS analysis pipeline: from genotyping to replication of association results

1. Sample QC

1.1. Origin

The first step consisted of identifying and eliminating samples that were not of European origin. For this purpose, we ran the genome function in each of our datasets (FM and controls). First, we extracted SNPs included in the 610-quad array from FM, CEU, Yoruba and Chinese-Japanese HapMap datasets (from HapMap 2 phase). HapMap 610-quad data were then merged with each of our datasets (controls and FM

610-quad data) in two separate files. A crucial step in this procedure was to get all the genotype data in the same strand in controls, FM and HapMap data (which all SNP alleles are in the positive strand). This was achieved with the PLINK flip strand function. Then, the Genome function (which calculates genome-wide identity by descent (IBD) given identity by state (IBS) information to define pairwise similarities between samples) was used to compare samples pairwise (FM samples with HapMap samples and controls with HapMap samples). An inbreeding coefficient (Pi_HAT) was then calculated:

PI_HAT P (IBD=2) + 0.5*P (IBD=1) (proportion IBD)

Then the output of this analysis was run into the identity by state function (cluster function) and samples were clustered depending on how many genotypes and alleles their shared (identity by state: IBS):

(IBS2 + 0.5 IBS1) N

IBS2: number of markers where pair of individuals share 2 alleles

IBS1: number of markers where pair of individuals share 1 allele

N, number of SNPs

The two principal components that best classified the samples into these clusters were then plotted in a multidimensional scaling plot with R. Those outliers (FM samples or controls) not-clustering with CEU samples were excluded (Figure 23).

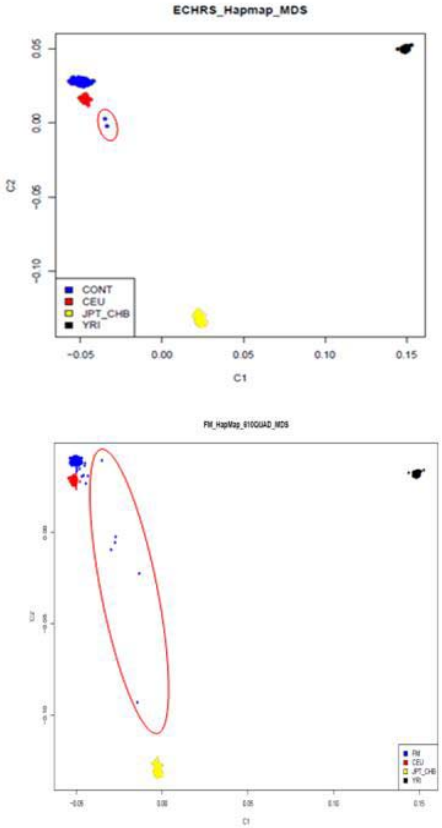


Figure 23: PCA plots for controls (top image) and FM samples. Samples that did not cluster with HapMap CEU samples where excluded from the analysis. Two control samples and six FM were excluded (red circles).

1.2 Call rate

PLINK's call rate function was used to obtain sample call rates for the FM and control datasets. These call rates were plotted in histograms with R software. The cut off value was defined by eliminating outliers (Figure24).

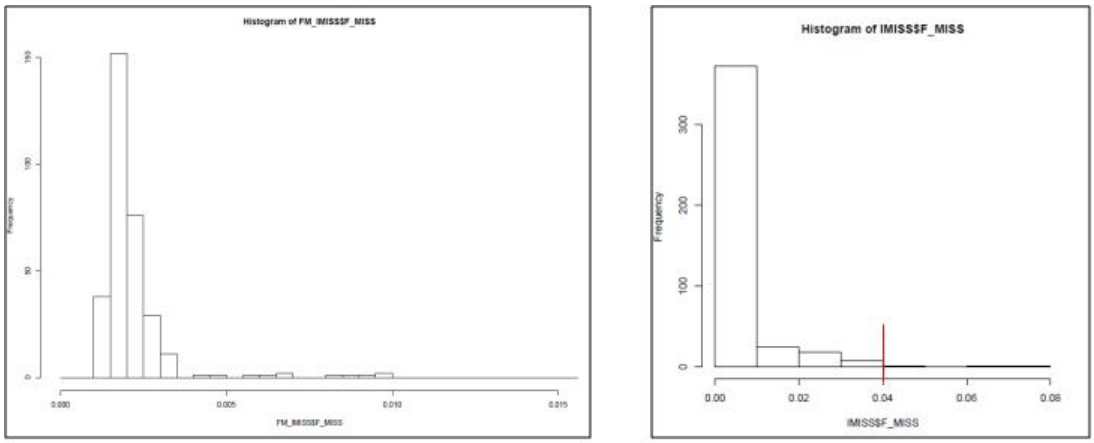


Figure 24: Histograms representing the genotyping rates of FM (left) and control samples (right). No FM samples were excluded; in the control dataset, the threshold (marked with a red line) was established at 96% and 3 individuals were excluded.

1.3 Heterozygosity

A low level of heterozygosity could be indicating inbreeding, and a high level, samples' contamination. An algorithm developed by the group of Dr Eleftheria Zeggini (Sanger institute) was used to calculate heterozigosity levels in FM samples and in controls were calculated. The heterozygosity values of the samples were then plotted in histograms (Figure 25). The cut off value was defined by eliminating outliers. For FM samples, cut off value was established at 30%: three samples were excluded. For controls, the threshold was 29% and two samples were excluded.

1.4 Gender check

Gender was imputed using genotype information from the X chromosome. Those samples discordant between phenotypic /database sex and imputed sex were excluded. Two FM samples and three controls were excluded for this reason.

1.5 Relatedness

Samples that presented a degree of relatedness higher than expected in a random sample were excluded. For this purpose we used the PI_HAT inbreeding coefficient values: samples presenting a PI_ HAT value higher or equal to 0.05 were excluded. Five FM samples and nine controls were excluded.

After all the sample-based filtering procedures, 13 samples were removed from FM dataset and 30 in the control dataset. The resulting filtered datasets included 582892 SNPs and 308 individuals in the FM dataset and 582892 SNPs and 395 individuals in the control dataset. These were then filtered at the SNP level.

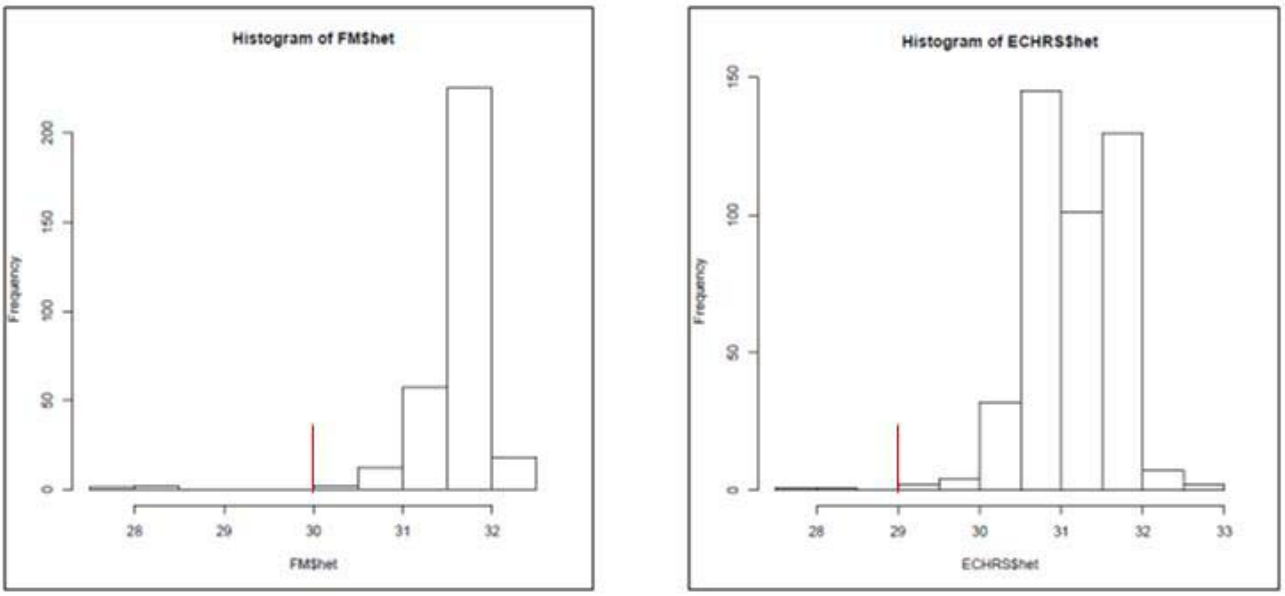


Figure 25: Histograms representing percentage of heterozygosity of FM (left) and control (right) samples. A red line indicates cut-off values.

2. SNPs QC

2.1 Genotyping rate

The Genotyping rates in both datasets were obtained with PLINK and plotted in histograms with R (Figure 26). Both for FM samples and for controls the cut-off value was established at 96%: 4846 SNPs were excluded in the FM dataset and 5716 SNPs in controls.

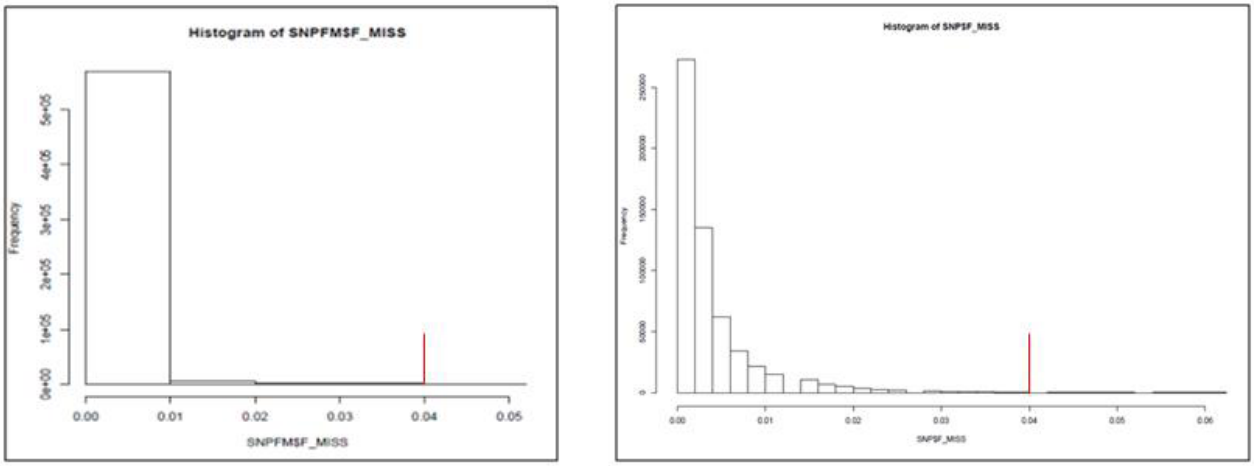


Figure 26: Histograms representing SNPs call rate in FM (left) and controls (right) datasets. A red line indicated cut-off value.

2.2 Hardy Weinberg Equilibrium

The threshold for Hardy Weinberg Equilibrium (HWE) in GWAS is not clearly established. Many investigators use a threshold of $10.e^{-8}$ claiming that they use the same multiple testing correction of association significance. Since we used a reduced sample number, we decided to be more stringent, and used a $10.e^{-4}$ threshold. This lead to the exclusion of 653 SNPs in the FM samples, and 883 SNPs in the control dataset.

2.3 Minimum Allele Frequency

SNPs with a minimum allele frequency (MAF) <5% were excluded: 64,610 in FM and 65,203 in controls. After QC at the sample and SNP levels, the FM dataset included 308 cases (8 males and 300 females) and 513,897 SNPs and the control dataset 395 controls (192 males and 203 females) and 512,615 SNPs.

Merging filtered FM and filtered controls

As FM cases and controls had been genotyped with different platforms, when merging both datasets we had to face the problem of flipped strands. In the first step of QC, when checking for the samples European origin, both FM and controls datasets were flipped in order to have all the SNPs in the positive strand as it is in the HapMap datasets. However, although PLINK detects flipped strands, it is not able to do so in the case of A/T and C/G SNPs. To solve this problem, we used the PLINK reference allele function (manually specifies major allele and minor alleles) on the overlapping SNPs between both filtered datasets (505,454 SNPs).

Then FM and controls datasets were merged. We ran the genome and cluster functions and, again, the two principal components that best classified the samples into these clusters were plotted in a multidimensional scaling plot with R. This was performed only with cases and controls, then adding CEU HapMap data and finally with all three HapMap populations (Figure 27).

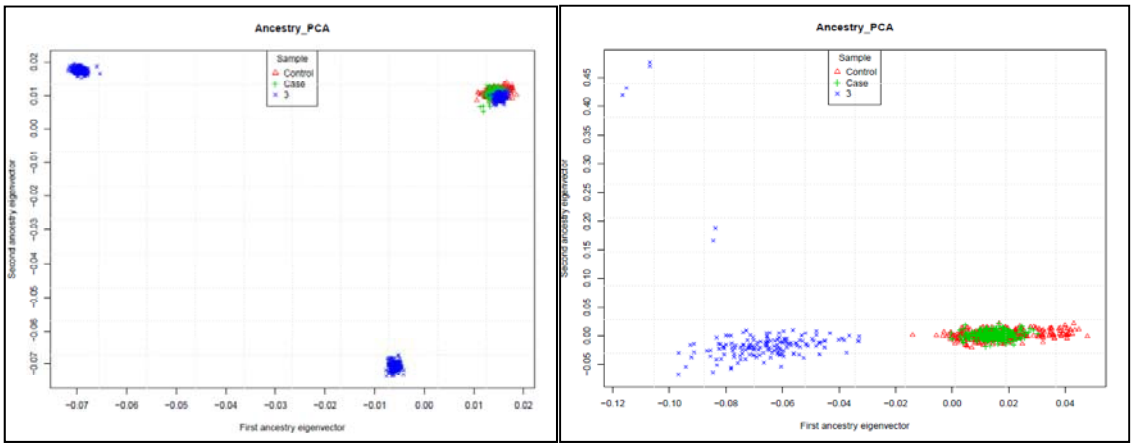


Figure 27: PCA plots of FM and control datasets with HapMap CEU, CHB_JPT and YRI populations (left), and with HapMap CEU (right). Both FM and control datasets cluster with CEU HapMap individuals.

The gender issue

Since 97% of the FM samples are female (and over 90% of FM cases in clinical practice) we decided to use only female controls to have a gender matched control set and to be able to evaluate the X chromosome (Figure 28). So in the end, 505,454 SNPs in 300 FM cases and 203 controls were considered for the association analysis.

Figure 28: PCA plot of the merged FM and controls only considering females. There was no evidence of population stratification: the genomic inflation in the females study was finally 1.013.

Association analysis

Allelic association analysis was performed with PLINK (95% CI). QQ plots and annotation of top SNPs were performed with the WGA software (http://compute1.lsrc.duke.edu/softwares/WGAViewer/mainmenu.php) (195) and Manhattan plot with Haploview software (196).

Checking SNPs clusters

SNP clustering for SNPS showing strongest association was only checked in FM cases, since we had no access to controls raw data. In normal conditions, three clusters are expected: one for each possible SNP genotype (homozygous for A allele, heterozygous and homozygous for B allele). The purpose of this step is to discard spurious associations caused by a secondary SNP or a CNV overlapping the SNP, which would lead to a bad clustering with more than three clusters.

Imputation

Imputation was performed for fine mapping of the GWAS results following the workflow summarized in Figure 29. In brief, PLINK files were divided into the different chromosomes. Then file formats had to be modified in order to run the imputation software. This was done with GTOOL (http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html). Then imputation was performed with Impute (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html) considering an interval buffer region of 500 kb and using as reference panels 1000Genomes and Hapmap3 CEU data. Association analysis of imputed data was performed with SNPTEST (http://mathgen.stats.ox.ac.uk/genetics) using a frequentist test with method_score option that takes into account genotype uncertainty. Infoscore measures the genotype uncertainty: 0.8 value would correspond to an uncertainty of 20%. Usually, infoscore> 0.8 (and for some groups even 0.5) is used as filtering criteria to select SNPs with a good imputation quality.

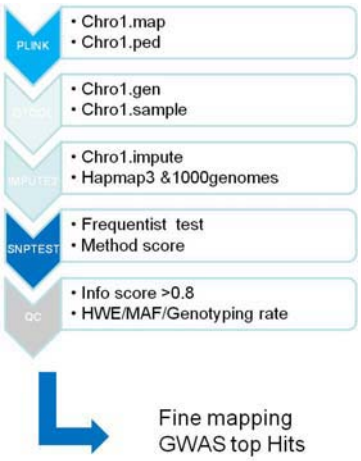


Figure 29: Workflow of the imputation analysis. At each step of the procedure, the programs used and finally the QC step are listed on the left side, while the format of the files, the reference panels used and the filters applied, are mentioned on the right side.

SNPs annotation

SNPs showing strongest association were annotated with WGA viewer software. Annotation included identifying the nearest gene and genetic position of the SNP (coding, intronic, intergenic). The SNP and its genomic regions relation to disease was evaluated with Decipher (http://decipher.sanger.ac.uk/) database.

Replication

In order to validate the GWAS results, 21 SNPs showing strongest association values were genotyped in a replication set from our cohort, consisting in 982 additional FM cases and 971 controls coming from the Banco Nacional de DNA de Salamanca. Controls were selected for having low levels of pain and fatigue as assessed by a questionnaire. Genotyping was performed by Kbiosciences, using their PCR SNP genotyping

system (KASPar®: Kbioscience, UK), which uses a competitive allele-specific PCR. As quality control, duplicates (2 HapMap samples in each plate) and negative controls where included. SNPs not fulfilling HWE, with a MAF<5% or a low genotyping rate (<95%), as well as samples with low genotyping rate (<95%), were excluded from the analysis. Association analysis was performed with Hmisc, Nnet and SNPassoc R packages (genotypic and multinomial analyses). Multinomial analysis was performed in order to evaluate the possible effect of FM clusters in SNP association.

The prioritization of the 21 SNPs to be genotyped in the replication cohort was based on several criteria:

- Pvalue
- Signals: clustering of SNPs with strong associations in a genomic region
- Function: SNPs located in genes or gene regions (as assessed by UCSC genome browser and WGA viewer software)

Imputation data was considered for fine mapping. For GWAS regions showing positive signals, understood as several SNPs showing a strong association in a 50 kb region, imputation results were evaluated. SNPs in a window span of 100kb were considered, and those showing a stronger association than the directly genotyped SNP where also included for replication.

Assessing SNP function in silico

In order to evaluate the function of associated SNPs we used different tools. First of all, we used Puppasuite (http://pupasuite.bioinfo.cipf.es/), an interactive web-based SNP analysis tool that allows for the selection of relevant SNPs within a gene, based on different characteristics of the SNP itself, such as validation status, type, frequency/population data and putative functional properties (pathological SNPs, SNPs disrupting potential transcription factor binding sites, SNPs located in intron/exon boundaries).

We also studied the SNPs' possible cis effects in gene expression levels with Genevar (http://www.sanger.ac.uk/resources/software/genevar/). This platform of database and web services is designed for integrative analysis and visualization of SNP-gene associations in eQTL (expression quantitative trait loci) studies. It evaluates the possible correlation between SNP genotypes and gene expression levels in different datasets: lymphoblastoid cell lines from 726 Hapmap3 individuals (197), skin, adipose tissue and lymphoblastoid cell lines from female twins (198), and in primary fibroblasts, lymphoblastoid cell lines and T-cells from umbilical cords of individuals of Western European origin (199). Briefly, data from expression arrays (Illumina human HT-12 and Illumina WG-6 v3 expression arrays) and from SNPs arrays (Illumina 1M duo and Illumina 550 K) are used to test for association between the SNP's genotype and mRNA levels. This is performed in a two-step process. First, SNP-gene associations surrounding eSNPs/lead SNPs among tissues/populations are investigated with the SNP centric analysis (cis-eQTL-SNP). Then SNP-probe

associations across cell types/populations are plotted (SNP-gene association analysis (eQTL - SNP-Gene)). We took into account all probes within a 1 MB window centred in the transcription start site of the gene and performed association analysis using Spearman Rank Correlation and setting the pvalue limit filter at 0.001.

Finally, after this year's publication of Encode results, we used the Regulome database (http://www.regulomedb.org/). RegulomeDB (200) is a database that annotates SNPs with known and predicted regulatory elements in the intergenic regions of the human genome. Known and predicted regulatory DNA elements include regions of DNAase hypersensitivity, transcription factors binding sites, and promoter regions that have been biochemically characterized to regulate transcription in different cell types (lymphocytes, glioblastoma cell lines, osteoblasts and stem cells, among others).

Pathway analysis of GWAS and replication results

SNPs showing strongest association were analyzed with Ingenuity Systems Pathway analysis (IPA) software (http://www.ingenuity.com/) and GeneSet analysis Toolkit v2 (http://bioinfo.vanderbilt.edu/webgestalt/). Briefly, for SNPs showing strongest allelic association in GWAS, the nearest gene Entrez Gene IDs were entered into the IPA database for a core analysis and into GeneSet analysis Toolkit for gene ontology (GO) biological processes, molecular function and cellular components analysis, and Kegg pathways analysis against human genome, applying Benjamini and Hochberg correction. Kegg pathway mapping is the process to map molecular datasets (in the present study genomics) to the Kegg pathway maps which are manually drawn pathways representing our knowledge on the molecular interaction and reaction networks for metabolism, cellular process.

CNV assessment: PennCNV

Illumina arrays, in addition to providing genotype calls, allow the inference of the proportion of hybridized sample that carries the B allele, which is called B allele frequency (BAF) (http://www.illumina.com/Documents/products/technotes/technote_cnv_algorithms.pdf). In a normal sample, discrete BAFs of 0, 0.5 and 1.0 are expected for each locus, representing each of the possible genotypes (AA,AB and BB). Deviations from this expectation are indicative of aberrant copy number. Furthermore, the logged ratio of observed probe intensity to expected intensity (LRR) is also measured. Any deviations from zero in this metric are suggestive of a region varying in copy number as well. Several algorithms handle these signal intensity data coming from SNP arrays to infer CNV calls. Of these, we decided to use PennCNV software, as it is one of the most accepted tools.

PennCNV implements a hidden Markov model (HMM) that integrates multiple sources of information to infer CNV calls for individual genotyped samples (161). It differs form segmentation-based algorithms in that

it considers SNP allelic ratio distribution and other factors, in addition to signal intensity alone. In addition, PennCNV can optionally utilize family information to generate family-based CNV calls by several different algorithms. Furthermore, PennCNV can generate CNV calls given a specific set of candidate CNV regions, through a validation-calling algorithm. It allows six-states definition of CNV events (Table 8).

Table 8: CNV States detected by PennCNV (taken from Wang et al).

Copy no. state	Total copy no.	Description (for autosome)	CNV genotypes
1	0	Deletion of two copies	Null
2	1	Deletion of one copy	A, B
3	2	Normal state	AA, AB, BB
4	2	Copy-neutral with LOH	AA, BB
5	3	Single copy duplication	AAA, AAB, ABB, BBB
6	4	Double copy duplication	AAAA, AAAB, AABB, ABBB, BBBB

In our PennCNV based study, we selected regions including at least three markers showing alterations in intensity data. Out of these, we only selected those aberrations present in at least 5% of the samples. We further selected those regions located in a gene (+/-50 kb). Then, CNV events having been previously reported as very frequent polymorphic events present in the general population and/or related to population differences were also discarded. Finally, out of the CNV regions that had passed all the selection filters, the validation of PennCNV results started by those regions that had been also detected by aCGH in our samples. Furthermore, we performed a complementary filtering procedure in order to detect rare and large events, which avoided plate or batch effects and were not present in the general population. Based on Zhang et al's (201) work, we followed the pipeline summarized in Figure 30.

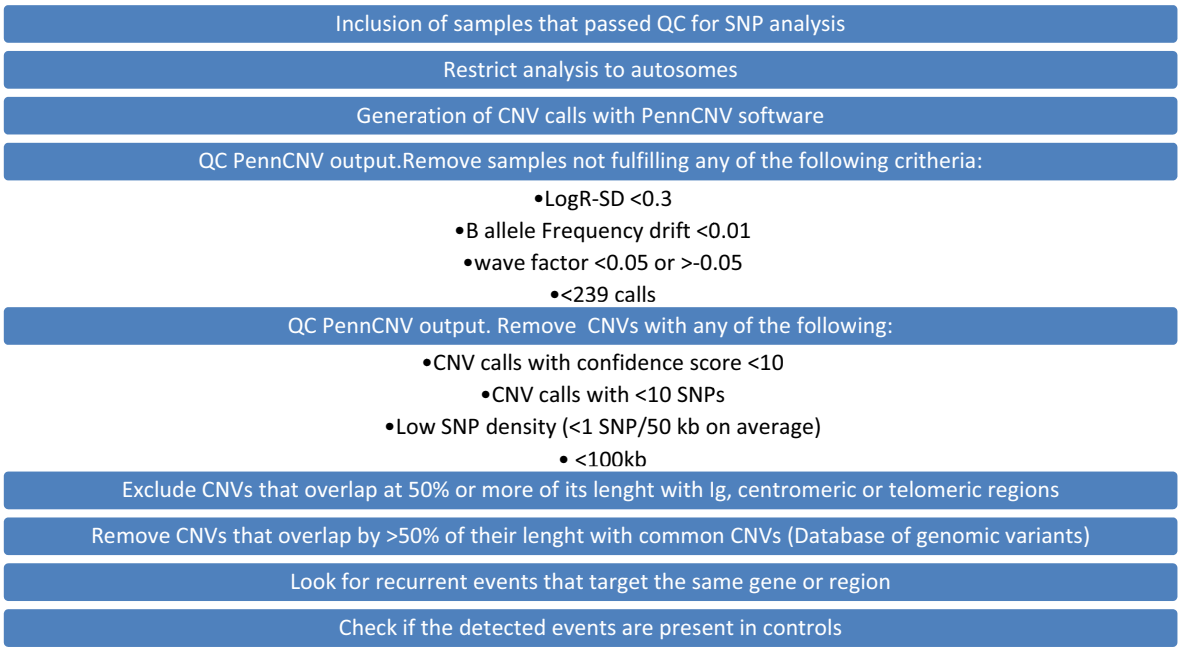


Figure 30: Detection of rare CNVs algorithm from PennCNV output.

CNV mosaicism assessment

We used the Mosaic Alteration Detection-MAD algorithm developed by Gonzalez et al (202) for the detection of CNVs in mosaic state by using BAF and LogR ratio. Briefly, when there is a CNV (insertion or deletion) there are changes in LogRatio and BAF that overlap .An altered BAF that doesn't overlap with significative changes in LogR (<0.2) is indicative of a CNV event in mosaicism. An abonormal average BAF for heterozygous SNPs (not centered at 0.5) with a normal average logR value around 0 is indicating a probable neutral copy number change with allelic imbalance suggestive of a uniparental disomy (UPD) in mosaicism; an abnormal average BAF accompanied by an altered logR ratio not reaching the chosen threshold for heterozygous deletions or duplications could be indicating a gain or loss in mosaic state. BAF values and BAF standard deviation (Bdev) can be used to infere the percentatge of cells carrying the mosaic event:

L (proportion of cells with a loss) = 2Bdev/ (0.5 - Bdev)

G (proportion of cells with a gain) = 2Bdev/ (0.5+ Bdev)

U (proportion of cells with copynumber neutral change- UPD) = 2Bdev

This algorithm allows the detection of four possible states: loss, gain, trisomy and lost of heterozygosity (LOH) due to identity by state.

ARRAY COMPARATIVE GENOMIC HYBRIDIZATION

aCGH experiments were performed with two sets of complementary platforms, Agilent® 400K and Agilent®1M using a pooling strategy. This strategy aims to dilute rare CNV polymorphisms due to inter-individual variability and highlight those common variants with a different frequency between cases and controls. Three pools of FM samples where designed: FM without fatigue (20 samples), FM with fatigue (20 samples) and early onset FM (30 samples). All the samples included in the pools had a family history of FM. FM pools were hibridized against one pool of controls (50 samples). Samples technical requirements included DNA integrity (large DNA fragments, 280-260 ratio around 1.8 and 260-230 ratio around 2) and a precise quantification ensuring that the same amount of each of the samples was included in the pool (concentration range 50-150ng/µl).

Experimental protocol

In order to assess DNA quality of the samples for inclusion in the pool, samples were run in a denaturing gel, DNA concentration was quantified by Picogreen and Nanodrop, and the 280-260 and 260-230 ratios were calculated. Samples not fulfilling quality control criteria were excluded from the pool. Equal amounts of each DNA were mixed into the pooled DNA used for the hybridization. A reference pool, generated in the same way, was created.

Each FM pool was hybridized against the control pool on the Agilent 400k array (direct and dye swap) and the Agilent 1M array (only direct as they were complementing 400K results) (Figure 31). 1µg of the reference and test pools was differentially labelled and both reference and one-test pools were competitively hybridized to a microarray. Then, in order to avoid bias due to labelling, the test and control pools were labelled with the opposite dye and hybridized onto a new array (dye swap). Hybridizations were performed according to Agilent Oligonucleotide Array-based CGH for genomic DNA Analysis (direct method) protocol (Version 4.0, June 2006). Data were filtered to exclude bad spots and to adjust too low or too high intensities to more reasonable values, and then normalized to correct systematic errors due to technical reasons instead of biological variability, so that the modal ratio for the genome was set to a standard value 0.0 on a logarithmic scale.

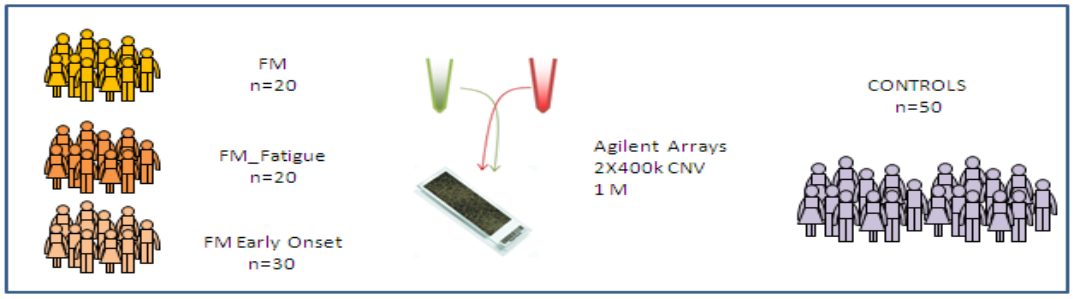


Figure 31: Array CGH using a pooling strategy. Three pools of FM samples with family history of FM were designed and each was dyed with a fluorophore and hybridized against a pool of pain-free controls on the 400k and 1M Agilent arrays. For the 400 K array the experiment was repeated exchanging the fluorophore between cases and controls in order to avoid fluorophore bias.

Analysis of arrays results

The resulting data was analyzed with Agilent's Genomic Workbench software, using the ADM-2 algorithm. Since the standard deviation of the arrays was lower than or equal to 0.15 (Table 9), the cut-off value for selecting probes showing differential hybridization was established at 0.3 (2 standard deviations). We selected regions with at least 3 aberrant probes and a mean/smoothed log2ratio>0.3. ADM-2 algorithm identifies all intervals with consistently aberrant low or high log2 ratios based on a statistical score. It automatically calculates the optimal size of an aberrant region and tests regions in which the statistical score pass a user-defined threshold. This statistical score is based on log2 ratios and number of probes. There is not a fixed window: it samples adjacent probes to arrive to a robust estimation of the boundaries of the aberrant region. Furthermore, it incorporates quality data of each log ratio (probe log ratio error). For this reason, it is more robust to noise in the data (presence of noisy probes), especially if the user is interested in detecting small aberrant regions. Finally, we selected only those regions fulfilling these criteria in both direct and dye-swap hybridizations.

Table 9: Standard deviation of 400k array hybridizations.

Hybridization	σ
FM-control	0.156
FM-control DS	0.151
FM_FC-control	0.146
FM_FC-control DS	0.142
FM_early-control	0.142
FM_early-control DS	0.135

Validation of array results

Most CNVs detected in the aCGH experiments were shown to be insertions-deletions (INDEL). In these cases, we first attempted to detect the breakpoints of the CNV, and then designed a multiplex PCR to genotype it in our entire cohort of cases and controls.

1.BREAKPOINTS DETECTION

We had to identify the breakpoints of the following CNVs: *ACACA* (detected both in 400K array and PennCNV) and *WWOX* (400K array). The pipeline summarized in figure 32 was followed-up. If the initial PCR amplified several products, we performed gel extraction of the band corresponding to the deleted allele. If gel extraction yield was insufficient for a subsequent sequencing PCR, we performed PCR subcloning. This also was performed when the presence of repetitive sequences flanking the products didn't allow the obtaining of a clean sequence. PCR conditions and the different experiments are described right afterwards.

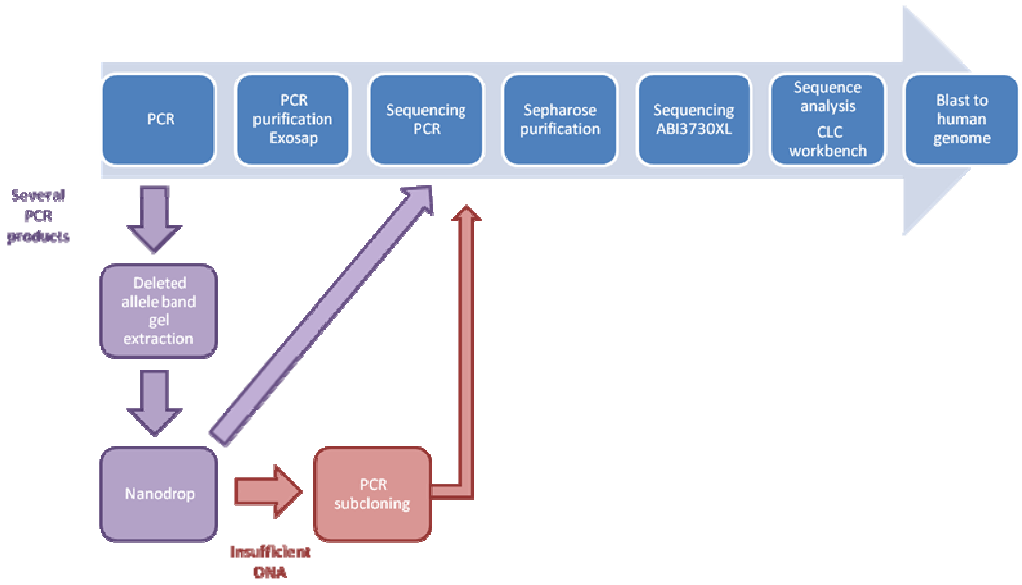


Figure 32: Pipeline for breakpoints detection. If the standard protocol (blue arrow) didn't allow breakpoints detection additional experiments (violet and red boxes) were performed.

1.1 PCR conditions

For breakpoints detection we used the same mix components for the two CNVs evaluated. PCR program differed in the annealing temperature and extension temperature and length, according to the expected size of the PCR products and the primers melting temperature (Table 10). Primers were designed with Primer3 software taking into account aCGH positive probes coordinates and overlapping CNVs described in the Database of Genomic Variants. Primers sequences are summarized in Table 11.

We took advantage of data available in Conrad et al work and included, as positive controls, HapMap samples with known genotypes for the CNVs (Table 12).

Table 10: PCR conditions for ACACA and WWOX CNVs breakpoints detection.

	ACACA	***WWOX***
PCR program	2' 94ºC	2' 94ºC
(30 cicles)	**30" 94ºC**	**30" 94ºC**
	30" 60ºC	**30" 60ºC**
	1' 72ºC	**45" 68ºC**
	7' 72ºC	7' 68ºC
Mix components	50 ng DNA	
	10x Roche® PCR reaction buffer + Mg^{+2}	
	0.2 mM dNTPs	
	0.4 pM/µl each primer	
	0.1 U/µl Taq Polymerase	
	H_2O to reach a final volume of 25 µl	

Table 11: Primers used for breakpoints detection

CNV	**Primers (5'-3')**	**Product size (bp)**
ACACA	**ACACA_DelF**: GGCCTCCTCTTCTAGCTGTTG **ACACA_DelR**: AACAGGTGCCCAATAAATGC	**≈ 1200**
WWOX	**WWOX_F1**: TGGGTAGGAATCCTGCAGAC **WWOX_R1**: TGCCTAAAAGCACACACTGC **WWOX_R2**: GGGCATCCCAGTTTTCTACC **WWOX_R3**: CCTGCTTCCTGAACATTCCT	**Depending on primers combinations**

Table 12: HapMap samples with different genotypes for WWOX and ACACA CNVs Conrad *et al.*.

CNV	Two copies	One copy	Zero copies
ACACA	NA10854; NA10855; NA10860	NA10847; NA10852	NA07055; NA07029; NA107048
WWOX	NA07019; NA06994	NA12056; NA12145; NA12864	NA06991; NA12761

1.2 Removal of PCR primers and reagents

5 µl PCR products were cleaned up with 2 µl of USB® Exo-SAP-IT® following an incubation of 15' at 37 º plus 15' at 80º.

1.3 Sequencing PCR

1 µl of PCR product after Exosap, was added to a mix of 1 µl Big Dye Terminator® v3.1 (Applied Biosystems), 1.5 µl 5X Buffer, 0.5 µl of either reverse or forward primer (10 µM) and 6 µl H_2O, and a sequencing PCR reaction was performed following the PCR program below (30 cicles):

30" 95º
30" 50º
3' 60º

For Minipreps products, the sequencing PCR was performed with 400 ng of DNA and one of the vector's primers (Table 13).

Table 13: pGEM easy T7 and SP6 universal primers sequence.

Primers (5'-3')
T7: TAATACGACTCACTATAGGG
SP6: ATTTAGGTGACACTATAG

1.4 Purification sequencing PCR

Sequencing PCR products were purified with sepharose (Sephadex®-G50) columns. Briefly, 800 µl of sepharose were pipetted into a column and centrifuged at 1000 g for one minute for sepharose compactation. Flow-through was discarded and 10 µl of water were added to the sepharose column and centrifuged for one minute again at 1000g. Finally, the column was introduced in a new eppendorf, the sequencing PCR loaded into the column and centrifuged for 1 minute at 1000g. The purified PCR was run in a capillary sequenced (3730XL Applied Biosystems).

1.5 Sequence analysis and blast to human genome

Sequencing results were analyzed with CLC workbench with standard settings. Only clean sequences were selected for blast analysis in UCSC Genome Browser.

1.6 PCR product gel extraction

PCR products were run on a 1.3% low-melting agarose gel, the bands corresponding to the deleted alleles were cut using a UV transilluminator and extracted with QIAquick® Gel Extraction Kit from Qiagen, using a microcentrifuge, according to product specifications.

1.7 Subcloning PCR products

The amount of PCR product (insert) to be included in the ligation step was calculated according the formula:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert:vector molar ratio} = \text{ng of insert}$$

Where the size of the vector is 3 kb, vector concentration 50 ng/ µl and insert:vector ratio is recommended to be 1

Ligation of the extracted PCR product was performed with New England Biolabs Ligase and pGEM easy vector according to manual instructions, with an overnight incubation in a 16 °C bath. The ligated product was then transformed in JM109 High Efficient Competent Cells that were plated in LB/ampicilin/IPTG/X-GAL plates overnight at 37 °C. Positive colonies were picked the day after and incubated for 1 hour at 37 °C in 100 µl of lysogeny broth (LB) medium (with ampicilin), and 1 µl of the incubation was used to perform the PCR to check for the inserts.

The 100 µl of LB medium (with ampicilin) positive for the insert, were incubated in 5 ml of LB medium (with ampicilin) overnight in a shaker at 37°C at 220 revolutions per minute (RPM). Plasmid DNA was purified the day after with Qiagen Miniprep kit according to the manufacturers protocol and then quantified with Nanodrop.

2. CNVs GENOTYPING

2.1 Multiplex PCR

Each genotyping reaction included both cases and controls in order to avoid possible bias, and negative (H_2O) and, when available, positive controls (HapMap samples with validated genotypes for each CNV (146)). PCR conditions are and primers are summarized in Tables 14 and 15.

Table 14: PCR conditions for CNV genotyping. Grey cells correspond to shared components in between reactions.

	ACACA	***GALNTL6***	***WWOX***	***MYO5B***	***PTPRD***	***NRXN3***
PCR program	2' 94ºC	2' 94ºC	2' 94ºC	2' 94ºC	2' 94ºC	2' 94ºC
(30 cicles)	**30" 94º**	**30" 94º**	**30" 94º**	**30" 94º**	**30" 94º**	**30" 94º**
	30" 60º	**30" 63º**	**30" 61º**	**30" 62º**	**30" 60º**	**30" 60º**
	1' 72º	**30" 72º**	**30" 72º**	**30" 72º**	**30" 72º**	**30" 72º**
	7' 72º	25' 72º	7' 72º	7' 72º	7' 72º	7' 72º
Mix components						
50 ng DNA		75 ng				
10x Roche® PCR reaction buffer+Mg^{+2}						
dNTPs (mM)	0.125	0.06	0.125	0.2	0.2	0.15
Del Primers (pM)	0.40	0.02	0.4	0.4	0.24	0.4
Non-Del primers (pM)	0.24	0.032	0.4	0.4	0.24	0.2
Taq Polymerase (U)	0.1	0.06	0.1	0.06	0.06	0.02
H_2O to volume of 25 µl						15µl

2.2 PCR products detection

ACACA and PTPRD PCR products were loaded in a 2% Agarose gel and visualized with a UV transilluminator (Gel Doc® (Bio Rad)).

WWOX and MYO5B PCR products were loaded in a 3% Agarose gel and visualized with a UV transilluminator (Gel Doc® (Bio Rad)).

GALNTL6 CNV and NRXN3_DEL were genotyped by multiplex PCR with 5' FAM modification, followed by capillary electrophoresis in a 3730XL automatic sequencer and analysis with the Gene Mapper package (Applied Biosystems, Foster City, CA). Analysis was performed with the Gene Mapper package (Applied Biosystems). Samples showing peak intensities below 1000 fluorescent units or ratios of deleted allele to non-deleted allele < 0.2 or > 5 were not considered for analysis. For capillary detection, NRXN3_DEL PCR reactions were diluted at 1:15, and 1 µl of PCR dilution was then added to 9 µl of a formamide/ROX mixture (950 µl + 20 µl per 100 samples), and samples were loaded into 3730XL. GALNTL6 PCR products were not diluted: 1 µl of the PCR was added to the formamide/ROX mixture.

Table 15: Primers for CNV genotyping PCR reactions.

Allele	Primers (5'-3')	Product size (bp)
ACAC Del	**ACACA_DelF**: GGCCTCCTCTTCTAGCTGTTG	**1164**
	ACACA_DelR: AACAGGTGCCCAATAAATGC	
ACACA Non-Del	**ACACA_F**: GAGCCCATTAATCCAGAAAGG	**449**
	ACACA_R: TGACTTAGTGCCCATTCAAGG	
GALNTL6 Del	**FF_Mi_A:** [6FAM]GCAAGTAATGCCCAAGGAAA	**250**
	RF_Mi_del2: AGAGCATAAACCTCACAGGAC	
GALNTL6 Non-Del	**FF_Mi_wt:** [6FAM]TGGTAATGAGCAGAGGAAAGG	**283**
	RF_Mi_wt5: TGAGCACTTACCCTGTCTGC	
WWOX Del	**WWOX_DelF**: ATCTGGCCATGTCCTCATTT	**192**
	WWOX_DelR: TGTGACCTGATAACCGCTGA	
WWOX Non-Del	**WWOX_F**: AATGGGAATCTTTGCCTGTG	**217**
	WWOX_R: ATGGCAACTGACTTGGGAAG	
MYO5B Del	**MYO5B_DelF**: AACAGGCTGTCTCTTCCATGA	**234**
	MYO5B_DelR: CAGGGGTGGTTAGAATGAGG	
MYO5B Non-Del	**MYO5B_F:** GAATGCATTTTGTCCAGCAGT	**201**
	MYO5B_R: CTCATAGAGGCGGTGTTCTTG	
PTPRD Del	**PTPRD_DelF**: GGGTGGTGGAAGGTGGTTAT	**450 Del**
	PTPRD_DelR: GGTCTGGCATTTTGACATGA	**980 Non-Del**
PTPRD Non-Del	**PTPRD _F:** GCCAATTTCAGATCCTCAGC	**219**
	PTPRD _R: TTAGTGGCGTTCACACATGG	
NRXN3 Del	**NRXN3_FDel**: CAGTCTTGACTGCTGGGTGAAC	**466**
	NRXN3_R: [6FAM]GTGACTGCTGATGAGCCACGC	
NRXN3 Non-Del	**NRXN3_FNodel**: GTGAGCACTCGATCCAGCATAA	**350**
	NRXN3_R: [6FAM]GTGACTGCTGATGAGCCACGC	

2.3 Veracode assay

NRXN3_del was also assessed with 3 SNPs included in a Veracode assay. Two of the SNPs were located within the deleted region (rs12894142 and rs12100748), and we designed a third SNP assay (NRXN3del) with each of its extension probes flanking the breakpoints of the CNV. A combination of the results for these SNPs was used to assess the genotype. A sample was considered as homozygous deleted when failing in both SNPs included in the deleted region and amplyfing in the breakpoints SNP; an heterozygous sample for the CNV was defined by presenting genotype for the three SNPs (the two inside the deletion having to be mandatorily homozygous); the homozygous non deleted samples were characterized by the failure of the breakpoints SNP and presenting genotype at the SNPs inside the CNV region (being either homozygous or heterozygous).

2.4 Statistical analyses

Allelic and genotypic association analyses were performed with R SNpassoc package. Association analysis of imputed data was performed with SNPTEST software. Quality control and association analyses of the Veracode assay were performed with PLINK software. Interaction between CNVs and NRXN3_DEL and GWAS results was tested with R software using a likelihood ratio test (LRT), which compared the full interaction logistic model, a model that includes an interaction term between SNPs, to the simpler additive model, a model that includes just the SNPs as additive effects.

3.EVALUATION OF CNV FUNCTIONAL CONSEQUENCES: NRXN3_DEL

3.1 Genomic characterization of NRXN3_del: Veracode assay

We selected 45 SNPs covering the genomic region of *NRXN3* using the Haploview software. These SNPs capture 42% of alleles in NRXN3α (1,460 Kb) and 66% in NRXN3β at $r^2 \geq 0.8$, based on the CEU HapMap genotyped SNPs (version 2, release 21). For the selection, potential functional variants, variants located near splice sites and SNPs that had been previously associated with disease (203) were forced to be included as TagSNPs. A total of 45 SNPs (41 TagSNPs and 4 singleton SNPs) and 882 FM samples and 889 controls were included for genotyping. SNPs were genotyped using *Illumina's VeraCode* technology (Illumina), at the CeGen's (CEGEN) Barcelona node genotyping facility, following the manufacturer's protocol. The assays, developed for the VeraCode beads, were detected by Illumina's BeadXpress Reader System, and data analysis was performed using Illumina's BeadStudio software (Illumina, Inc., San Diego, CA, USA). As a quality control, 5% of the genotyped samples were duplicated, and no-template controls and six HapMap trios (NA10840-NA12286-NA12287, NA12766-NA12775-NA12776, NA12818-NA12829-NA12830, NA12832-NA12842-NA12843, NA12865-NA12874-NA12875,NA12877-NA12889-NA12890) were included.

3.2 Evaluation of mRNA consequences of NRXN3_DEL

Since *NRXN3* is a gene mainly expressed in the CNS, in order to evaluate NRXN3_del consequences at the mRNA level, neuronal cell lines available in the laboratory were tested for NRXN3_del genotype: two neuroblastoma cell lines (SH-SY5Y and SK N SH), a neuronal stem-cell cell line and two glioblastoma cell lines (T98G and U87). Neuroblastoma and neuronal stem cells were heterozygous for NRXN3_del, while U87 was homozygous deleted and T98G homozygous non-deleted. Therefore, U87 and T98G were selected to evaluate NRXN3_del possible functional consequences.

T98G and U87 CULTURE and DNA and RNA EXTRACTION

T98G and U87 glioblastoma cell lines were plated and grown in Dulebecco's modified Eagle Medium (DMEM) supplemented with 10% of heat inactivated (45 min at 56ºC) fetal bovine serum and 1% of penicillin and streptomycin. Medium was replaced twice a week. Cells were passed when reaching confluency (1:5 every 48 hours approximately). After two passes cells where checked for the presence of mycoplasma infection with Venor®Gem Mycoplasma detection kit (Minerva biolabs). Briefly, 2 µl of the medium were collected and used as a template in a mycoplasma specific PCR. Thus, mycoplasma infection was discarded in both cell lines. Then, cells were tripsinized and pellets for RNA and DNA extraction were collected.

DNA extraction was performed with Wizard® genomic DNA purification kit (Promega) according to the manufacturer's protocol. RNA extraction was performed with Qiagen RNAse plus kit according to product's protocol and including a DNAse purification step. Both DNA and RNA were then quantified with Nanodrop.

As both U87 and T98G DNA and RNA pellets were obtained from a different cell aliquot, with a different pass from the ones from which DNA had been extracted and NRXN3_del had been evaluated, NRXN3_del was evaluated again confirming U87 and T98G genotypes for this variant.

NRXN3 TRANSCRIPTS IN HUMAN BRAIN SAMPLES, AND GLIOBLASTOMA AND HAPMAP CELL LINES

Reverse transcriptase (RT) reaction was performed with 1 µg of RNA with superscript III (SST III) first strand from Qiagen, including a negative control (without SST) following manufacturer's instruction. RT started with 5' at 65ºC, followed by 1' on ice, then 60' at 50ºC, followed by 15' at 70ºC and a final incubation of 15' at 70ºC. 1 µl of cDNA was used in a PCR with *ACTIN* primers and another reaction with the forward primer in *NRXN3* exon 18 and the reverse in exon 21 in order to capture the two possible transcripts (with or without exon 20) resulting from alternative splicing in splice site 4 (Figure 33).

ACTIN housekeeping PCR reaction started with a 5' denaturation step at 95ºC, followed by 30 cycles of 30'' at 95ºC, 30" at 60ºC, and 1' at 72ºC, and a final extension step of 7' at 72ºC. Amplification reactions were performed with 1 µl of cDNA template, 10X Roche® PCR reaction Buffer+Mg, 0.15 mM dNTPs,0.4pM of each of the primers (Table 16) , 0.1 U/µl Taq Polymerase, and H_2O to reach a total volume of 25 µl.

Table 16: Primers for expression housekeeping gene (*ACTIN*) for checking the performance of the RT reaction and *NRXN3* primers for transcripts evaluation (taken from Occhi *et al.* work (204))

Primers (5'-3')	**Product size (bp)**
ACTIN_F**: CTGGAACGGTGAAGGTGACA** ACTIN_R**: GGGAGAGGACTGGGCCATT**	**195**
NRXN3_F18**: TCTTTGGGAAAAGTGGTGGG** NRXN3_R21**: ACCAAGATGCCATCCTTCAC**	**540** **640**

.

NRXN3 mRNA PCR reaction started with a 3' denaturation step at 94ºC, followed by 37 cycles of 30'' at 94ºC, 1' at 65ºC, and 1' at 72ºC, and a final extension step of 7' at 72ºC. Amplification reactions were performed with 1 µl of cDNA template, 10X Roche® PCR reaction Buffer+Mg, 0.16 mM dNTPs, 0.4 pM of each of the primers (Table 17), 0.2 U/µl Taq Polymerase, and H_2O to reach a total volume of 25 µl. 5 µl of the PCR product was loaded on a 2% Agarose Gel in order to check RT reaction (ACTIN) and NRXN3 transcripts.

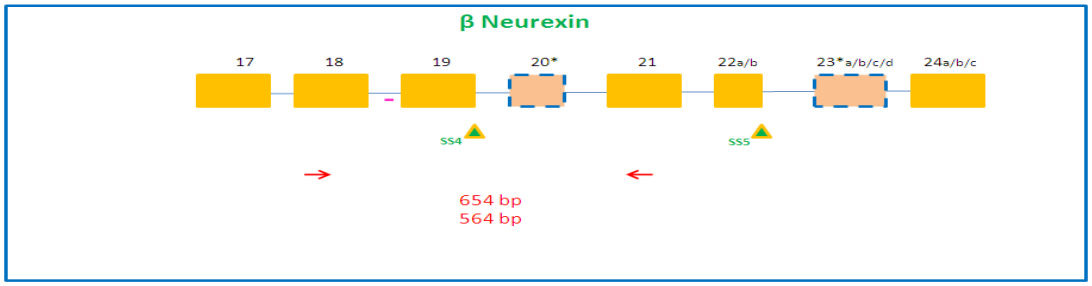


Figure33: βNeurexin 3 scheme including exons and the two canonical sites of alternative splicing. As NRXN3_del (pink line) is located in intron 18-19, we wanted to assess whether it may be influencing alternative splicing at splicing site 4, which may generate two different isoforms (with or without exon 20). SS4: splicing site 4; SS5: splicing site 5. Red arrows represent PCR primers. Exons with blue dashed lines may be discarded through alternative splicing. Note that both the numbering of exons and splicing sites don't start at number one because they also belong to the larger αNeurexin3 isoform and are therefore numbered consecutively after αNeurexin3 exons1-16.

CLONING AND SEQUENCING OF NRXN3 TRANSCRIPTS

In order to check that the two transcripts obtained corresponded to the isoforms with and without exon 20, we sequenced both products from NRXN3 mRNA PCR reaction. For this we gel-extractedthe correct band, purified it and subcloned it, and then performed sequencing PCR reactions.

RELATIVE QUANTIFICATION OF RTqPCR NRXN3 TRANSCRIPTS in U87 and T98G CELL LINES

Quantification of *NRXN3* transcripts in the two glioblastoma cell lines was performed with Taqman specific gene expression assays, purchased from applied biosystems webpage, following manufacturer's instructions. In addition to a specific assay for each of the NRXN3 transcripts resulting from SS4# (Figure 34), a NRXN3 assay measuring a transcript not affected by alternative splicing was purchased in order to evaluate the gene expression in the cell lines. First, a reverse transcription (RT) reaction was performed with SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR (Applied Biosystems) according to the manufacturer instructions. RT reaction started by 10' at 25ºC, followed by 30' at 55ºC and ice chilling. Finally, RNAseH was added and the reaction was incubated 20' at 37ºC. Serial dilutions of the resulting cDNA were then performed to check the efficiency of the different assays in the two glioblastoma cell lines (Figure 35). Once the efficiency was confirmed as being 2, reactions including two housekeeping genes (HPRT1 and TBP) and both *NRXN3* assays in both cell lines were included in a single 384-well plate experiment, in order to minimize bias due to different reactions and plates. This was replicated in a second experiment with a new RT reaction. qPCR reaction consisted of 40 cycles of 15'' at 95º and 1' at 60ºC. The 10 µl reaction consisted of 5 µl of TaqMan universal PCR master mix no amperase UNG, 3.5 µl of H_2O, 0.5 µl of Taqman Gene expression assay and 1 µl of RT product. Two independent RT reactions were performed for each sample, and then, for each RT, samples were run in quadruplicate. Reactions were run in an Applied Biosystems 7900HT Fast RealTime thermocycler. Results were analysed with SDS V 2.1 and RQ manager (Applied Biosystems).

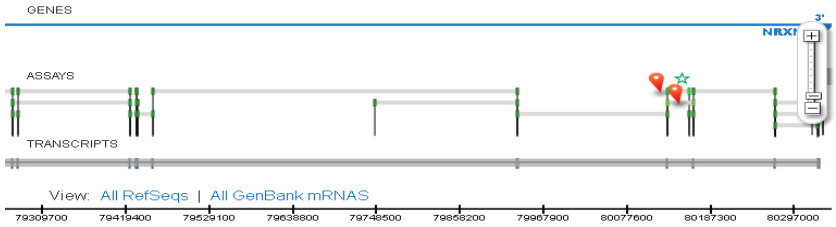


Figure 34: Genomic location of NRXN3 Taqman gene expression assays (NCBI B37; image taken from Taqman assays webpage Applied Biosystems).

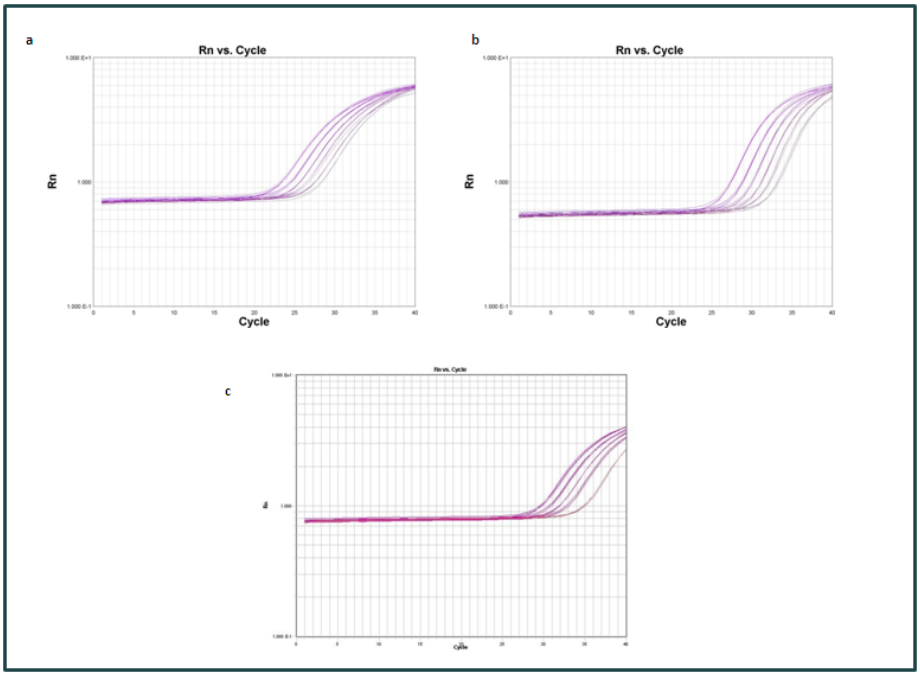


Figure 35: Amplification curves in U87 cell line. a) HPRT1 gene expression assay; (b)TBP gene expression assay;c) NRXN3 gene expression assay. Images taken from RQ manager software (Applied Biosystems).

Relative quantification was calculated with the $2^{-\Delta\Delta Ct}$ method to compare the expression values of both NRXN3 transcripts normalized to two housekeeping genes HPRT1 and TBP, using the followint formula(205):

$$\textbf{RATIO} = (\text{Etest})^{\Delta Ct\ (\text{Control-Sample})} / (\text{Ereference})^{\Delta Ct\ (\text{Control-Sample})}$$

Then, we quantified NRXN3 transcripts in both glioblastoma cell lines and compared the ratio of each NRXN3 transcript in each cell line.

IN SILICO EVALUATION OF PTBP BINDING SITES IN NRXN3_DEL

The candidate intronic regulatory sequences were gathered from Ladd et al (206) and were aligned against the NRXN3_del genomic sequence obtained from UCSC genome browser (http://genome.ucsc.edu/) using clustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Conservation of regulatory sequences present in NRXN3_del sequence was checked with UCSC genome browser conservation track.

EVALUATION OF PTBP2 EXPRESSION IN T98G and U87 CELL LINES AND MUSCLE

We designed primers to check PTPBP2 expression in T98G and U87 cell lines (Table 17). Forward primer was designed in exon 2 and reverse in exon 4 with a genomic span of 7921 bp. PTBP2 mRNA PCR reaction started with a 3' denaturation step at 94ºC , followed by 30 cycles of 30'' at 94ºC, 30'' at 60ºC ,and 30'' at 72ºC, and a final extension step of 7' at 72ºC. Amplification reactions were performed with 1 µl of cDNA template

(T98G, U87, SHSY and muscle), 10X Roche® PCR reaction Buffer+Mg, 0.125 mM dNTPs,0,4pM of each of the primers , 0.1 U/μl Taq Polymerase, and H_2O to reach a total volume of 25 μl.

Table 17: PTBP primers sequence for checking the expression of the gene in U87 and T98G cell lines.

Primers (5'-3')	**Product size (bp)**
PTBP2_F: GATGGTGCTCCTTCTCGTGTA **PTBP2_R**: TGCACTCTCGCTAACTGTGG	297 bp

5 μl of the PCR product were loaded on a 2% Agarose Gel in order to check the presence of PTBP2 PCR product. PCR product was then purified and sequenced.

RESULTS

CLUSTER ANALYSIS

In order to reduce FM heterogeneity we classified clinical data (48 variables) into simplified dimensions that defined FM subgroups. 1,446 FM samples of Caucasian origin were included in the study. Of these, 97.2% were women with a mean age of 49±10 years; more than two thirds (70.5%) were married; more than a third (37.5%) had a high school degree and an additional 19,3% had gone to university; and 80% of the patients had a paid employment.

An initial set of 559 unrelated FM cases was considered for the analysis; the same cluster analysis was cross-validated in a second set of 887 cases. In the first dataset of 559 unrelated FM cases, the selected 48 clinical variables clustered into three independent dimensions: FM symptoms and their characteristics (Dimension 1: "symptomatology"), familial and personal comorbidities (Dimension 2: "comorbidities") and FM core clinical scales (Dimension 3) (Figure 36, Table 18). The composition of the resulting dimensions was homogeneous: only four variables (trembling, personal history of chronic pain, and SF-36 mental and physical subscales) clustered in apparently unrelated dimensions. Since their weights within the respective dimension were among the lowest, their effect on the subsequent FM classification was reduced. Principal component analysis indicated the presence of two correlated factors that explained 31.59% of total variability.

This clustering of the variables was replicated in the second cohort of 887 patients. Given this confirmation, a global analysis was performed using the whole cohort to obtain a global weight for each variable in order to perform patient classification.

Only two dimensions ("symptomatology" and "comorbidities") were considered for the construction of FM subgroups. For each FM sample, a composite index was calculated for these two dimensions. The resulting indexes were used to classify 1,398 out of the 1,446 FM samples, due to missing data.

On one hand, these dimensions were considered as more reliable as they included more variables with higher weights. On the other hand, core clinical scales were not included in the FM subgrouping because they were subsequently used for the assessment of the resulting subgroups and disease severity. Using the scales in this way allowed us to maximize the information from the scales instead of using them as dichotomized variables.

Table 18: Summary of the three different dimensions that emerged after cluster analysis.

VARIABLE	VALUES	DIMENSION	WEIGHT	NEIGHBOUR
DIMENSION 1: Symptomatology				
Widespread pain	0=No 1=yes	1	0.417062	3
Muscle weakness	0=No 1=yes	1	0.403494	3
Post exercise fatigue	0=No 1=yes	1	0.401630	3
Morning stiffness	0=No 1=yes	1	0.386802	3
Muscular contractures	0=No 1=yes	1	0.370318	3
Concentration problems	0=No 1=yes	1	0.363702	3
Memory complaints	0=No 1=yes	1	0.343472	3
Onset	0=Progressive	1	0.327471	3
Sleep Disturbances	0=No 1=yes	1	0.295485	3
Forgetfulness	0=No 1=yes	1	0.232666	3
Migratory joint pain	0=No 1=yes	1	0.232439	3
Headache	0=No 1=yes	1	0.196143	3
Pain subtle movements impairment	0=No 1=yes	1	0.192320	3
Intestinal dysfunction	0=No 1=yes	1	0.188124	3
Visual accommodation impairment	0=No 1=yes	1	0.134756	3
Trigger	0=No 1=yes	1	0.128889	3
Dizziness	0=No 1=yes	1	0.106305	3
Excessive Perspiration	0=No 1=yes	1	0.103443	3
Months of pain (96; p_{25}: 48; p_{75}: 156)	0≤96 1≥96	1	0.097555	3
Personal history of chronic pain	0=No 1=yes	1	0.081029	3
Palpitations	0=No 1=yes	1	0.079645	3
Age of onset (38; p_{25}: 30; p_{75}: 45)	0≥38 1≤38	1	0.039180	3
DIMENSION 2: Personal and family				
Posttraumatic stress disorder	0=No 1=yes	2	0.574638	3
Personality disorders	0=No 1=yes	2	0.573018	3
Family history of autoimmune disorders	0=No 1=yes	2	0.555378	3
Family history of chronic fatigue syndrome	0=No 1=yes	2	0.554531	3
Panic attacks	0=No 1=yes	2	0.491861	3
Family history of fibromyalgia	0=No 1=yes	2	0.466244	3
Blackouts	0=No 1=yes	2	0.456721	3
Facial oedema	0=No 1=yes	2	0.432674	3
Connective disorder	0=FM	2	0.395716	3
Adjustment disorder	0=No 1=yes	2	0.363677	3
Previous Personal history psychopathology	0=No 1=yes	2	0.321596	3
Major depression	0=No 1=yes	2	0.295372	3
Family history of chronic pain	0=No 1=yes	2	0.188152	3
Impaired urination	0=No 1=yes	2	0.155414	3
Spine osteoarthritis	0=No 1=yes	2	0.116031	3
Life quality SF-36 physical subscale (27; p_{25}: 22; p_{75}: 32)	0≥27 1≤27	2	0.071334	3
Life quality SF36 mental subscale (35; p_{25}: 25; p_{75}: 48)	0≥35 1≤35	2	0.004091	1
DIMENSION 3: Scales				
HAD depression subscale (10; p_{25}: 7; p_{75}: 14)	0≤10 1≥10	3	0.266515	2
Fibromyalgia Impact Questionnaire (FIQ) (74.66; p_{25}: 63.05;	0≤74.66	3	0.249480	
Fatigue Impact Scale (FIS) (66; p_{25}: 56.50; p_{75}: 75.00)	0≤66 1≥66	3	0.241993	1
Pain level (VAS 1-10 cm) (7.5; p_{25}: 6.5; p_{75}: 8.5)	0≤7.5 1≥7.5	3	0.196313	2
Pittsburgh Sleep Quality Index (PSQI) (14; p_{25}: 11; p_{75}: 17)	0≤14 1≥14	3	0.194950	1
Fatigue level (VAS 1-10 cm) (8; p_{25}: 6.4; p_{75}: 9)	0≤8 1≥8	3	0.191182	1
HAD anxiety subscale (12; p_{25}: 8; p_{75}: 15)	0≤12 1≥12	3	0.105136	1
Number of Tender Points (13; p_{25}: 13; p_{75}: 18)	0≤16 1≥16	3	0.055568	1
Trembling	0=No 1=yes	3	0.043791	1

The variables included in each dimension are listed and sorted by their weighted contribution. For continuous variables, the median was used as cut-off value for binary codification. Weight represents the relationship between the similarity of the variable to the other variables of the dimension and its similarity to the remaining variables. Neighbour refers to the closest dimension, excluding the one in which the variable was included.

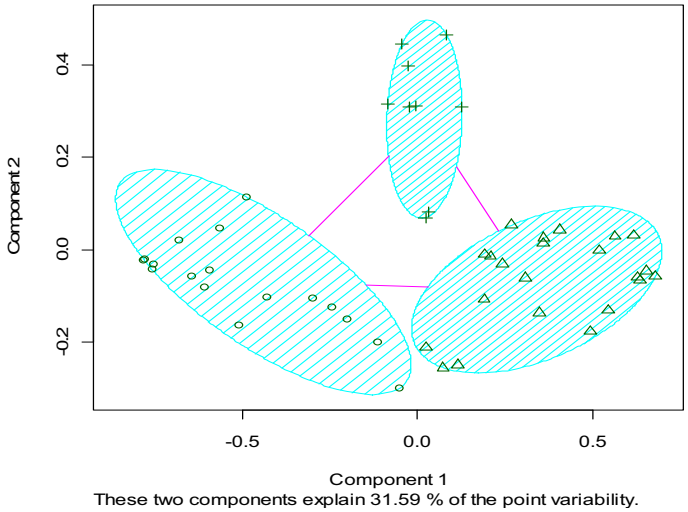


Figure 36: Clustering of variables into three dimensions. Crosses, triangles and circles represent variables assigned to each of the dimensions.

Cases were classified into three subgroups: low symptomatology and low levels of familial and personal comorbidities (Cluster 1; 283 cases, 20.2%), high symptomatology and high comorbidities (Cluster 2; 357 cases, 25.6%), and high symptomatology but low comorbidities (Cluster 3; 758 cases, 54.2%) (Figure 37).

The resulting FM subgroups presented no differences in terms of gender, age, marital status, or employment. However, patients having higher levels of education (high school degree and above) were more represented in cluster 1 than the ones having lower levels (65% vs. 35%; p=0.015 chi-square test). This symptom-based classification correlated with the data from the scales measuring pain, fatigue, psychiatric symptoms, and their impact in life, since individuals belonging to the low symptomatology and low comorbidities group also had lower medians for the scales (Table 20).

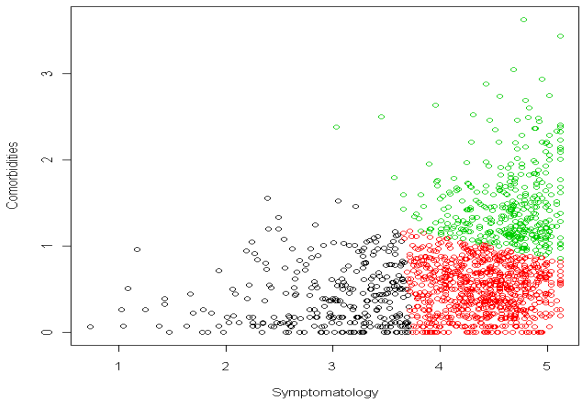


Figure 37: Subgrouping of fibromyalgia (FM) samples based on scores in "comorbidities" (Y axis) and "symptoms" (X-axis). Circles represent FM patients and are colored by cluster based on the classification by Kmeans. Black circles are samples of Cluster 1 (Low levels of symptoms and low levels of comorbidities), green circles Cluster 2 (High levels of symptoms and high levels of comorbidities), and red circles Cluster 3 (High symptoms and low comorbidities).

When the relationship between FM subgroups and core measures of severity was analyzed, Cluster 1 was markedly different from Cluster 2 and Cluster 3 in all scales, identifying the less affected group. Cluster 2 was more affected than Cluster 3, but differences were not as significant as when comparing Cluster 1 with the other two groups (Table 19).

The FM subgroup with high pain and comorbidities was also the one with the highest level of fatigue (fatigue VAS was higher than pain VAS) and, in fact, 20% of the patients in this cluster fulfilled also chronic fatigue syndrome criteria (FM + CFS) whereas in the other two clusters only 8% (Cluster 1) and 11% (Cluster 3) of the patients were fulfilling CFS criteria. Although pain VAS was higher in the high pain and comorbidities cluster, the number of tender points was not significantly different between both high pain clusters.

Table 19: Medians of different pain, psychiatric and quality of life scores in each of the fibromyalgia clinical subgroups. p-value of multinomial analysis.

VARIABLE	FM Cluster 1 (N=283)	FM Cluster 2 (N=357)	FM Cluster 3 (N=758)	P-value
Fibromyalgia Impact Questionnaire (FIQ)	59.46±19.53	77.69±13.67	73.16±14.38	<2.2E-16
Fatigue Impact Scale (FIS)	65.10±12.26	69.02±10.81	53.37±16.46	<2.2E-16
Pain level (VAS 1-10 cm)	6.30±1.96	7.88±1.52	7.43±1.65	<2.2E-16
Fatigue level (VAS 1-10 cm)	6.33±2.43	7.98±1.76	7.43±1.80	<2.2E-16
Number of Tender Points	13.54±3.61	15.54±2.61	15.24±2.77	1.26E-15
Life quality SF-36 physical subscale	30.7±10.50	26.71±6.73	26.86±7.66	7.42E-12
Life quality (SF36) mental	42.12±14.45	31.73±12.57	36.08±13.57	<2.2E-16
HAD anxiety subscale	9.54±4.63	13.03±4.35	12.12±4.33	<2.2E-16
HAD depression subscale	7.48±4.58	11.48±4.52	10.39±4.53	<2.2E-16
Pittsburgh Sleep Quality Index (PSQI)	11.28±4.41	15.35±3.44	14.27±3.89	<2.2E-16
Years of disease evolution	11.14±9.64	13.60±10.80	11.63±18.50	0.07

GENOMEWIDE ASSOCIATION STUDY

SNP genotyping analysis

1. Population structure

There was no evidence of population stratification in the samples considered for association analysis, after removal of PCA outliers, as illustrated by the QQ plot (Figure 38) and the genomic inflation value (λ=1.013).

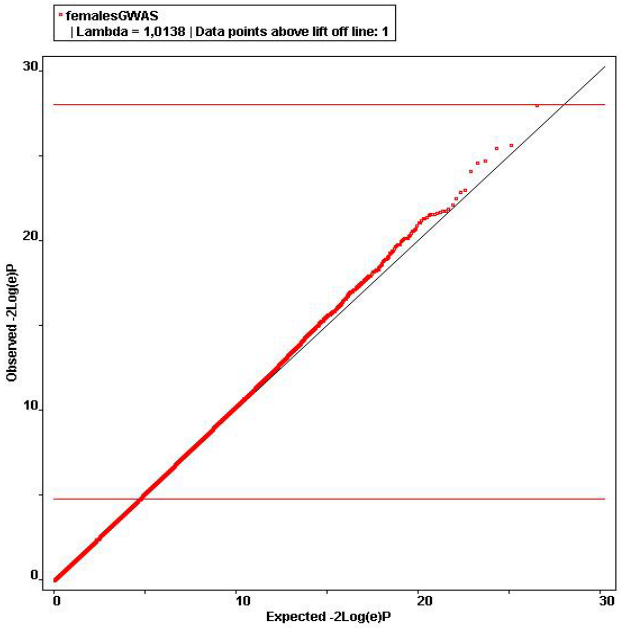


Figure 38: QQ plot of FM GWAS. Observed pvalues are plotted against expected pvalues in an association study of 500 k SNPs. The almost perfect correlation between observed and expected values was indicative of absence of population stratification as proven by a genomic inflation (λ) value of 1.013. Figure obtained with WGA viewer software.

2. Genome-wide association study findings

After QC procedures, 505,454 SNPs were considered for the anaysis in 300 FM and 203 controls. We performed allelic association analyses to find loci associated with FM. No SNP reached GWAS significance. 8 SNPs showed a pvalue <1X10^{-5} and 69 had a pvalue <1X10^{-4}. SNPs with better pvalues are summarized in Table 20. Manhattan plot (Figure 39) showed possible signals at chromosome 3 and X.

Table 20: SNPs showing the strongest allelic associations (pvalue<10-4).

SNP	P-value	Chromosome	Coordinate (Hg18)	Gene/Region
rs12556003	2,14X 10-6	X	138743267	MCF2
rs12704506	3,20X 10-6	7	89621311	STEAP1/STEAP2
rs11923054	3,52X 10-6	3	167051769	ZBBX
rs2858166	3,61X 10-6	X	100875273	ARMCX6
rs10782344	3,63X 10-6	6	156778660	RP11-518I13.1
rs1998709	7,62X 10-6	10	95884574	PLCE1
rs2194390	8,06X 10-6	2	50902931	NRXN1
rs2701106	8,94X 10-6	12	114697547	TBX5
rs9525923	1,11X 10-5	13	44783715	RP11-478K15.2
rs1347532	1,12X 10-5	16	60615455	RP11-51O6.1
rs12486010	1,53X 10-5	3	166942627	ZBBX
rs7616572	1,61X 10-5	3	167046536	ZBBX
rs10894241	1,75X 10-5	11	130635852	AP003486.1
rs11925091	2,00X 10-5	3	166944651	ZBBX
rs17512210	2,21X 10-5	17	11230466	SHISA6
rs5951332	2,35X 10-5	X	100743826	ARMCX4
rs7060491	2,35X 10-5	X	100754149	ARMCX4
rs9381682	2,48X 10-5	6	48620238	AL391538.1
rs17689185	2,50X 10-5	16	77525081	AC025284.1

rs11127292	2,60X 10-5	2	2029943	MYT1L
rs6523526	2,64X 10-5	X	100917910	ARMCX2
rs3784820	2,76X 10-5	16	1569252	IFT140
rs6621083	2,81X 10-5	X	100760626	OTTHUMG00000022030
rs10432656	2,82X 10-5	2	33375032	OTTHUMG00000152118
rs2071222	3,00X 10-5	X	100617372	LRG_128
rs11971008	3,05X 10-5	7	82136025	CACNA2D1
rs858939	3,10X 10-5	2	50971951	NRXN1
rs11187789	3,23X 10-5	10	95871655	RP11-162K11.4
rs13068321	3,46X 10-5	3	167013777	ZBBX
rs963618	3,79X 10-5	X	100743037	ARMCX4
rs9296606	3,91X 10-5	6	48640714	AL391538.1
rs12770855	4,05X 10-5	10	31120198	ZNF438
rs10821659	4,06X 10-5	10	61793424	ANK3
rs265015	4,12X 10-5	4	96360796	UNC5C
rs882847	4,39X 10-5	17	4382729	SPNS3
rs10507243	4,43X 10-5	12	114708798	TBX5
rs9565180	4,53X 10-5	13	76231470	OTTHUMG00000017093
rs4680657	4,61X 10-5	3	166894537	AC112501.2
rs1994979	4,68X 10-5	17	4350990	SPNS3
rs259154	4,72X 10-5	7	89626822	OTTHUMG00000065036
rs13238853	4,73X 10-5	7	135959346	OTTHUMG00000155618
rs4148965	4,75X 10-5	18	9109484	NDUFV2
rs6043433	5,00X 10-5	20	15659486	MACROD2
rs6537129	5,49X 10-5	4	143779613	INPP4B
rs9299090	5,68X 10-5	9	9264932	AL353733.1
rs5951269	5,99X 10-5	X	100778274	ARMCX4
rs11869601	6,19X 10-5	17	11234035	SHISA6
rs9410632	6,27X 10-5	9	90400909	CTSL3
rs259152	6,42X 10-5	7	89626611	OTTHUMG00000065036
rs8034595	6,43X 10-5	15	96719229	AC016251.2
rs4986649	6,43X 10-5	X	100736761	ARMCX4
rs265018	6,52X 10-5	4	96362497	UNC5C
rs7022749	6,61X 10-5	9	90405495	CTSL3
rs309853	6,62X 10-5	8	29873603	OTTHUMG00000163815
rs10507833	6,97X 10-5	13	76226139	OTTHUMG00000017093
rs4910595	7,15X 10-5	11	4049129	STIM1
rs11602757	7,15X 10-5	11	4053881	STIM1
rs2920137	7,15X 10-5	11	4079318	STIM1
rs6719219	7,19X 10-5	2	2010779	MYT1L
rs1938204	7,24X 10-5	6	48787015	AL391538.1
rs6083017	7,28X 10-5	20	23119766	RP4-737E23.4
rs12588013	7,29X 10-5	14	62724837	AL390816.1
rs9643612	7,39X 10-5	8	50430756	RP11-738G5.1
rs2065703	7,53X 10-5	20	31966698	CDK5RAP1
rs7314743	7,55X 10-5	12	114718647	TBX5
rs2009626	7,84X 10-5	3	187600404	RP11-44H4.1
rs981524	7,94X 10-5	14	33186257	AKAP6
rs6778044	8,16X 10-5	3	187594092	RP11-44H4.1
rs5951340	8,24X 10-5	X	100771055	ARMCX4
rs1323851	8,36X 10-5	1	64450437	ROR1
rs12744386	8,80X 10-5	1	24168019	HMGCL
rs6966421	9,18X 10-5	7	155329530	CNPY1
rs6556373	9,25X 10-5	5	158359476	EBF1
rs2997370	9,42X 10-5	6	48778106	AL391538.1
rs10184672	9,42X 10-5	2	11198448	AC062028.1
rs2009627	9,68X 10-5	3	187600359	RP11-44H4.1
rs5991939	9,73X 10-5	X	100712422	ARMCX4

SNPs are listed on descending order based on their pvalue. SNPs selected for replication appear in bold.

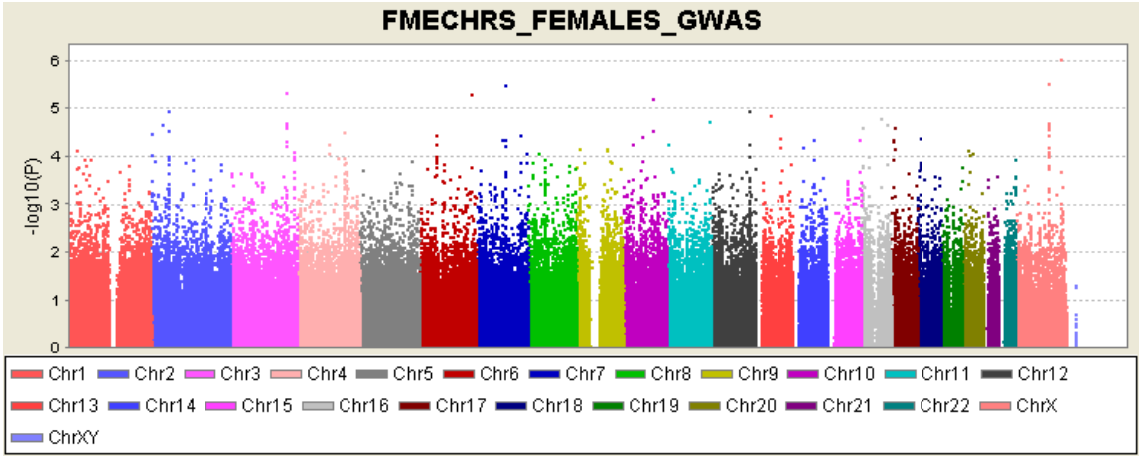


Figure 39: Summary of FM genome-wide association scan results. Negative LOG_{10} p-values across the genome and by chromosome are shown.

3.Pathway analysis of top associated SNPs

3.1 Ingenuity Pathway Analysis

The top 77 SNPs (pvalue <1X10^{-4}) listed in table 21, were selected to perform pathway analysis. Out of the 77 gene IDs introduced, only 53 were mapped. IPA pathway analysis identified two top networks that were overrepresented in our geneset: reproductive system disease (16 genes) and neurological disease (12 genes) (Figure 40).

The analysis of canonical pathways showed overrepresentation ($p<0.05$) of leucine degradation I, ketogenesis, p7036K signaling, P13K Signaling in B-lymphocytes and D-myo-inositol (1,4,5)-trisphospate biosynthesis. The top molecular and cellular function was cellular development with 8 molecules (p=1.24E-04-4.84E-02).

ID	Molecules in Network	Score	Focus Molecules	Top Functions
1	**ADAMTS18, ANK3, ARMCX2, ARMCX6,** C6orf72, **CDK5RAP1, CNPY1,** DHRS1, **EBF1, HMGCL, IFT140, INPP4B,** KIAA2013, **MACROD2,** miR-124-3p (and other miRNAs w/seed AAGGCAC), miR-193b-3p (and other miRNAs w/seed ACUGGCC), miR-26a-5p (and other miRNAs w/seed UCAAGUA), miR-4727-5p (and other miRNAs w/seed UCUGCCA), miR-495 (and other miRNAs w/seed AACAAAC), miR-9-5p (and other miRNAs w/seed CUUUGGU), **MYT1L***, NDUFA6, NDUFB10, NDUFB11, **NDUFV2,** PIGU, REEP4, **ROR1, SHISA6***, SLC10A4, SLC16A13, SPG11, **SPNS3***, STX10, UBC	39	16	Reproductive System Disease, Hereditary Disorder, Cancer
2	Actin, **AKAP6,** ATP6AP2, Beta ark, **BTK,** C1orf216, **CACNA2D1,** CACNA2D4, ERK, FAIM, FAM49A, FLVCR1, GALR1, GNGT1, HS3ST5, IBTK, **LMO7*, LTBP1, MCF2,** miR-133a (and other miRNAs w/seed UUGGUCC), miR-377-3p (miRNAs w/seed UCACACA), NFkB (complex), NMUR2, **NRXN1*, PLCE1,** Ras homolog, S1PR4, **STIM1***, SYT13, **TBX5***, TMEM87B, **UNC5C***, WASP, WHSC2, **ZBBX***	27	12	Neurological Disease, Carbohydrate Metabolism, Small Molecule Biochemistry

Figure 40: Networks identified by IPA software form GWAS top associated SNPs (image taken from IPA software output). Genes included in our geneset appear in bold.

3.2 Geneset

Only 31 of the 77 gene IDs were mapped in Geneset analysis Toolkit to perform GO analysis. This analysis identified two main molecular functions: protein binding and ion binding (Figure 41), with 16 and 14 genes, respectively; in these categories, the analysis also identified metal ion binding as the only statistically significant overrepresented molecular function (p-adjusted<0.05, Benjamini correction) (Figure 42). Most of the mapped genes (18/31) encoded membrane or membrane related proteins, thus GO identified the membrane as the top cellular component (Figure 41), and a statistically significant overrepresentation of genes in the calcium channel complex. No biological process showed a statistically significant overrepresentation in the list of genes used for the analysis. The most relevant (although non-significant) biological processes related to our gene list included calcium mediated signalling and calcium ion transport, cardiac muscle development and (Figure 42).

Finally, Kegg pathways analysis (with N mapped gene Ids) indicated enrichment in inositol phosphate metabolism, phosphatidyl inositol signalling and metabolic pathways.

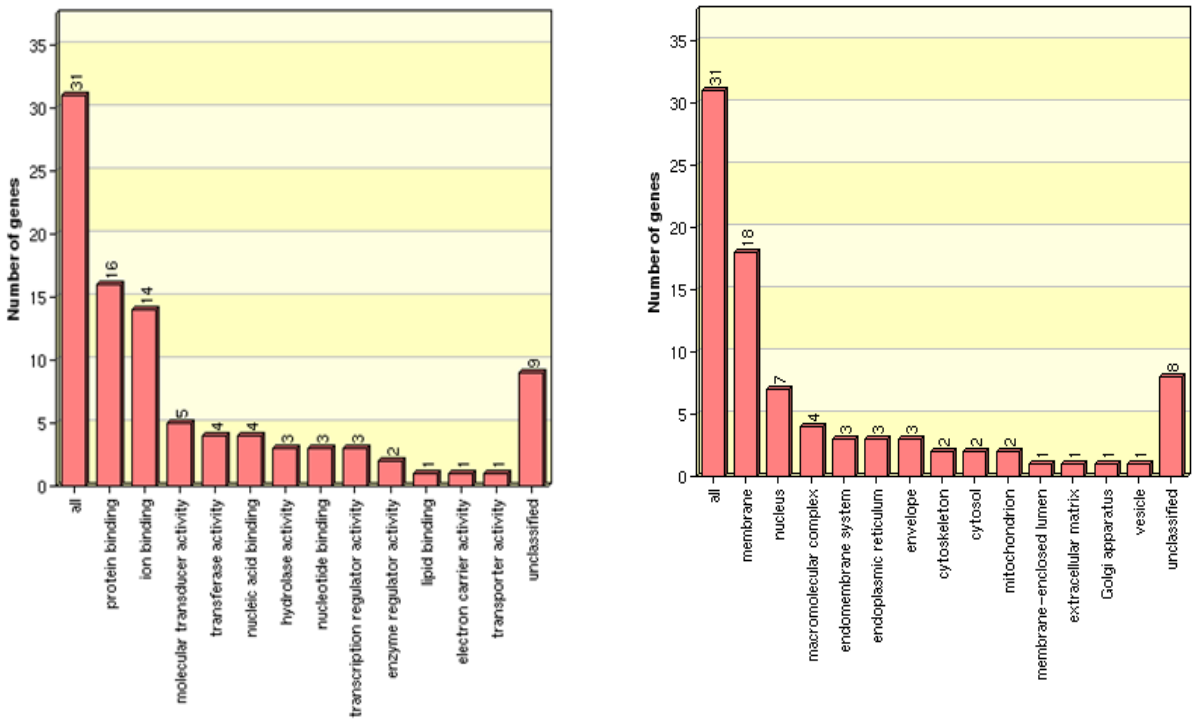


Figure 41: Bar chart of molecular function categories (left image). Number of genes is plotted against each category. Bar chart of cellular component categories (right image). Number of genes is plotted against each category. Plots taken from Geneset GO output.

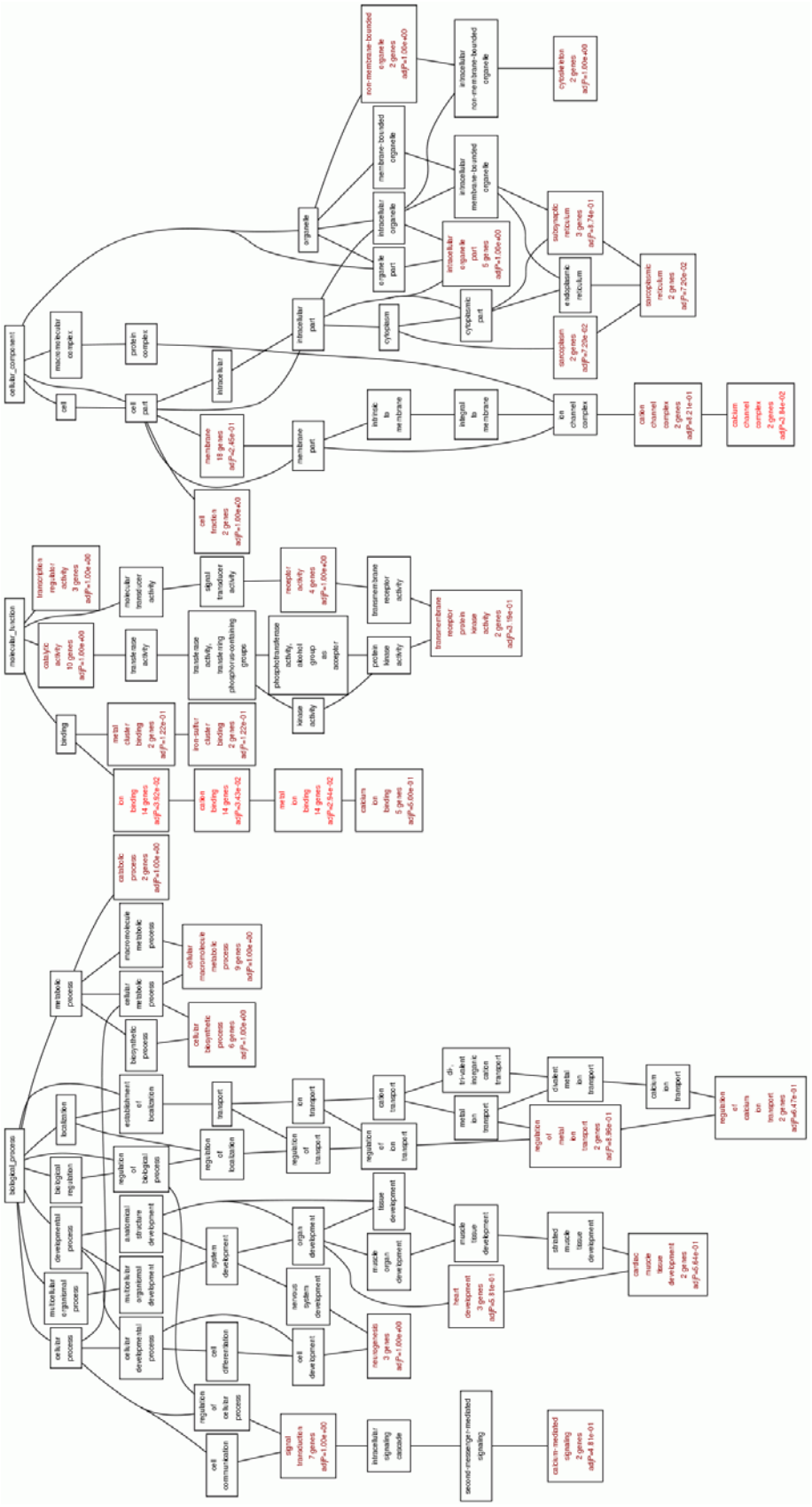


Figure 42: Geneset output from GO analysis. Most represented biological processes, molecular functions and cellular components are represented. Brown categories are top ten GO categories; statistically significant enriched GO categories (p-adjusted<0.05, Benjamini correction) appear in red.

4. Imputation

505454 SNPs considered in GWAS for association analysis were used for imputation, and 8126238 SNPs were imputed. Of these, 5012343 SNPs were selected for having a good imputation quality (frequentist info score> 0.8) and finally 4540904 for having a minimum allele frequency higher than 0.05. Since the number of cases and controls was small, imputation data were only used for fine mapping of signals detected in direct genotyping. Three imputed SNPs, showing stronger association than directly genotyped SNP, were included for replication.

5. GWAS replication

After QC procedures, 20 SNPs in 968 cases and 937 controls were considered for the association analysis (rs290761 was discarded for not fulfilling HWE, $p<0.05$). We performed allelic association in the whole replication set and afterwards only in female samples. None of the selected SNPs was significant in the replication analysis, since the correction for multiple testing established the significance threshold at 0.0023 (Bonferroni correction) (Table 20).

rs1998709 (p=0.025 in replication cohort), rs7963168 (p=0.039 in the replication cohort), rs12704506 (p=0.01 in the females replication cohort) and rs265015 (p=0.016 in the females replication cohort) showed nominally significant associations in the replication study. However, the association for these SNPs in the replication cohort was for the opposite allele of the one associated in the GWAS analysis. In fact, when adding the GWAS data to the replication data, these associations were lost.

We then selected the 4 SNPs showing the strongest combined p-values and performed an association test in the subsets identified by the aforementioned cluster analysis. We first performed a multinomial analysis in order to test for a possible association between each SNP and the FM clusters generated using the entire FM cohort (GWAS samples and samples included in the replication set) (Supplementary tables 1-4). This showed that rs11127292 genotypes in cluster 3 were different from those in cluster2, although this result was not statistically significant. Nevertheless, we decided to perform association analysis for rs11127292 separately in each FM cluster, both in the GWAS samples, in the replication cohort and with a final joint analysis (Table 21). In this analysis, rs11127292 showed a stronger association in female FM cases belonging to cluster 3 than those belonging to cluster2, although the observed differences were not statistically significant (although the OR point estimate of cluster2 association wasn't included in the cluster 3 OR 95% confidence interval OR 95% confidence intervals overlapped between groups).

Table 21: SNPS selected for replication

SNP	Rank	Type	Gene/region	P GWAS	P repli	CombinedP	P repli fem	Combined P fem
rs12556003	1	intron_variant	MCF2	2,14X10-6	-	-	0.132	1.98X10-4
rs12704506	2	intergenic_variant	STEAP1/STEAP2	3,20X 10-6	0.181	0.34	0.010	0.9
rs11923054	3	intron_variant	ZBBX	3,52X 10-6	0.458	0.0059	0.580	0.0047
rs2858166	4	5KB_downstream	ARMCX6	3,61X 10-6	-	-	0.695	0.008
rs10782344	5	intergenic_variant	RP11-518I13.1	3,63X 10-6	0.547	0.0091	0.526	0.004
rs1998709	6	intron_variant	PLCE1	7,62X 10-6	0.025	0.873	0.14	0.3397
rs2901761	IM	Intron_variant	PLCE1	5,11X 10-7	‡	‡	‡	‡
rs2194390	7	intron_variant	NRXN1	8,06X 10-6	0.803	0.073	0.702	0.049
rs2701106	8	intergenic_variant	TBX5	8,94X 10-6	0.225	0.431	0.287	0.194
rs7963168	IM	intergenic_variant	TBX5	1,49X10-8	0.039	0.63	0.08	0.28
rs17512210	15	intron_variant	SHISA6	2,21X 10-5	0.749	0.103	0.851	0.049
rs9381682	18	intergenic_variant	-	2,48X 10-5	0.738	0.029	0.156	7.38X10-4
rs11127292	20	intron_variant	MYT1L	2,60X 10-5	0.183	0.002	0.039	1.76X10-4
rs12770855	32	intergenic_variant	ZNF438	4,05X 10-5	0.422	0.01	0.149	0.001
rs10821659	33	intron_variant	ANK3	4,06X 10-5	0.182	0.002	0.110	6.22X10-4
rs265015	34	intron_variant	UNC5C	4,12X 10-5	0.073	0.781	0.016	0.852
rs9565180	37	intron_variant	LMO7	4,53X 10-5	0.898	0.093	0.242	0.267
rs6043433	43	intron_variant	MACROD2	5,00X 10-5	0.641	0.174	0.288	0.899
rs6131711	IM	intron_variant	MACROD2	9,92X10-8	0.872	5X10-4	0.596	9X10-4
rs11602757	57	intron_variant	LRG_164	7,15X 10-5	0.905	0.056	0.795	0.077
rs981524	67	intron_variant	AKAP6	7,94X 10-5	0.132	0.003	0.139	0.001

SNPs are listed in ascending order according to allelic association GWAS pvalue (with the exception of the three imputed SNPs). Type indicates the relative position of the SNP with respect to the nearest gene, and gene provides the gene in which the SNP is located or the nearest gene in a 500 kb window. IM: imputed SNP. Type: SNP location upon the gene according to WGA viewer classification; for rs12556003 and rs2858166, which are located onthe X chromosome, allelic association was only performed in females' subset. ‡rs2901761 was not in HWE nor in the whole control set (pvalue=0.01) neither in females controls (pvalue=0.005).

In fact, both in the GWAS and replication datasets, the frequency of the protective allele A in FM cases belonging to cluster 3 was slightly lower than in all the cases. As a consequence, when performing the joint analysis including all female cases (GWAS and replication) belonging to cluster 3, the overall significance was increased (p=6.2X10-5). Association analysis results in the different FM subsets for rs11127292 are summarized in table 22.

Table 22: rs11127292 allelic association in the different FM clusters in GWAS, replication and joint cohorts

rs11127292	F_FM	F_CONTROLS	P-value	OR (95%CI)
GWAS (300 FM vs 203 C)	0.051	0.125	2.6X10-5	0.37 (0.23-0.60)
Replication fem (940 FM vs 592 C)	0.091	0.114	0.03	0.77 (0.61-0.98)
GWAS+ replication fem (1240 FM vs 795 C)	0.081	0.117	1.76X10-4	0.67 (0.54-0.82)
GWAScl3 (196 FM vs 203 C)	0.045	0.125	6.21X10-5	0.33 (0.19-0.58)
Replicationcl3 (450 FM vs 592 C)	0.083	0.114	0.019	0.70 (0.52-0.94)
GWAScl3+replicationcl3 (646 FM vs 795 C)	0.071	0.117	4.28X10-5	0.58 (0.44-0.75)
GWAScl1+replicationcl1 (240 FM vs 795 C)	0.085	0.117	0.05	0.70 (0.49-1.006)
GWAScl13+replicationcl13 (886 FM vs 795 C)	0.075	0.117	4.03X10-5	0.61 (0.48-0.77)
GWAScl2+replicationcl2 (304FM vs 795 C)	0.092	0.117	0.09	0.76 (0.55-1.04)

F, frequency of the effect allele (minor allele; A or T in rs1112792); FM, female individuals; C, controls.

Finally, we evaluated possible SNPs in LD with these 4 SNPs showing the strongest pvalue. Only rs11127292 and rs10821659 showed strong LD (r^2>0.8) with other markers, as assessed by Haploview software (Supplementary figures 1&2). rs11127292 was included in a block and defined an haplotype with rs1978703 and rs1213578, tagging rs6719219 and rs11685526.

6. Functional analysis of SNPs selected for replication

We evaluated *in silico* the possible functional consequences of the SNPs selected for replication with three different tools: Puppasuite, Genevar and Regulome. Of the 24 SNPs (21 plus three in LD), Puppasuite only found functional relevance for two SNPs, rs981524 and rs7963168, which are highly conserved (Figure 43).

Genvar was used to perform cis-eQTL-SNP analysis, finding a significant association of rs12704506 with *STEAP2* transcripts levels as assessed by two expression probes in the cis-eQTL-SNP analysis (Figure 44). SNP-probe association's plot was not available for this dataset.

rs7906905 (in almost perfect LD with rs1082659, r^2=0.95) showed a significant association with *FAM13C* transcripts levels (one expression probe) in the cis-eQTL-SNP analysis (Figure 44). SNP-probe association plot in the skin of female twins showed a correlation between the SNP and *FAM13C* gene expression (Figure 45).

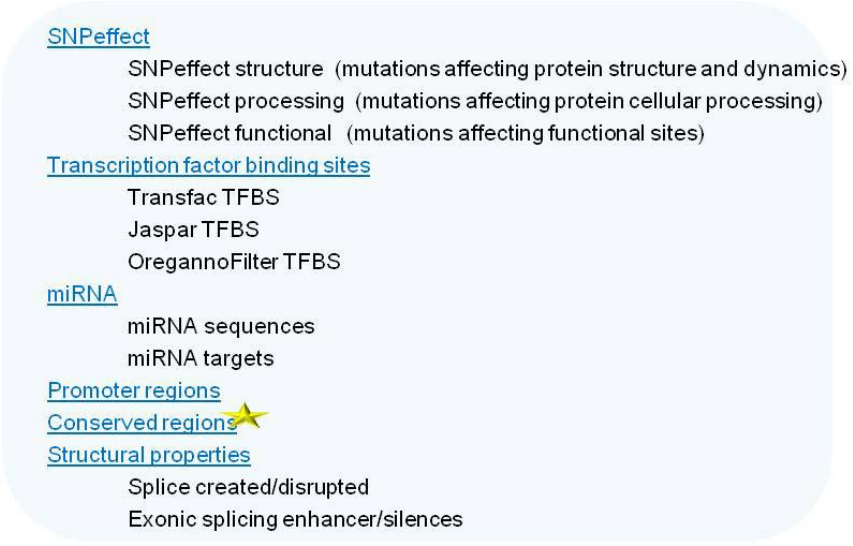
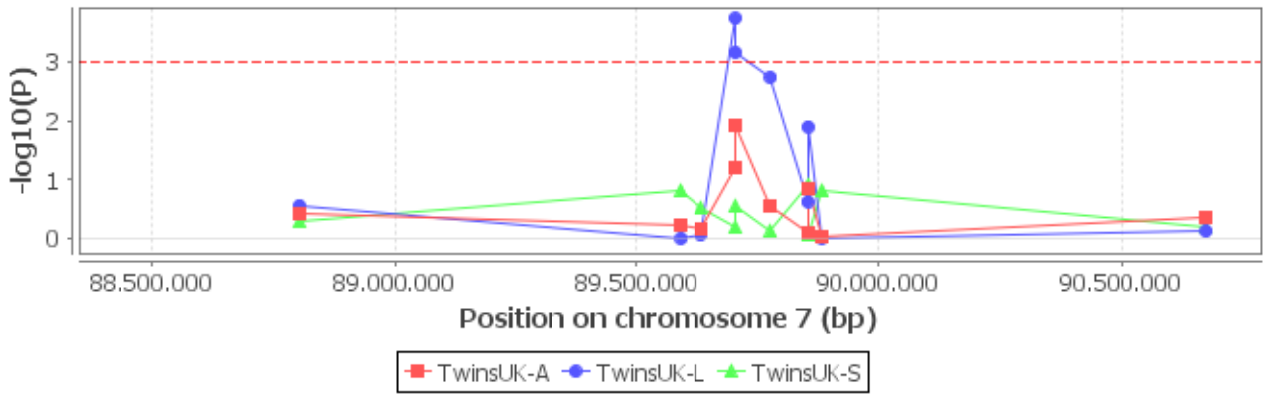


Figure 43: Summary of SNP properties evaluated by Pupasuite in SNPs showing a nominal association in the replication corhort. Only two SNPs appeared to be in conserved regions (as marked with a star).

a)



b)

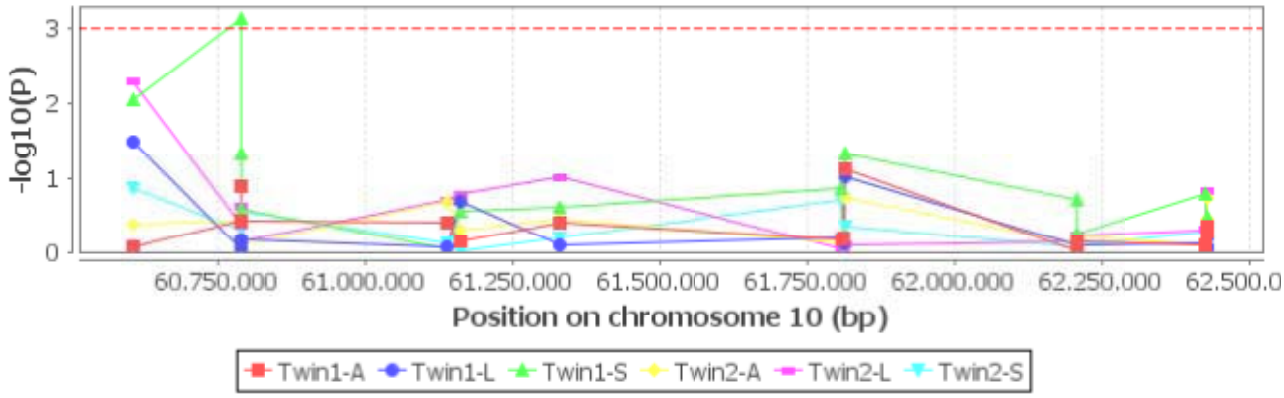


Figure 44: a) rs12704506 showed a statistical significant association with *STEAP2* expression levels in two expression array probes in lymphoblastoid cell lines of female twins. b) rs7906905 showed a statistical significant association with *FAM13C* expression levels (one probe) in the skin of female twins. Log10 of pvalue of the association between the SNPs are plotted against chromosomal position (1 MB around the SNP). Dashed line represents threshold for statistically significant association (p<0.001).

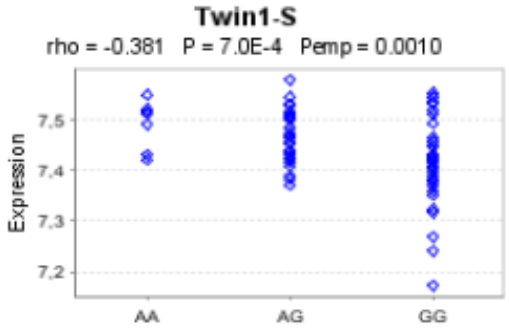


Figure 45: SNP-probe association plot for rs7906905 and ILMN_2285280 expression probe from *FAM13C* gene in female twin's skin. Expression levels are plotted against SNP genotypes.

rs11127292 showed no significant association with any expression probe. However, both this SNP and rs11685526 (in almost perfect LD with rs11127292, r^2=0.90) presented a no statistically significant correlation (p>0.001), at one expression probe, with *STNG2* gene expression in lymphocytes (see Annexes).

Regulome analysis was only performed for 10 of the 24 SNPs, as the others did not have available information. The identified functional marks were TF binding sites or DNAse peak/histone modifications with minimal binding evidence (Table 23).

Table 23: Results from regulomedb analysis.

SNPS	Functional mark
rs2901761	TF binding or DNase peak
rs17512210	TF binding or DNase peak
rs265015	TF binding or DNase peak
rs9565180	TF binding or DNase peak
rs2858166	TF binding or Histone modifications
rs7963168	TF binding or Histone modifications
rs9381682	TF binding or Histone modifications
rs11127292	TF binding or Histone modifications
rs6043433	TF binding or Histone modifications
rs11602757	TF binding or Histone modifications

CNV Assessment: PENNCNV

1. Raw data analysis

We detected a total of 18648 CNV events in 317 FM samples, with a median size of 20 kb and including a mean number of 16 SNPs (Table 24). The distribution of these CNVs among the different chromosomes was, in global, in accordance with the chromosome size (Figure 46), although there was a higher number for CNV events in chromosome 6, probably due to the HLA region.

Table 24: PennCNV results description of CNVs detected.

	Min	1st Quartile	Median	Mean	3rd Quartile	Max
NºCNVs/sample	26.00	51.00	59.00	58.83	67.00	99.00
Size (bp)	33	7212	20540	48810	54240	4451000
Nº SNPs	3.00	6.00	9.00	16.42	18.00	912.00
Copy number	0	1.0	1.0	1.86	3.0	4.0

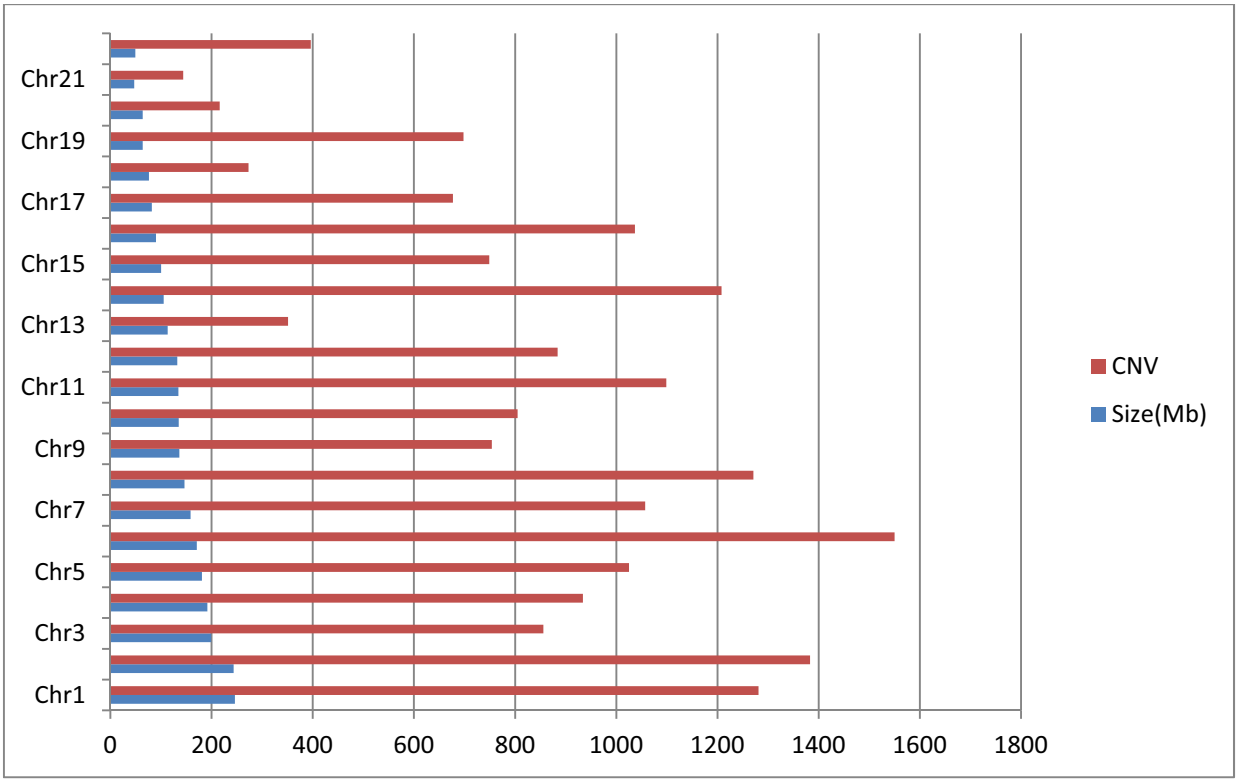


Figure 46: Chromosomal distribution of CNVs detected by PennCNV (red) and chromose sizes (data from UCSC 2003).

Out of the detected CNVs, we selected 94 that were present in at least 5% of the samples (Table 25) and then followed a prioritisation pipeline for CNV validation (Figure 47). We ended up with two CNVs that were also identified in the aCGH analysis, located in *GNG1* and *ACACA*.

A 3 Kb region within *GNG1* genomic region was detected in aCGH as a gain (Supplementary table 5), whereas PennCNV detected 59 samples with only one copy. Furthermore, although the aberration detected overlapped with previously described CNV regions, this overlap was only partial, despite the fact that both the genotyping array and the aCGH design included probes fully overlapping the published CNVs. These circumstances led us to prioritize the validation of the other CNV variant.

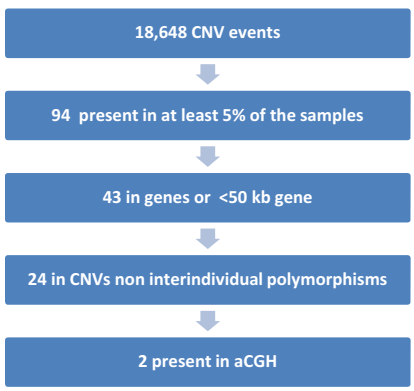


Figure 47: Pipeline for the selection of CNVs to be validated. We considered interindividual polymorphisms as those CNVs previously described as population dependent, mainly affecting genes involved in gene-environment interaction (*GSTT, ADAM*, *LCE3*, *BTLN*, *Cit-p450*, amylase metabolism, HLA, AB, antigen recognition).

Table 25: CNV regions detected in at least 5% of the samples.

Chromosome	Start	End	N SNPs	Length (bp)	State	Samples	Gene
chr1	1128776	1138758	19	9983	state5,cn=3	47	TNFRSF18
chr1	1131250	1138758	12	7509	state5,cn=3	24	TNFRSF18/TNFRSF4
chr1	1199499	1237468	37	37970	state5,cn=3	21	SCNN1D/ACAP3
chr1	1206579	1230886	22	24308	state5,cn=3	26	SCNN1D/ACAP3
chr1	16026472	16027531	4	1060	state1,cn=0	31	FBLIM1 50 kb upstream
chr1	111180501	111189246	6	8746	state2,cn=1	17	50kb upstream CD53
chr1	150828032	150850302	4	22271	state1,cn=0	9	LCE3C
chr1	167495768	167505182	3	9415	state1,cn=0	59	NME7
chr1	173064490	173068262	6	3773	state2,cn=1	58	RABGAP1L
chr1	246812825	246859583	14	46759	state2,cn=1	48	OR2T10
chr2	4191253	4200019	7	8767	state2,cn=1	78	Intergenic
chr2	34551325	34582939	13	31615	state2,cn=1	197	Intergenic
chr2	38809481	38818580	12	9100	state5,cn=3	30	GALM
chr2	41092148	41101972	10	9825	state2,cn=1	31	Intergenic
chr2	52613957	52637176	4	23220	state1,cn=0	33	Intergenic
chr2	87334785	87356521	5	21737	state5,cn=3	33	RMND5A clone similar to anaphase
chr2	89027949	89090896	6	62948	state5,cn=3	81	ab parts
chr2	146583025	146592386	3	9362	state1,cn=0	38	Intergenic
chr2	180123158	180129913	6	6756	state2,cn=1	25	ZNF385B
chr2	208061364	208066083	9	4720	state1,cn=0	60	Intergenic
chr3	37954886	37961253	6	6368	state2,cn=1	21	CTDSPL
chr3	53003023	53013826	6	10804	state1,cn=0	42	SFMBT1
chr3	65166887	65187636	12	20750	state2,cn=1	27	Intergenic
chr3	89485137	89499861	7	14725	state2,cn=1	40	EPHA3
chr3	100428761	100430538	4	1778	state1,cn=0	21	Intergenic
chr3	116143746	116150586	6	6841	state2,cn=1	28	ZBTB20
chr3	129886959	129894714	5	7756	state5,cn=3	21	Intergenic
chr3	163613385	163625169	9	11785	state2,cn=1	31	Intergenic
chr3	164004033	164101579	8	97547	state1,cn=0	37	BC073807
chr3	192548086	192552678	8	4593	state2,cn=1	70	CCDC50
chr4	10006425	10009254	6	2830	state1,cn=0	32	Intergenic
chr4	34469747	34499424	4	29678	state1,cn=0	26	Intergenic
chr4	63820936	63833261	6	12326	state2,cn=1	129	Intergenic
chr4	69064675	69117497	14	52823	state2,cn=1	58	UGT2B17
chr4	69124109	69163188	20	39080	state2,cn=1	116	Intergenic
chr4	115398433	115401739	6	3307	state1,cn=0	74	Intergenic
chr5	12864506	12866338	3	1833	state1,cn=0	30	Intergenic
chr5	19411830	19412007	3	178	state1,cn=0	26	Intergenic
chr5	46376015	46435031	7	59017	state6,cn=4	32	Intergenic
chr5	151495149	151499003	10	3855	state2,cn=1	26	AK001582
chr5	155410444	155421495	7	11052	state2,cn=1	83	SGCD
chr5	180309991	180341436	5	31446	state1,cn=0	29	BTNL8
chr6	19151979	19156771	5	4793	state2,cn=1	36	Intergenic
chr6	31388080	31396472	15	8393	state2,cn=1	40	HLAB
chr6	32561832	32592346	27	30515	state1,cn=0	249	DRB5
chr6	58856097	58878583	5	22487	state5,cn=3	44	Intergenic
chr6	67074215	67105019	22	30805	state2,cn=1	49	Intergenic
chr6	77498434	77510033	5	11600	state1,cn=0	82	Intergenic
chr6	79029649	79090197	37	60549	state2,cn=1	122	Intergenic
chr7	57951757	57985270	3	33514	state5,cn=3	28	Intergenic
chr7	61789417	61797361	5	7945	state2,cn=1	31	Intergenic
chr7	61849664	61909571	13	59908	state6,cn=4	51	Intergenic
chr7	93165330	93168493	6	3164	state2,cn=1	59	GNG1
chr7	109229030	109238466	4	9437	state1,cn=0	48	Intergenic
chr7	115724188	115727149	3	2962	state1,cn=0	49	CAV2
chr7	141412174	141435188	13	23015	state2,cn=1	32	MGAM
chr8	584761	588391	3	3631	state1,cn=0	70	ERICH1
chr8	7200170	7235238	8	35069	state5,cn=3	33	Intergenic

chr8	39351896	39499553	32	147658	state2,cn=1	193	ADAM
chr9	6691130	6695824	6	4695	state2,cn=1	43	Intergenic
chr9	22486640	22489958	3	3319	state1,cn=0	29	Intergenic
chr9	23352799	23361855	4	9057	state1,cn=0	75	Intergenic
chr9	44917247	44956619	4	39373	state2,cn=1	35	DQ594366/FAM27E3
chr10	6059021	6060197	7	1177	state5,cn=3	29	IL15RA
chr10	20890630	20897371	12	6742	state2,cn=1	65	GENE DESERT
chr10	38839401	38943702	8	104302	state2,cn=1	23	SPLICED est
chr10	66980652	66983043	5	2392	state1,cn=0	47	gene desert
chr11	18906668	18916600	6	9933	state2,cn=1	30	MGPRX1
chr11	54700151	54738983	6	38833	state5,cn=3	28	Intergenic
chr11	55124465	55165276	31	40812	state2,cn=1	89	OR4C11
chr11	81178281	81194909	11	16629	state2,cn=1	52	Intergenic
chr12	33192424	33197122	7	4699	state1,cn=0	98	Intergenic
chr12	34724272	34744278	4	20007	state5,cn=3	47	Intergenic
chr12	36627461	36651570	7	24110	state6,cn=4	43	Intergenic
chr12	39161618	39162250	7	633	state2,cn=1	71	Intergenic
chr13	33038341	33041447	4	3107	state2,cn=1	55	STARD13
chr14	40679974	40738084	13	58111	state2,cn=1	71	intergenic
chr14	105421439	105630045	35	208607	state5,cn=3	242	Abparts
chr14	105702178	105849653	45	147476	state5,cn=3	111	Abparts
chr15	19800798	19885553	21	84756	state5,cn=3	39	OR4N4
chr15	45168269	45182235	4	13967	state5,cn=3	33	Intergenic
chr15	54579805	54587099	9	7295	state2,cn=1	34	Intergenic
chr15	95616714	95633191	10	16478	state2,cn=1	42	Intergenic
chr16	2638870	2639784	6	915	state1,cn=0	32	Intergenic
chr16	33778130	33813896	14	35767	state5,cn=3	83	Intergenic
chr17	32831694	32832761	3	1068	state1,cn=0	41	ACACA
chr18	65359372	65362926	8	3555	state2,cn=1	24	DOK6
chr19	20391627	20507201	27	115575	state2,cn=1	36	ZNF826
chr19	46073341	46073380	5	40	state2,cn=1	50	CYP2A7
chr19	58212895	58244108	12	31214	state5,cn=3	53	Intergenic
chr20	28039018	28136181	18	97164	state2,cn=1	32	Intergenic
chr20	52081230	52088118	12	6889	state2,cn=1	25	BCAS1
chr22	21484058	21484640	7	583	state5,cn=3	42	ab parts
chr22	22676385	22717669	9	41285	state5,cn=3	89	GSTT

Coordinates for breakpoints, number of SNPs included in the CNV, CNV size, state (copies), number of samples presenting the aberration and gene/genomic position are provided.

ACACA

PennCNV identified a 1 kb deleted region (0 copies) in 41 FM samples. This was included in a larger CNV region detected by Conrad *et al*. (146), which had aCGH support in our datasets. By aCGH, this region appeared to have a higher frequency of the deleted allele in cases than in controls (Table 26).

Table 26: LRR of 400K array hybridizations in ACACA PennCNV region.

PROBE	CHR	START	END	GENE	FM vs C	FM vs C_DS	FM_FC vs C	FM_FC vs C_DS	FM_EARLY vs C	FM_EARLY vs C_DS
A_16_P03240326	chr17	32830147	32830206	ACACA	-0.4694158	0.5565521	-0.6497874	0.56127566	-0.14497282	-0.038216703
A_16_P20637974	chr17	32830696	32830755	ACACA	-0.333385	0.5299692	-0.34776998	0.48764184	0.12682688	0.27565214
A_16_P03240328	chr17	32831133	32831192	ACACA	-0.53137225	0.40327814	-0.3466659	0.4100373	-0.26633722	0.043815494
A_16_P20637978	chr17	32831699	32831758	ACACA	-0.5339292	0.303255	-0.59006333	0.34813052	-0.15061097	-0.0153433215
A_16_P20637980	chr17	32832198	32832257	ACACA	-0.44436887	0.49498904	-0.42604992	0.56295055	-0.10720204	0.11112005

Two FM pools (with and without fatigue) presented a deletion both in direct and dye-swap hybridizations.

First, we attempted to determine the breakpoints of this CNV region. We designed PCR primers taking into account aCGH results rather than PennCNV coordinates, as aCGH results overlapped with CNVs previously described (146, 207) and the smaller PennCNV region was included in this CNV region and couldn't be overlapping it completely as there were no Illumina's probes in the remaining region (Figure 48). We used CEU HapMap samples for which *ACACA* CNV genotypes had been determined in the publication by Conrad *et al*. (146).Only homozygous deleted and heterozygous samples amplified a 1200 bp product (Figure 49) which was sequenced. As the resulting sequences didn't allow breakpoints detection, we performed subcloning of the deleted PCR products and sequenced the purified plasmid. We mapped the CNV's breakpoints, characterizing it as a 3.1 kb deletion (chr17:32829763-32832899) (Figure 48), located in an intronic region within the *ACACA* gene. We then developed a multiplex PCR assay for genotyping and used it to genotype our cohort.

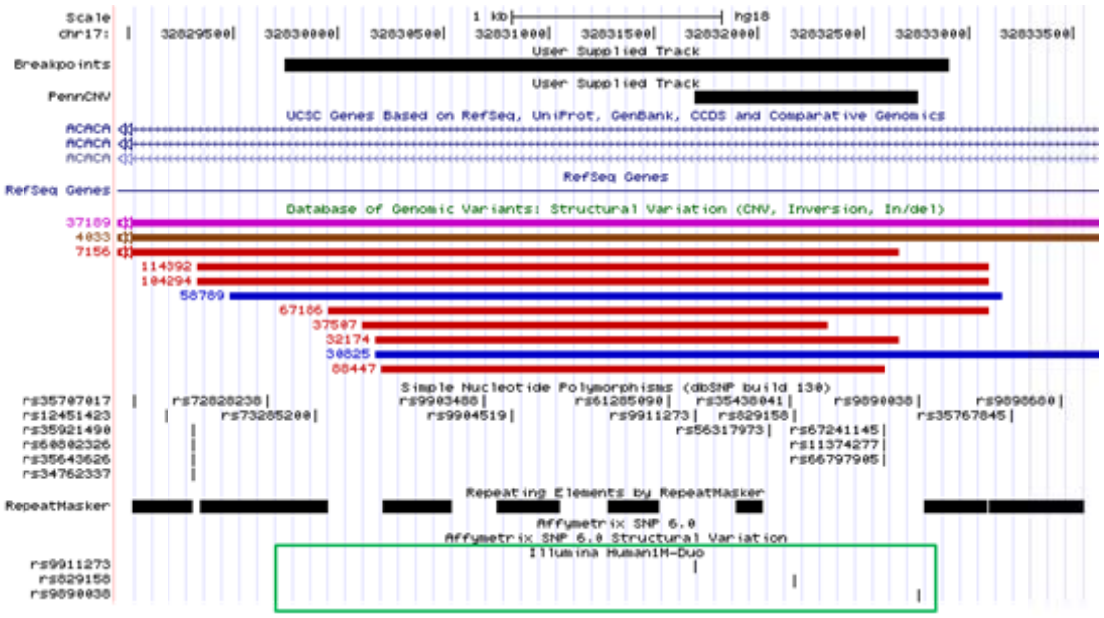


Figure 48: Genomic location of the CNV with defined breakpoints and the variant predicted by PennCNV. PennCNV only detected part of the variation because of the absence of Illumina probes in the upstream region of the CNV (green box) image taken from UCSC genome browser).

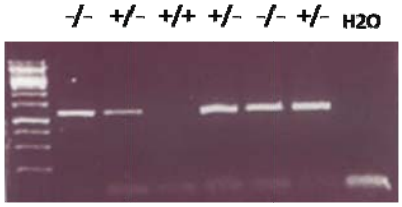


Figure 49: Homozygous deleted and heterozygous HapMap samples for the *ACACA* CNV amplified a 1200 bp product (2% agarose gel).

After genotyping 200 FM cases and controls, no differences were observed between the two groups (Table 27; genotypic association, Fisher test p=0.2). What is more, the results from these 400 individuals were in contradiction with the expected results from aCGH, as the deletion appeared more often in control samples (Del: 27.9% FM, 31.5% controls). Given these discouraging results, the analysis of this region was not pursued further.

Table 27: *ACACA* CNV genotyping results.

ACACA_CNV GENOTYPES	CONTROLS (N=211)	FM (N=209)
Del/Del	23 (10.9%)	24 (11.4%)
Del/NoDel	87 (41.2%)	69 (33.4%)
NoDel/NoDel	101 (47.8%)	116 (55.5%)

2. Analysis of rare large events

PennCNV detected the occurrence of 25 rare, large events (Supplementary table 6). Only two of these were identifed in at least two samples, and both were also detected in a set of controls of European origin (data coming from another study that was ongoing in the laboratory). Therefore, no rare CNV was selected for follow-up.

Assessment of CNVs in mosaic state

The MAD algorithm identified 25 non-LOH aberrations were detected by the algorithm. Out of these, 9 were recurrent (Table 28)

Table 28: Recurrent non-LOH aberrations in mosaic state.

Chromosome	Coordinates(Hg18)	Gene	Aberration	Samples
chr6	168278364-168297270	70 kb up FMRD1	Gain	10
chr10	135107081-135190557	-	Gain	7
chr10	44584427-44679489	-	Gain	11
chr10	47013328-47173875	ANTXRL	Gain	29
chr12	31105446-31293957	OVOS2	Gain	14
chr12	7914482-7962082	SLC2A14	Gain	11
chr17	74880641-74899077	HRNBP3	Gain	8
chr17	41602941-41706070	-	Gain	47
chr22	24086674-24173884	LRP5L	Gain	9

Five of them were located in genes. The most recurrent was located in *ANTXRL* (anthrax toxin receptor-like), which was a gene of unknown function. The aberration within *OVOS2* (Figure 50) was present in 14 samples (in 8 cases it included the whole gene and in 6 samples it involved a region 13 kb upstream of the gene).

Between 21 and 48% of the cells (depending on the sample) were carrying the gain. The region overlapped with previously reported CNVs (Figure 51). 11 samples presented a 48kb gain including the two first introns and exons of *SLC2A14.* This gene encodes a ubiquitously expressed glucose transporter. Between 22 and 43% of the cells (depending on the sample) were carrying the gain. Nine samples presented a gain in chr22: 24086674-24173884, including part of *LRP5L* (low density receptor related protein) and 8 samples carried a CNV in chr17: 74880641- 74899077, located in the intronic region of *HRNBP3* (hexaribonucleotide binding protein 3).

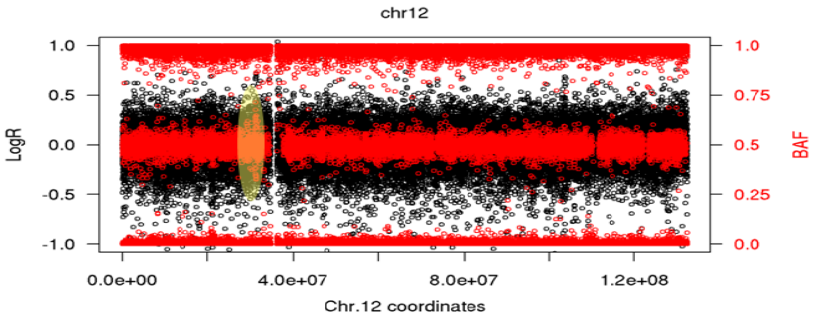


Figure 50: Aberration detected in *OVOS2.* The region (marked by a yellow oval) is identified as it presents changes in BAF without significant changes in LogR. In the image, the BAF (red dots) and LogR (black dots) are plotted against chromosomal position.

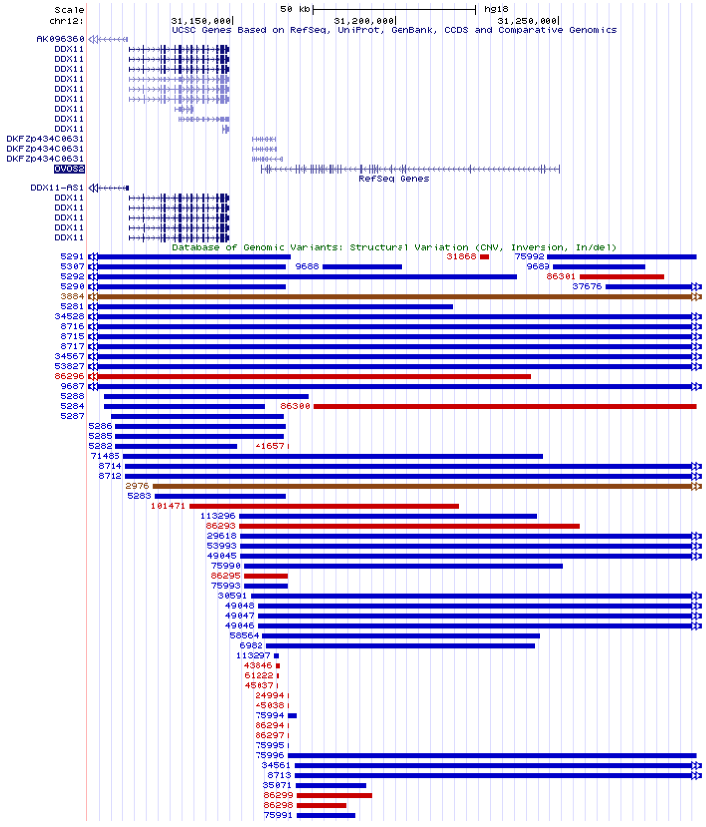


Figure 51: OVOS2 mosaic event overlapped with previously described CNVs. Image taken from UCSC genome browser.

ARRAY COMPARATIVE GENOMIC HYBRIDIZATION

400K array

1. Results

Using the ADM2 algorithm provided by Agilent's software, we detected seven regions showing differential hybridization in FM samples in comparison to controls both in direct and dye swap hybridizations (Table 29). Three regions (*WDR60*, *DOCK5* and *SIRPB1),* were not considered for replication since they persistently appeared in all the aCGH experiments performed in the laboratory (data not shown).

Table 29: 400k-CNV aCGH results. Genomic location is based on build Hg 18. DS: Dye swap.

Chromosome	Cytoband	Gene	Start	End	Probes	Log2ratio_Direct	Log2ratio_ DS	Pools
chr4	q34.1	GALNTL6	173661791	173666272	8	-0,996797	1,197165	FM early
chr7	q36.3	WDR60	158400565	158402804	9	0,483664	-0,318777	FM early
chr8	p21.2	DOCK5	25122432	25126488	6	-1,579404	1,207151	FM_fatigue
chr9	p23	PTPRD	10394403	10395130	3	-2,224173	1,828	FM
chr14	q31.1	NRXN3	79175885	79184422	18	-0,54372	0,584219	FM_fatigue FM early
chr16	q23.1	WWOX	76929398	76941774	11	-0,462331	0,542376	FM early
chr20	p13	SIRPB1	1511432	1531941	28	-0,408943	0,470049	FM_fatigue

2. Validation of aCGH results

Four regions coming from ADM2 analysis were selected for validation. An additional region, *MYO5B*, detected by the less restrictive ADM1 analysis was selected for validation because it constituted a good candidate for FM. The breakpoints for four of them had been defined previously and breakpoints coordinates were available in public databases (dbSNP and UCSC genome browser; hg18): *GALNTL6:*rs67651552 (chr4:173661608-173670094); *PTPRD***:** rs71315285 (chr9:10394565-10395094); *MYO5B:* rs72192652 (chr18:45948972-45952380) and *NRXN3:* ss49993191 (chr14:79175982-79184862).For the remaining region, *WWOX*, we had to identifie the indel's breakpoints. Then, for all regions, we designed multiplex PCR experiments to genotype them, first in a subset of 300 FM samples and 300 controls and, if there was validation of aCGH findings, in our entire case control cohort (Figure 52).

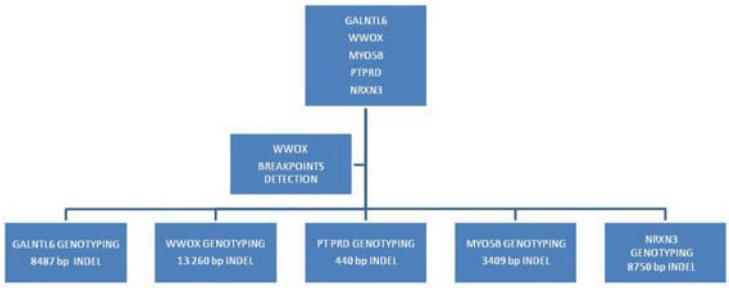


Figure 52: Five regions from aCGH results were selected for replication. Size of the variants is indicated (bp).

2.1 *GALNTL6*

The first region selected for validation was a 5 kb deletion detected in the FM_early pool that overlapped with a previously described CNV (chr4:173661608-173670094 (hg18)). Although the overlap wasn't total, this could be explained by a low density of array probes in the CNV flanking regions (Supplementary table 7). Taking advantage of a multiplex PCR assay designed in the laboratory (Figure 53), we genotyped a initial subset of 417 controls and 460 FM samples, and found nominal association at the genotypic and allelic levels (p=0.04 OR (95%CI): 0.78 (0.62-0.99) genotypic association, log additive model; p=0.04 OR (95%CI) 0.77(0.60-0.99) allelic association (Fisher test)). This wasn't confirmed when we increased the sample size to 893 controls and 1095 FM cases (p=0.55, genotypic association, log-additive model; p=0.56 allelic association (Fisher test)). Since this region was detected only in the FM_EARLY pool (FM patients with age of onset of the disease before 20 years old), we considered the possibility that the association was specific to the early onset cases. We performed a second analysis considering only early onset cases (age of onset<20) and, although we found a higher proportion of homozygous deleted samples in this subset, this was still not statistically significant (p=0.29, genotypic association, log additive model; p=0.33 allelic association (Fisher Test)) (Table 30).

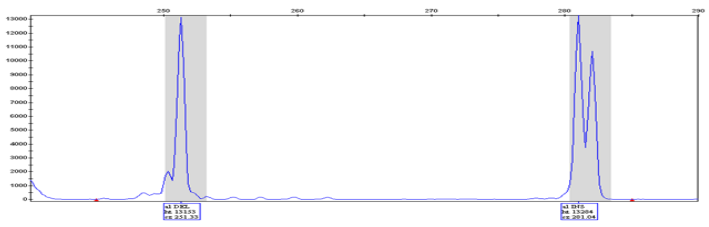


Figure 53: Gene mapper image of a heterozygous sample for *GALNTL6* CNV genotyping. Peaks intensities are plotted against the PCR products sizes (in bp).

Table 30: Genotypic distribution of *GALNTL6* CNV in a) The initial subset of cases and controls b) controls, FM cases and FM of early onset.

a)

GALNTL6 CNV GENOTYPES	CONTROLS (N=417)	FM (N=460)
Del/Del	273 (65.5%)	744(70.4%)
Del/NoDel	121 (29.0%)	317(26.7%)
NoDel/NoDel	23 (5.5%)	34 (2.8%)

b)

GALNTL6 CNV GENOTYPES	CONTROLS (N=893)	FM (N=1095)	FM early (N=101)
Del/Del	597 (66.9%)	744 (67.9%)	74 (73.3%)
Del/NoDel	265 (29.7%)	317 (28.9%)	23 (22.8%)
NoDel/NoDel	31 (3.5%)	34 (3.1%)	4 (4.0%)

2.2 *WWOX*

In the 400K_CNV aCGH experiments, analysis by the ADM2 algorithm detected an aberrant region at chr16q23.1 spanning over 11 probes. This region was shown as deleted in the FM_Early pool versus the controls both in direct and dye swap hybridizations. The region, of almost 14 Kb, overlapped with a known CNV (146). Taking into account the positions of aCGH probes (Supplementary table 8) and the chromosomal positions for the published CNV, we designed a PCR experiment in order to detect the CNV's breakpoints. As the FM_Early pool presented a loss, we genotyped the samples included in this pool, expecting to detect the deleted allele in homozygous or heterozygous states. Three combinations of primers were tested. Single band products were directly sequenced, while for unspecific PCR reactions, the PCR product was loaded in a low melting agarose gel, for band extraction and purification, and the resulting purified product was then sequenced (Figure 54).

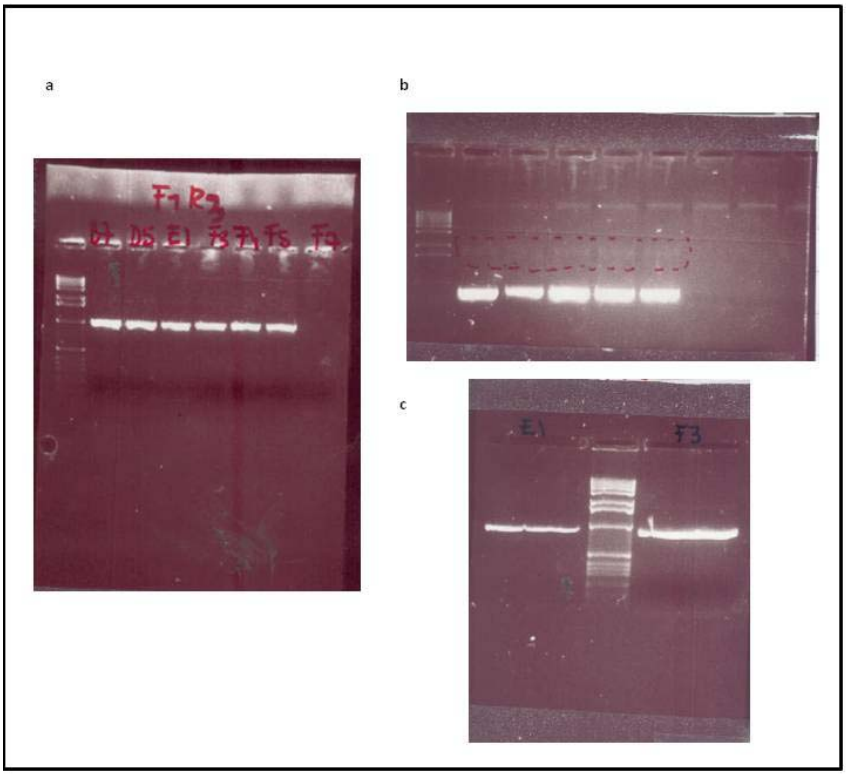


Figure 54: *WWOX* CNV breakpoints detections PCRs loaded in 2% agarose gel. a) F1R3 combination of primers resulted in a product of approximately 1000 bp. b) F1R1 combination did not amplify and F1R2 resulted in a product of approximately 1000 bp but with the presence of unspecific bands (red dashed boxes). c) Two of the samples of F1R2 reaction were loaded in a 1.3% low melting agarose gel for band isolation.

We detected the breakpoints for the CNV, characterizing it as a 13Kb deletion located in the intronic region of *WWOX* (Chr16: 76929139-76942400 (hg18), and we designed a multiplex PCR assay to genotype it (Figure 55). We genotyped 619 FM samples and 691 controls and found no statitiscally significant differences. In fact, there was a slight, non significant, increase of the non-deleted allele and the homozygous non-deleted genotype in cases (allelic association Fisher test p=0.10, OR (95%CI)=1.13 (0.97-1.32); genotypic association log additive model p=0.10, OR (95%CI)=0.88 (0.76-1.32); table 31) which was contradictory to the aCGH results .

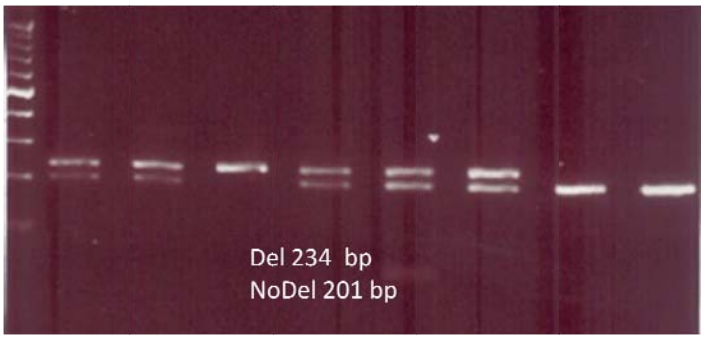


Figure 55: WWOX multiplex PCR loaded in a 3% agarose gel.

Table 31: Genotypic distribution of WWOX_INDEL in cases and controls genotyped with a multiplex PCR.

WWOX GENOTYPES	CONTROLS (N=691)	FM (N=619)
NoDel/NoDel	172 (24.9%)	169 (27.3%)
Del/NoDel	340 (49.2%)	314 (50.7%)
Del/Del	179 (25.9%)	136 (22.0%)

2.3 *PTPRD*

The CNV in *PTPRD* was a 500 bp deletion compared to the reference genome. This deletion was genotyped by a multiplex PCR. Since the deletion was small, the assay for the deleted allele (450 bp) generated a 980 bp product for the non deleted allele. For this reason, we designed a specific assay for the non deleted allele, generating a smaller product (219 bp) that ensured multiplex PCR with a balanced amplification of both alleles (Figure 56). PTPRD_INDEL genotyping of 283 controls and 303 cases didn't support aCGH findings (Supplementary table 9). We didn't find any homozygous non deleted samples and the distribution of the other two genotypes did not differ between cases and controls (Table 32; p=0.7 genotypic association log additive model, allelic association Fisher test p=0.08).

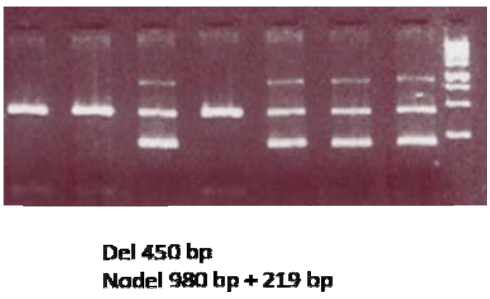


Figure 56: PTPRD_INDEL multiplex PCR loaded in a 2% Agarose gel. Two products correspond to the non deleted allele and one to the deleted allele.

Table 32: Genotypic distribution of PTPRD_INDEL among cases and controls.

PTPRD GENOTYPES	CONTROLS (N=283)	FM (N=303)
Del /Del	200 (70.7%)	218 (71.9%)
Del/Nodel	83 (29.3%)	85 (28.1%)

2.4 *MYO5B*

A 3kb gain in *MYO5B* intronic region was detected in FM pool (both in dyrect and dye swap hybridizations) with ADM1 algorithm (Supplementary table 10). As it overlapped with a previously described indel, we used the breakpoints coordinates to design a multiplex PCR(Figure 57). We genotyped it in 350 FM samples and 306 controls and found no statitiscally significant differences (allelic association Fisher test p=0.11; genotypic association log additive model p=0.9) (Table 33).

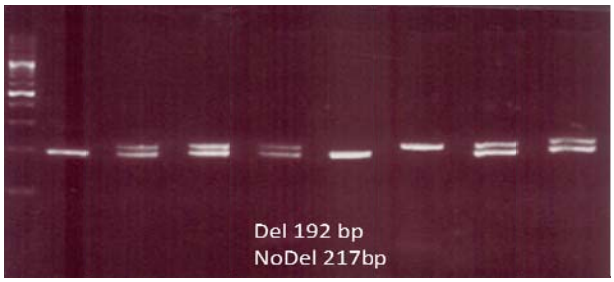


Figure 57: MYO5B_INDEL multiplex PCR product loaded in a 3% agarose gel.

Table 33: Genotypic distribution of MYOB_INDEL in cases and controls.

MYO5B_INDEL GENOTYPES	CONTROLS (n=306)	FM (n=350)
NoDel/NoDel	142 (46.4%)	161 (46%)
Del/NoDel	133 (43.5%)	153 (43.7%)
Del/Del	31 (10.1%)	36 (10.3%)

2.5 *NRXN3*

A deletion in an intronic region of *NRXN3* was the only 400K array result that was present in two FM pools (Supplementary table 11). In order to validate this enrichment of NRXN3_DEL in FM samples, we first verified the published breakpoints (chr14:79175982-79184862 (hg18)) by sequencing a HapMap homozygous deleted sample and found a slightly smaller event (chr14:79176042-79184805 (hg18)) (Figure 58). Then we designed a multiplex PCR and genotyped the CNV (Figure 59) in an initial subset of 359 FM cases and 378 controls, confirming the association of the deleted allele with FM (genotypic association,

recessive model (homozygous deleted as risk genotype) p=0.0037, OR (95%CI) = 1.74(1.19-2.54); allelic association p=0.12 Fisher Test) (Table 34). We completed the genotyping of our entire cohort with both the multiplex PCR assay and a SNP-based assay included in a Veracode experiment, finding a trend to association (genotypic association, dominant model p=0.064 OR 1.18 (0.99-1.40, allelic association Fisher Test p-value=0.07, OR (95%CI) = 1.12 (0.98-1.27)). Since 97% of FM cases were females, we performed the same analysis only considering female cases and controls and the assiociation became statistically significant (genotypic association, recessive model (homozygous deleted as risk genotype) p=0.021, OR 1.46 (1.05-2.04), log additive model p=0.014, OR 1.22 (1.04-1.47), allelic association Fisher Test p=0.015 OR (95%CI) = 1.22 (1.03-1.43) (Table 35).

We performed association analysis on the subsets generated by the cluster analysis that identified three FM subgroups. We found that there were differences among clusters, with an enrichment of the deletion in clusters 1 and (in a minor extent) 3 (interaction p-value=0.046, supplementary table 12). We performed association analyses for each of the FM clusters against the controls and found that association was statistical significant in FM samples with low levels of comorbidities (clusters 1 and 3) (genotypic association, dominant model (homozygous deleted and heterozygous as risk genotypes) p=0.009, OR (95%CI)=1.28 (1.06-1.54); allelic association Fisher test p=0.019, OR (95%CI)=1.17 (1.021-1.34)) and the association was stronger if considering only females (genotypic association, recessive model (homozygous deleted as risk genotype) p=0.019, OR (95%CI) = 1.49 (1.06-2.11), log additive model p=0.004 OR (95%CI)=1.28 (1.08-1.51), allelic association Fisher Test p=0.004 OR (95%CI) = 1.27(1.07-1.50) (Tables 36)

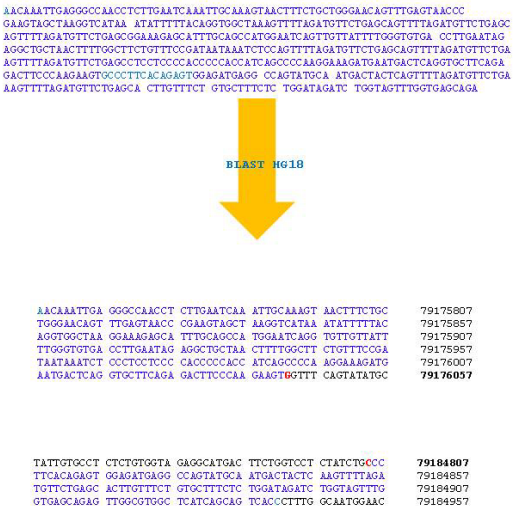


Figure 58: NRXN3_DEL breakpoints detection. NRXN3 deleted allele sequence was blasted against the human reference genome (Hg18). A 8870 bp deletion (chr14:79176042-79184805) was identified.

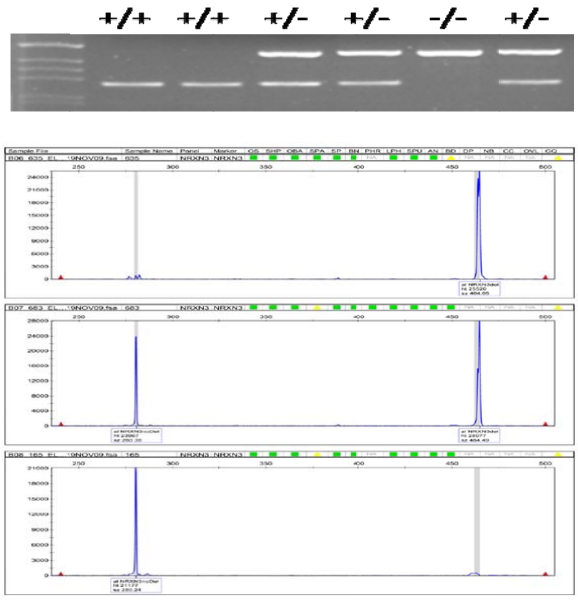


Figure 59: NRXN3_DEL genotyping. Multiplex PCR products run in a 2% agarose gel (top). Gene mapper images from homozygous deleted, heterozygous and homozygous non-deleted genotypes for NRXN3_DEL CNV (bottom).

Table 34: a) Genotypic distribution of NRXN3_DEL among cases and controls initial dataset. b) Allelic distribution of the CNV in the same subset of samples.

a)

GENOTYPES	CONTROLS N=378	FM N=359
Del/Del	55 (14.6%)	82 (22.8%)
Del/NoDel	169 (44.7%)	130 (36.2%)
NoDel/NoDel	154 (40.7%)	147 (40.9%)

b)

ALLELES	CONTROLS	FM
Nodel	477 (63.1%)	424 (59%)
Del	279 (36.9%)	294 (41%)

Tables 35: a) Genotypic distribution of NRXN3_DEL among cases and controls. b) Allelic distribution of the CNV in all the samples. c) Genotypic distribution of NRXN3_DEL among female cases and controls. d) Allelic distribution of the CNV in all the female samples.

a)

GENOTYPES	CONTROLS N=862	FM N=1397
Del/Del	118 (13.7%)	208 (14.9%)
Del/NoDel	386 (44.3%)	657 (47.0%)
NoDel/NoDel	362 (42.0%)	532 (38.1%)

b)

ALLELES	CONTROLS	FM
NoDel	1106 (64.1%)	1721 (61.7%)
Del	618 (35.9%)	1073 (38.3%)

c)

GENOTYPES	CONTROLS fem N=445	FM fem N=1358
Del/Del	48 (10.8%)	204 (15.0%)
Del/NoDel	207 (46.5%)	640 (47.1%)
NoDel/NoDel	190 (42.7%)	514 (37.8%)

d)

ALLEL ES	CONTROLS fem	FMfem
NoDel	587(65.9%)	1668(61.4%)
Del	303(34.1%)	1048(38.6%)

Tables 36: NRXN3_DEL association across the different clusters. a) Genotypic distribution among the different FM Clusters. b) Allelic distribution Clusters 1 and 3.

a)

GENOTYPES	CONTROLS N=862	FMCl1 N=277	FMCl2 N=341	FMCl3 N=730	FMCl1+Cl3 N=1007	FM N=1397	FMfemCl1+Cl3 N=980
Del/Del	118 (13.7%)	50 (18.1%)	49 (14.4%)	104 (14.2%)	154 (15.3%)	208 (14.9%)	150 (15.3%)
Del/NoDel	386 (44.3%)	132 (47.7%)	149 (43.7%)	357 (48.9%)	489 (48.6%)	657 (47.0%)	476 (48.6%)
NoDel/NoDel	362 (42.0%)	95 (34.3%)	143 (41.9%)	269 (36.8%)	364 (36.1%)	532 (38.1%)	354 (36.1%)

b)

ALLELES	CONTROLS	FM (cl1+cl3)	CONTROLSfem	FM(cl1+cl3)fem
NoDel	1106 (64.1%)	1217 (60.4%)	587 (65.9%)	1184 (60.4%)
Del	618 (35.9%)	797 (39.6%)	303 (34.1%)	776 (39.6%)

3. CNVs interaction evaluation

We evaluated a possible interaction between NRXN3_DEL and *GALNTL6* CNV and *WWOX* CNV and found that there was no evidence of interaction (Figure 60).

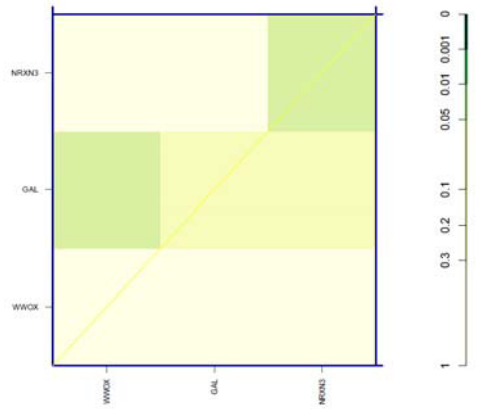


Figure 60: Plot representing interaction of NRXN3, GALNTL6 and WWOX CNVs. There was no evidence of statistical significant interaction. Multiple comparison results are presented in the plots considering the codominant model of genetic action for both tested pair CNV combinations. Statistical significances are represented by colour scale; dark-green indicates greater statistical significance and yellow-white indicates less statistical significance. The upper triangle in the matrix contains the p-values for the interaction. The diagonal contains p-values from a Likelihood Ratio Test (LRT) for the crude effect of each CNV. The lower triangle contains p-values from a LRT, comparing the two-CNV additive likelihood to the best of the single-SNP.

4. NRXN3_DEL GWAS SNPs interaction

We evaluated a possible interaction between NRXN3_DEL and the four GWAS SNPs presenting a nominal association in the joint analysis. We then performed the same analysis considering only females, in order to include the X chromosome variant rs12556003. There was no evidence of interaction; females there ws evidence of an additive effect of rs10821659 and rs11127292 (p=0.04) (Figure 61).

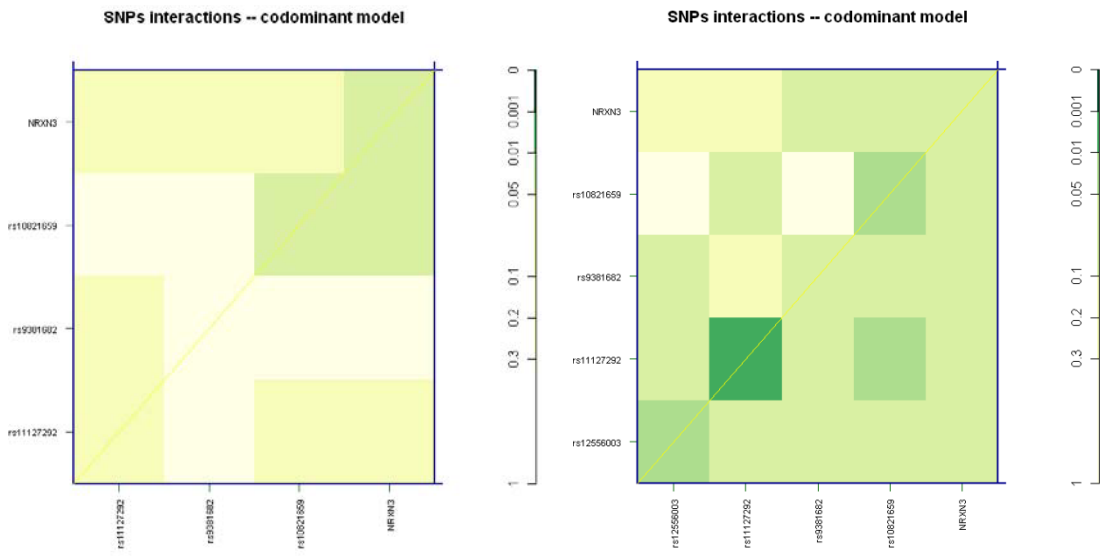


Figure 61: Plots representing interaction of NRXN3_DEL and rs11127292, rs9381682, rs10621659 in the whole cohort (left plot) and in the females subset (Codominant model).

aCGH 1 M array

For the analysis of the 1M aCGH array we used the same selection criteria as for the analysis of the 400K array, except that we only had data from direct (instead of direct and dye-swap) experiments. Five regions were detected as potentially deleted or gained between cases and controls (Table 37) in one or more pools. The CNV located near *RXRA* was discarded since FM and FM_FC presented a gain whereas FM_early presented a loss; a careful examination of the aCGH results showed that the altered probes for *SCAPER* and *SLC12A7* CNVs were actually the same probe that was duplicated. Finally *SHANK3* and *SCLA9A3* regions, were also discarded because not all probes in the regions were actually. For all these reason, we didn't undertake the validation of 1 M array aCGH results.

Table 38: 1M aCGH results. Regions detected by ADM2 algorithm, for regions including at least 3 probes and with a logRatio >0.3.

Chromosome	Cytoband	Gene	Start	End	Probes	Log2ratio Direct	Pools
chr9	q34.2	RXRA	136471996	136472396	5	0.475	FM early FM_fatigue FM
chr15	q24.3	SCAPER	74671426	74680639	4	-0.73	FM early FM
chr5	P15.33	SLC9A3	528014	567430	18	-0.41	FM early FM
chr5	p15.33	SLC12A7	1107647	1147435	15	-0.38	FM early
chr22	q13.33	SHANK3	49514241	49518515	4	-0.63	FM early

Evaluation of the possible functional consequences of NRXN3_DEL

1.Veracode assay

A veracode assay including several functional and tagging SNPs in the *NRXN3* gene was designed to test for association with other *NRXN3* polymorphisms that could have been tagged by the deletion. After quality control, 34 SNPs in 869 FM samples and 843 controls were included for analysis, as 4 SNPs failed HWE, 7 SNPs presented missingness >0.05 and 12 a MAF<5%, while 59 individuals were removed for having a low genotyping rate. After correcting for multiple testing (34 markers, p<0.00147, Bonferroni correction) no SNP showed significant association (supplementary table 13). The three assays included to genotype the NRXN3_DEL presented a 100% concordance with the results from the multiplex PCR assay. No other SNP appeared to be a proxy of the deletion. We then applied cluster analysis and found no interaction with SNPs association and the different clusters.

2.mRNA experiments

We performed a *NRXN3* PCR with cDNA obtained from two glioblastoma cell lines with a different homozygous genotype for the deletion, U87 (Del/del) and T98G (Nondel/Nondel), and including as well a neuroblastoma cell line (SH-SY) that was heterozygous for the deletion. Two products were expected, and in three independent experiments (from three different RT reactions) we found that there was a different ratio of the two products between cell lines (Figure 62).

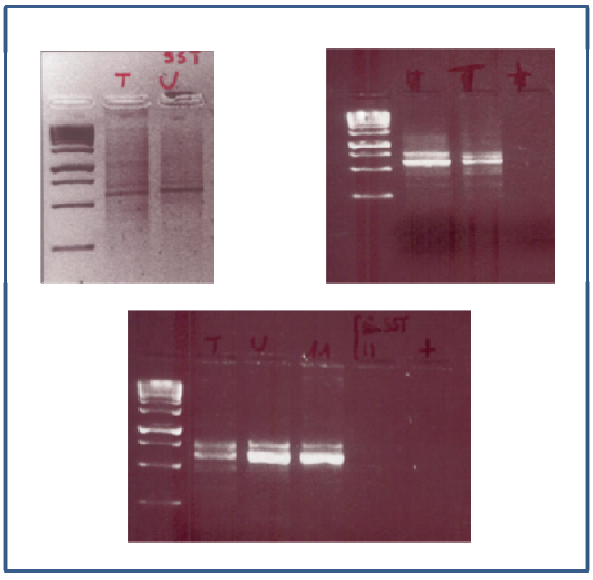


Figure 62: mNRXN3 PCR in three independent RT. T: T98G cell line; U: U987 cell line; 11:SHSY cell line: 11sin SST: RT reaction blank (without superscript); †: PCR blank.

We subcloned and sequenced each of the products and confirmed that they corresponded to the expected products: *NRXN3* with and without exon 20 (Figure 63).

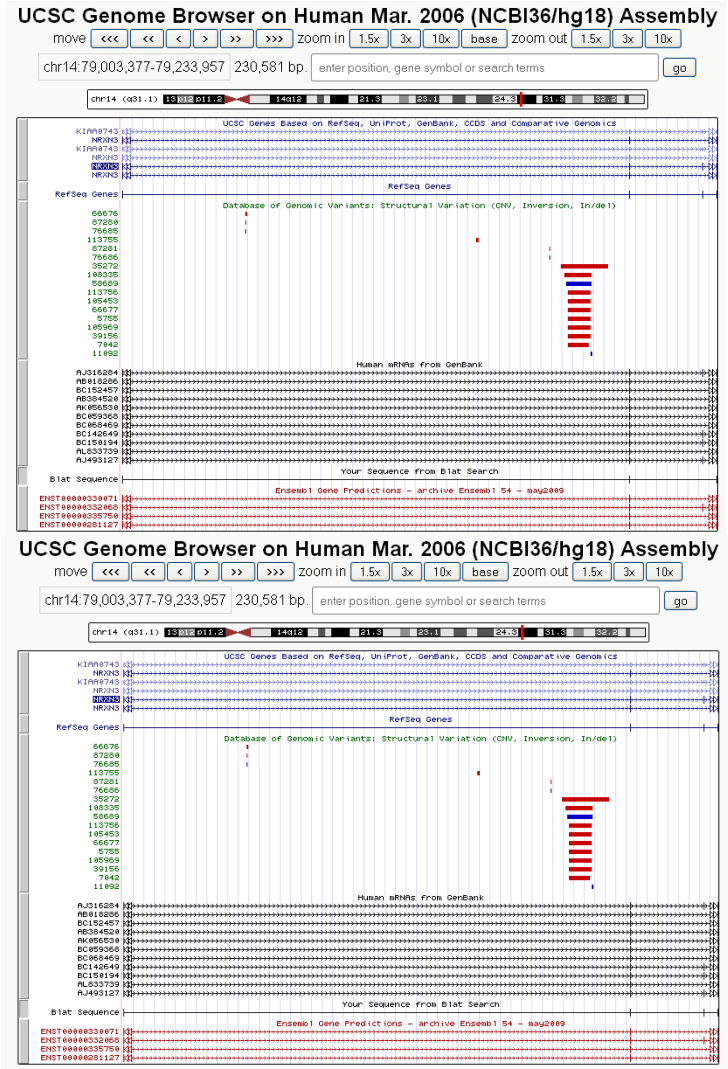
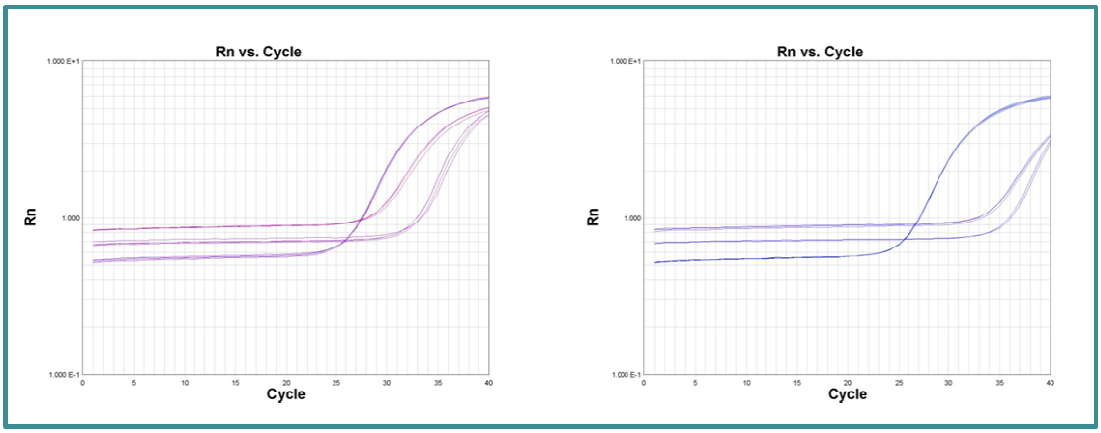


Figure 63: Genomic position of blat results of mRNA *NRXN3* sequenced products. Top image corresponds to the smaller PCR product not including exon20 and bottom image to the larger fragment including exon20. Images taken from UCSC genome browser

To quantify these differences we performed quantitative RTqPCR using commercial Taqman expression assays. Since T98G and U87 cell lines came from individuals of different gender, we used a housekeeping gene not located on the X chromosome: TBP. We performed the same experiment using cDNA from two different RT experiments and confirmed the different ratio between the *NRXN3* isoforms with and without exon 20 (Figure 64) (Table 38). Whereas both isoforms where highly expressed in U87 cell line, the transcript not including exon20 (exon 1921) was 3.63 times more expressed than the one including exon 20. These results showed an enhanced exon 20 skipping in U87, carrying the NRXN3_DEL in homozygosity. We

therefore explored a possible mechanism that could link NRXN3_DEL to alternative splicing at splicing site 4 (responsible for exon20 skipping). We found in the literature a link between alternative splicing at splicing site 4 and PTBP2 (polypirimidine binding protein 2) molecules (alternative splicing inhibitors) (208), through the specific binding of these factors in introns flanking exon20. As the paper identified PTBP2 binding sites in introns 21 and 19 we evaluated *in silico* the presence of these motifs in the deleted region and found some of these motifs. In fact, they corresponded with some of the regions with high conservation rate inside NRXN3_DEL (Figure 65).



Assay	T98G Ct	U87 Ct
TBP	26.03	26.59
Exon1921	34.23	29.18
Exon20	35.79	32.59

Figure 64: U87 (left) and T98G (right) amplifications curves for TBP, Eonx1921 and Exon21 assays. Ct for each assay in each of the two cell lines are indicated in the table below.

Table 38: Ratio of Exon 20 and Exon 1921 expression (referedto TBP housekeeping gene) in U87 cell line respect copared to T98G cell line.

	Exon20/TBP	Exon1921/TBP	Exo1921/Exon20
Ratio U87/T98G	13.23	48.08	3.63

$RATIO_{EXO20} = (Efficiency\ Exon\ 20)^{\Delta Ct\ (T98G-U87)} / (Efficiency\ TBP)^{\Delta Ct\ (T98G-U87)}$;

$RATIO_{EXO1921} = (Efficiency\ Exon\ 1921)^{\Delta Ct\ (T98G-U87)} / (Efficiency\ TBP)^{\Delta Ct\ (T98G-U87)}$

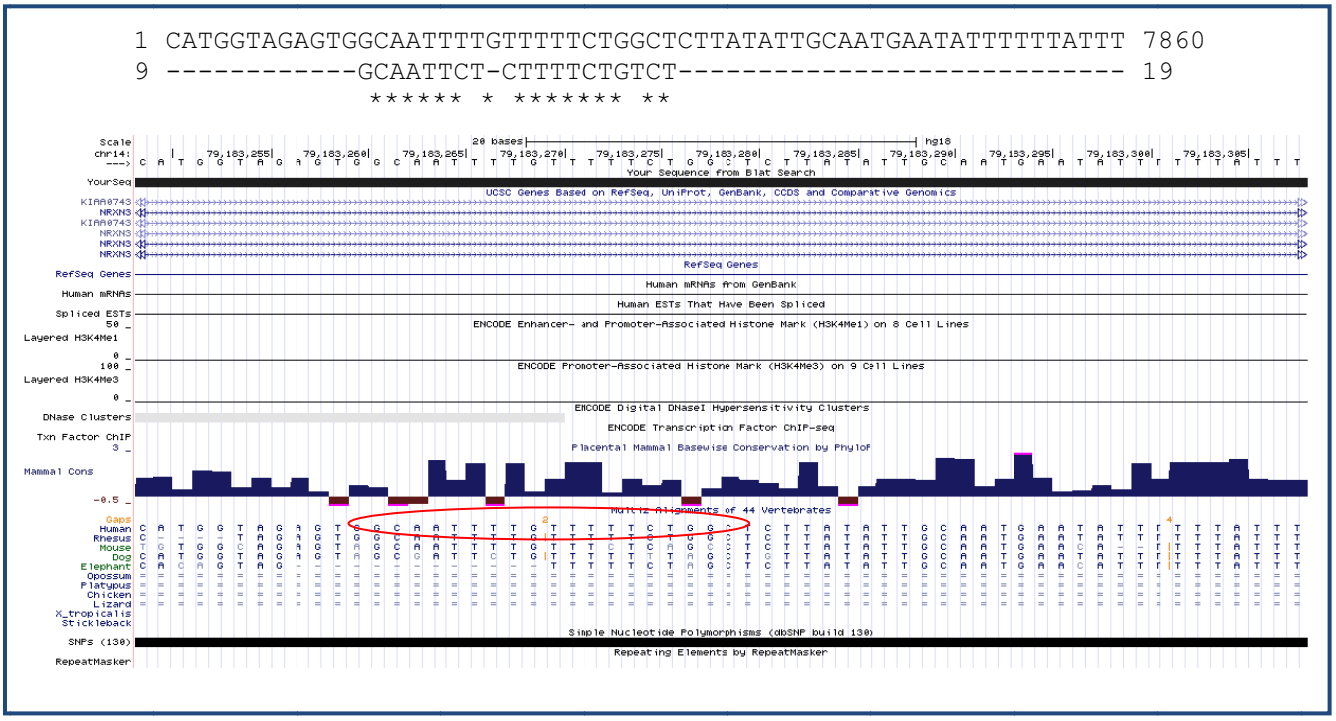


Figure 65: PTBP2 silencer binding site sequence in NRXN3_Del. Top of the figure, blat result of GCAATTCTCTTTTCTGTCT sequence in among NRXN3_del; below, this NRXN3_Del sequence (red circle) had high levels of conservation among species (image taken from UCSC genome brower Hg18).

In order to confirm whether exon20 skipping was really linked to the presence of NRXN3_DEL, we performed the same experiment in cDNA coming from HapMap cell lines with known genotypes for the deletion. We did no longer observe a relation between the amount of each of the transcripts and the genotypes for the NRXN3_DEL (Figure 66). To completely discard the possible effect of NRXN3_DEL in exon20 skipping in neural tissue (since the possible splicing inhibitors being involved in alternative splicing PTBP2 could be tissue specific) we genotyped 25 human brain samples for NRNX3_DEL and performed the same PCR in cDNA of three of the samples presenting the three possible NRXN3_DEL genotypes. This confirmed the absence of a direct correlation between NRXN3_DEL and exon 20 skipping (Figure 66).

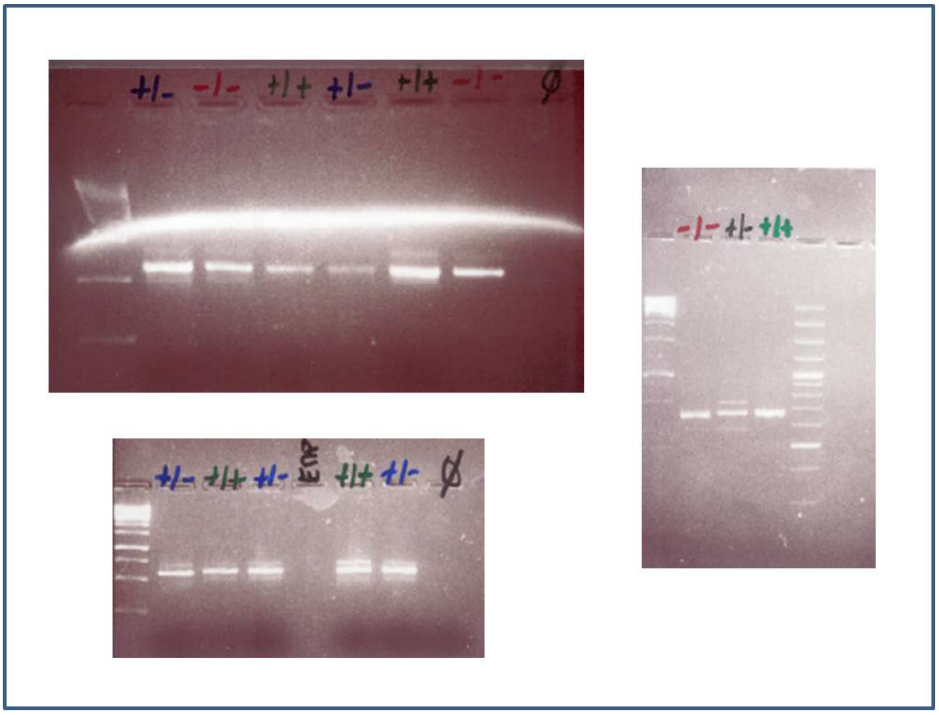


Figure 66: *NRXN3* mRNA PCR in HapMap cell lines and three human brain samples with different genotypes for NRXN3_DEL.

DISCUSSION

In this thesis we have tried to evaluate, with two whole genome approaches, the genetic contribution to FM susceptibility. This constitutes the first attempt to dissect the possible role of SNPs and CNVs in FM ethiopathogenesis in a comprehensive way. Since FM is a very heterogeneous disorder, we first attempted to identify homogeneous subgroups with a two-step clustering procedure.

Identification of fibromyalgia subgroups through cluster analysis of clinical data

We have performed a cluster analysis of clinical data and subsequent FM subgrouping in a large cohort of FM patients. These findings were replicated in a second larger cohort, thus conferring a stronger robustness to our results. To our knowledge this is the first study to perform a two-step clustering process to define variables' dimensions and subsequently identify FM clinical subgroups. The inclusion of personal and family history of comorbidities and the collection of data through direct physician examination constitute also novel contributions to FM cluster analysis.

Based on our data, the variables were grouped in three dimensions: FM symptoms and their characteristics (Dimension 1), familial and personal comorbidities (Dimension 2), and scales (Dimension 3). This clustering of FM clinical variables into different dimensions is in agreement with previous studies (179). In fact, the resulting dimensions are not completely unexpected, as some of the observed clustering can be attributed to the variables referring to the same symptom or organ (i.e. muscular symptoms in dimension 1).

A novelty of our findings is that pain symptoms were grouped into the same dimension as cognitive symptoms. Since symptoms within a cluster may share a common identifiable etiology (172), this clustering could be highlighting the CNS implication in the physiopathology of FM (6). Nevertheless, previous studies did not show the clustering of FM core symptoms and cognitive symptoms. In some cases, cognitive symptoms were not considered (175-177), or physical and psychological symptoms were considered separately (180) but in the study by Rutledge *et al.,* where both were considered, pain and cognitive symptoms did not cluster together. A possible explanation for these contradictory results is that Rutledge *et al.* evaluated in fact the patients' management of the symptoms and which ones they wanted to be improved, and not only the presence of the symptom. Another possible explanation could be that their data were collected in a different way, through online questionnaires instead of through a physician's interview.

The clustering of family history and personal comorbidities into a second dimension is also consistent with the fact that a family history of FM is linked with a more severe disease with more comorbidities (209). Personal history of chronic pain before extensive pain, however, clustered in the first dimension, although

presenting a low weight in the dimension. The fact that the history of chronic pain did not cluster with other comorbidities could be indicating that it may be considered as a supporting FM core symptom. In fact, it is difficult to identify the real onset of fibromyalgia, and whether the previous regional chronic pain belongs to the disease itself.

Finally, the cluster of scale-type variables in the third dimension could be due to the nature of the variables themselves rather than their clinical value. This might be indicating a limitation of the clustering analysis. However, we have to take into account that scales have been reduced to a binomial variable, which could have reduced their clinical value. In any case, this last dimension was the one with the fewer variables and the lowest weights, making the dimension less reliable than the other two. Since the main purpose of clinical scales is to measure severity (constituting also a screening method), they were finally used to evaluate differences among the resulting FM subgroups.

Once the clustering of the samples was performed based on their indexes in the first and second dimension only, the samples formed three groups: low symptomatology and low familiar and personal comorbidities (Cluster 1), high symptomatology and high familiar and personal comorbidities (Cluster 2) and high symptomatology but low familial and personal comorbidities (Cluster 3).

We did not observe differences in age or gender among the subsets. It has to be taken into account that due to the reduced number of males included in the study (49 individuals), we were underpowered to exclude a possible difference in the distribution of males and females among subgroups. We found no differences between groups in age, neither in age at the time of recruitment nor in the age of onset. This is in agreement with previous studies (173, 180).

Patients included in Cluster 1 showed markedly lower values for all scales. This could be pointing to a milder FM form, or to a less evolved disease. However, we evaluated time of evolution of the disease and did not observe statistical differences between the three clusters, although patients belonging to Cluster 2 seemed to have a longer disease evolution. This would indicate that the differences among the clusters are not related to disease staging, and could point to the identification of different clinical subsets. A prospective study would be useful to elucidate if the less affected Cluster 1 could have a better prognosis.

FM patients belonging to Cluster 2 (high symptomatology and high comorbidities) were also the ones with the highest levels of pain. This should not be unexpected, as most comorbidities are psychiatric, and individuals with pain have been shown to be more prone to depression, because of pain's adverse effects on mood and physical function (172). Furthermore, the number of depressive symptoms is a factor that has been associated with the development of chronic widespread pain (210), and previous cluster analyses in breast cancer showed that depression, fatigue and pain were all significantly correlated to each other (and

to total health status) (211). Comorbid medical conditions could also be responsible for a greater severity of FM symptoms in Cluster 2. However, differences between Clusters 2 and 3 in FM core measures observed in our study, although statistically significant, were limited. This could be indicative of a limited influence of personal or family history of chronic pain and psychopathology in disease severity, showing that the presence of high number of symptoms is the main marker of disease severity.

The FM subgroups that arise from our study are similar to the ones described in previous studies (175, 180, 212). Giesecke *et al.,* for instance, also identified three subsets of patients, mainly based on pain and psychopathology: 1) moderate mood ratings, moderate levels of catastrophizing and perceived control over pain and low levels of tenderness; 2) elevated mood ratings, high levels of catastrophizing and low level of perceived control over pain and high levels of tenderness; and 3) normal mood ratings, low levels of catastrophizing and high level of perceived control over pain and extreme tenderness. Their groups show similarities with our findings, despite analyzing different variables. Nevertheless, our results and Giesecke's results are not directly comparable, as they used measurements of experimental pain and some variables not included in our study. It is also difficult to compare our results with those of the study by Rehm *et al*., as they used specific pain characteristics that were not evaluated in our work. Finally, in the study by Wilson *et al*., they conducted a cluster analysis on the physical and psychological symptoms to identify subgroups. Subgroups I and IV of Wilson's classification seem to correspond to groups 2 and 1 of our classification, respectively. It would be interesting to test the degree of similarity between the resulting groups of each study. In their study, as in the present study, the group of patients with lower levels of symptomatology had higher education levels. Differences in the study design (use of web-based surveys, and exclusion of personal and family comorbidities) could explain the resulting four groups instead of three. However, both the study by Wilson *et al.* and the present study highlight the importance of symptoms in FM patient classification. This is, indeed, in agreement with the new diagnostic and classification criteria of fibromyalgia (14). Nevertheless, comparison of classifications of different studies cannot be properly performed without actually applying all the analysis to the same set of patients and directly comparing the composition of each of the resulting subgroups. In fact, it would be of interest to find a consensus regarding criteria for subgrouping of fibromyalgia patients. Performing the various types of classification in the same set of patients could help in identifying the most relevant criteria. These subgroups could then be reproduced in different patient populations and used for response analysis in future trials.

We propose a definition of FM subgroups based on two clinical dimensions resulting from a clustering analysis, although it presents several limitations. We cannot discriminate the relative weight of personal and family history of comorbidities, since they cluster together. Also, we do not know if some group of symptoms (characteristics of pain, cognitive or autonomic symptoms) were more relevant than others. Finally, the

transformation of clinical scales into binomial variables could have limited their value. It would be possible to circumvent this last issue by implementing another statistical methodology allowing their inclusion in the cluster analysis as continuous variables.

In spite of the described limitations, the resulting subgroups have been used in the replication of GWAS data and in the association study of CNV regions detected in aCGH, observing that the associated variants detected behaved differently among the different FM subgroups.

Evaluation of SNPs contribution to fibromaylgia susceptibility: a genomewide association study

We have performed a genome-wide association study considering over 500 000 SNPs in 300 FM cases and 203 controls. To our knowledge, this is the first GWAS performed in FM. We have not identified any SNP reaching GWAS association threshold, althought Manhattan plot showed possible signals on chromosomes 3 and X coming from SNPs with pvalues $<10^{-4}$. These signals were not replicated but four of the 21 most significantly associated SNPs chosen for replication showed nominal association in the replication cohort. Out of these four SNPs, three were located in intronic regions: rs12556003 (*MCF2* gene), rs9381682 (*ANK3* gene) and rs11127292 *(MYT1L* gene), and rs9381682 was intergenic.

The proto-oncogen *MCF2* is a member of a large family of GDP-GTP exchange factors that modulate the activity of small GTPases of the Rho family. It is ubiquitously expressed and is involved, among other functions, in ovogenesis, apoptotic process and dendrite proliferation in the CNS. *MCF2* variants have been associated with susceptibility to autism and schizophrenia (213).

ANK3 (ankyrin 3) belongs to a family of proteins that are believed to link the integral membrane proteins to the underlying spectrin-actin cytoskeleton. Ankyrins play key roles in activities such as cell motility, activation, proliferation, and maintenance of specialized membrane domains. Ankyrin 3 is a scaffold protein that has many essential functions in the brain, including organizational roles for subcellular domains in neurons such as the axon initial segment and nodes of Ranvier. It orchestrates the localization of key ion channels and GABAergic presynaptic terminals, and it is also involved in creating a diffusion barrier that limits transport into the axon and helps define axo-dendritic polarity. It is postulated that ANK3 similar structural and organizational roles at synaptic terminals. Variants in *ANK3* have been associated to schizophrenia, bipolar disorder (214) and autism (215).

rs11227292 showed a nominal association in the whole cohort, and , when considering only females with low levels of comorbidities (clusters 1 and 3), association got stronger both in GWAS and replication

samples, and the joint pvalue was improved. rs11127292 (chr2:2008950-2008950; Hg18) is in located in the third intron of *MYT1L* (myelin transcriptor factor 1-like). *MYT1L* encodes for a postmitotic neuronal specific zinc finger protein and belongs to the myelin transcription factor 1 (Myt1) gene family (*MYT1*, *MYT1L* and *NZF3*). It spans over 500 kb, includes 25 exons and is involved in neuronal differentiation by recruiting hystone deacetylase (216). Studies in rats have shown that it is mainly expressed during development (217), and it is still detected in the adult brain at low levels. It can induce differentiation into neurons in human embryonic and postnatal fibroblasts and in pluripotent cells, through the microRNA machinery (218) and in combination with other transcription factors (219, 220). *MYT1L* variants have been associated with neuropsychiatric disorders: rare CNVs and SNPs are involved in schizophrenia (221, 222), a duplication in autism (223), and various SNPs in Chinese Han cases of major depression (224). None of these previously associated SNPs were in LD with rs11127292 and, in fact, this variant was more strongly associated with FM cases with low levels of psychiatric comorbidities. This could be indicating that the identified association is really related to FM and not to FM psychiatric comorbidities, such as major depression. This, in addition to the role of *MYT1L* role in neuronal differentiation, makes rs11127292 a good candidate SNP for FM, which is characterized by a CNS dysfunction (87). The potential functional effect of this intronic variant was evaluated by exploring LD in the region. No functional variants in *MYT1L* were in LD with rs11127292, as it was only linked with intronic variants rs6719219 and rs11685526.

Regarding possible regulatory effects in other genes, Genvar analysis revealed a possible correlation with *SNTG2* (syntrophin gamma 2) expression in lymphocytes for both rs11127292 and rs11685526. This gene encodes a protein belonging to the syntrophin family. Syntrophins are cytoplasmic peripheral membrane proteins that bind to components of mechanosensitive sodium channels and to the carboxy-terminal domain of dystrophin and dystrophin-related proteins. In particular, the PDZ domain of SNTG2 interacts with a proteic component of a mechanosensitive sodium channel that affects channel gating. According to Genvar findings, rs11127292 AA genotype (A being the risk allele for FM) is linked to lower expression levels of *SNTG2*. *SNTG2* variants have been also associated to autism (225) and to suicide attempts in major depression (226). Thus, rs11127292 implication in FM susceptibility could also be due to changes in *SNTG2* expression levels. An interesting approach to validate this would be to evaluate whether there are different *SNTG2* CNS expression levels in FM patients compared to controls (by measuring its levels in the cerebrospinal fluid). This could also be further related to the specific genotypes of this SNP in these samples.

This potential role for the CNS in FM genetics susceptibility was not only highlighted by these three SNPs: IPA analysis of the 77 GWAS SNPs with most significant pvalues revealed neurological disease as one of the two top networks. Furthermore, Geneset GO analysis results showed ion binding (and calcium in particular) as one of the two main molecular functions, calcium channel complex as the top molecular component, and

regulation of calcium mediated signalling and calcium ion transport and neurogenesis, among the top ten biological functions. Since calcium channels are essential for neurotransmitter release, the overrepresentation of this system constitutes an additional argument supporting a role for CNS in FM.

Although all the results reported above suggest FM association with genes expressed in the CNS, we have to take these results with care since this GWAS presents several limitations.The first issue that we need to address is the reduced number of samples that we were considering. In fact, to achieve 80% power to detect low effect variants (which are the ones expected in complex disorder such as FM) it is necessary to take into account thousands of samples. Nevertheless, it has been possible to detect associations through GWAS in smaller sample sizes, as long as the cohorts of cases were very homogeneous, and, in these cases, the detected variants showed a higher effect. (227). In any case, in our study we would have been able to detect strong associations, such as the ones that have been reported in autoimmune disorders in the HLA *locus*, and we have not seen any evidence pointing to such an association. This is an important finding since many attempts have been undertaken in order to find HLA associations with FM, in particular with those alleles previously associated with rheumatic disorders (particularly, with the rheumatoid arthritis HLA-DR4 allele (112)). HLA-DR4 association to rheumatoid arthritis presented an OR ranging from 2.36 to 4.1 (143). In our study, with 300 FM samples and 203 controls, for MAF≥0.15 we had over 90% power to detect associations with OR≥2.0, as calculated with Quanto software (http://hydra.usc.edu/gxe/). Thus, we should have detected an association at the level of HLA_DR4. Nonetheless, association to the HLA locus can not be completely discarded, since HLA association for disorders not classically considered as autoimmune (such as schizophrenia), that actually have HLA risk alleles, have a much lower OR, of 1.11-1.25 (228).

Since we only had access to this sample cohort, in order to maximize our power, we decided to use a very well characterized set of patients, specifically selected for having low levels of psychiatric comorbidities and a clear FM diagnosis. This was, in fact, confirmed by the enrichment of cluster3 individuals (high symptomatology and low comorbidities) in GWAS samples (supplementary table 13). In retrospect, it might have been better to have the clustering results before starting the genetic study, in order to center the GWAS on cluster 3, but this was not possible since the collection of samples was done in parallel with the development of the project, and thus, the clustering analysis could not have been done with the initial set of samples. The reduced number of samples with GWAS data could explain why we were not able to detect an association reaching GWAS threshold. It could also result in the detection of false positives among the most associated variants, and this would lead to a lack of replication of these variants. In our case, none of the variants was strictly replicated when considering the whole cohort, as the joint analysis of the GWAS and replication cohorts did not result in an improved pvalue. Nevertheless, four of the SNPS selected for replication presented a nominal association in the replication cohort that was in the same direction that in

the GWAS. What is more, the stratified analysis of rs11127292, considering FM sample clusters, showed a replication of the GWAS results. This highlights the importance of a precise phenotyping of the samples in extremely complex disorders like FM.

An interesting finding is that association of replicated SNPs was stronger when only considering females. In fact, Manhattan plot detected a signal on X chromosome. Interpretation of this is difficult: FM is a disorder that affects women in over 90% of the cases and there is controversy as to whether pain sensibility differs among FM women and men. Small studies have been undertaken in order to evaluate this potential difference, but they show contradictory results (229, 230). Gender differences in pain sensitivity are in fact a matter of discussion. Chronic pain is more frequent in females than in males, but studies addressing different pain sensitivity among genders are not conclusive. There have been reported ,women factors such as the hormonal cycle, that may play a role in pain response and therefore complicate the comparison with men. However, some researchers have explored possible underlying mechanisms to explain sex differences in pain response and neurochemical differences have been identified (NMDA receptors, cytokine expression…). In fact, genes with a sex-dependent effect on pain have been reported. The melaconocortin-1 receptor (*MC1R*) gene, for instance, has been reported to mediate female specific mechanisms of analgesia, whereas toll like receptor 4 (*TLR4*) gene has been involved in spine pain processing only in males, as reviewed in Mogil *et al.* (231). Thus, our GWAS results would be in agreement with these reported sex differences in genes involved in pain perception.

In order to completely confirm our findings and overcome the reduced sample size, it would be necessary to replicate the detected associations in another cohort. This is challenging in FM since, to our knowledge, this cohort is the largest existing collection of fibromyalgia DNA samples. We only know of another FM cohort similar in size, and we are working with them in performing a reciprocal replication study, once an appropriate agreement is reached, since it involves several pharmaceutical companies. Other statistical methods such as common polygenic variation (228) could also help in detecting FM associated variants of low effect, although we don't know whether this statistical method could be implemented in a cohort of our sample size.

Evaluation of the contribution of copy number variants to fibromyalgia using genome-wide association study data

In addition to evaluate genetic association through SNPs, the GWAS genotyping data was used for the identification of CNVs using the PennCNV algorithm. The analsysis of CNVs through SNP arrays presents several limitations, which we have observed in our study. First of all, available algorithms used to infer CNVs from SNPs arrays need the use of more than one method to minimize the error. Furthermore, their accuracy

to predict copy number state is limited, and the resolution for detecting small variants is very low. In our case, since we evaluated a phenotype without broad clinical characteristics, such as schizophrenia, autism or mental retardation, we do not expect it to be associated to very large causative events, and it is well known that small events are missed by this technology. In particular, the platform that we have used (Illumina 1 M duo) is not particularly enriched in CNVs probes; it has a smaller number of CNVs probes compared to newer arrays such as the Illumina Omni 1. After the filtering procedure we have only followed up one of the detected CNVs, located in the *ACACA* gene. The analysis of this CNV with a direct genotyping method did not replicate the high frequency of the homozygous deleted samples observed in FM GWAS data. In fact, the presence and genotype of the CNV in the GWAS samples in which it had been detected was not confirmed by direct genotyping. Therefore, there were no differences between cases and controls when direct genotyping of the CNV was performed. This illustrates the limitations of SNP-arrays (or maybe this particular algorithm) in copy number prediction. From 10 samples that appeared as homozygous deleted from PennCNV results, only one presented this genotype by direct multiplex PCR genotyping.

Finally, we evaluated the possible presence of CNV events in mosaic state. CNV in mosaic state could constitute a good pathophysiological mechanism for FM. Changes could be due to interaction with environmental /social factors and could appear across the individual's lifetime. However, the main limitation that we had to face was that the detection of mosaicism should be ideally performed in different tissues, involving target tissues if possible: CNS and muscle and comparison across the different tissues and across time. We did not have access to this type of data, and we performed this analysis based on SNP genotype data. We finally considered interesting two of the putative mosaic aberrations, both located on chromosome 12. One was located in *SLC2A14*, a glucose carrier gene located 30 kb away from SNPs that have been associated with Alzheimer's disease in non-ApoE ε4 allele carriers (232, 233). Furthermore, glucose carrier's genes have been associated with neuropsychiatric disorders such as adult attention-deficit hyperactivity disorder (234). *SLC2A14,* a gene involved in CNS related disorders could therefore constitute a good candidate for FM, characterized by a dysfunction in pain processing. The other CNV region in mosaic state included the full coding sequence of the *OVOS2* gene. *OVOS2* (ovostatin2) encodes a protein able to inhibit all four classes of proteinases by a unique 'trapping' mechanism. It is ubiquitously expressed and it has shown to be overexpressed in ageing muscle and nucleus pulposus cells (235). Changes in *OVOS2* copy number could be implicated in FM etiopathogenesis as a complementary factor (initiating or perpetuating) to an altered pain transmission. Nevertheless, we did not perform further experiments such as MLPA or quantitative PCR in order to validate these results. Although implication of both genes in FM susceptibility seems reasonable, we did not follow them up because there were not additional evidences (SNPs in GWAS or CNVs in aCGH) supporting them as FM candidate genes, and additional tissues of the patients were not

available. This exploratory analysis should be pursued in muscle and nervous tissue in order to complete a successful validation.

Contribution of copy number variants to fibromyalgia: array comparative genomic hybridization

We have also explored the possible contribution of CNV to FM genetic susceptibility using array comparative genomic hybridization. To this purpose, we have performed aCGH with two types of arrays (Agilent 400K and 1M) using a pooling strategy. We detected seven regions that appeared to have different frequencies in cases and controls, and five of them were selected for follow up. Of these, only one was actually validated as a susceptibility factor for FM. The deletion in the *NRXN3* intronic region showed an association with susceptibility for FM in females, being stronger in cases with low levels of comorbidities.

GALNTL6 emerged as differential hybridized in the early onset pool. We initially validated this aCGH finding in a subset of 300 cases and 300 controls but association was lost when enlarging the replication cohort. We were not able to replicate this finding when only considering the early onset samples. Since the number of early onset cases was small (100 when considering those with an age of onset before 20 years old as the ones included in the aCGH pool), we also explored *GALNTL6* variant in the samples with an age of onset lower than 30 (30 representing the cohort $p_{25,}$ the median age of onset being 38) with no evidence of association either. Finally, we evaluated this variant in the different FM clusters but there was no interaction that could be explaining the initial replication that was not confirmed with further genotyping.

We also intended to replicate *WWOX*, *PTPRD* and *MYO5B* aCGH results. *WWOX* encodes for a protein with an oxidoreductase putative function and its expression is upregulated in early stages of development of the peripheral and central nervous system in mouse embryos. This protein regulates a wide variety of metabolic processes such as steroid metabolism, apoptosis, and tumour growth suppressor. It has been associated with different types of solid and blood tumors (236), it has been shown to be downregulated in a population of hippocampal neurons of individuals affected with Alzheimer's disease (237) and upregulated after nerve injury (238) .

PTPRD encodes for a member of the protein tyrosin phosphatase family, which may be implicated in neurite growth and neurons axons guidance, as suggested in studies of its orthologs in chicken and fly. It has been associated with several tumors and neuropsychiatric disorders, and has been shown to play a role in the development of inhibitory synapses (239).

Myosin Vb is expressed in several neuronal populations and associates with the alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate-type glutamate receptor subunit GluR1, mediating its transport (240). *MYO5B*

variants have been associated with bipolar disorder (241),and another member of the gene family, *MYO18B* has shown to be associated with schizophrenia (228).

We could not replicate these three variants. Since there was no evidence of association after genotyping a subset of 300 cases and 300 controls we did not continue the follow-up experiments. We cannot completely exclude the participation of these CNVs in the disease, as they could be variants with a low effect on disease which would need larger cohorts for validation. Considering all the results emerging from SNPs and CNVs analysis in FM, these could constitute good FM candidates, as they have been shown to play a role in the CNS.

In order to explore a possible additive effect of these variants, explaining a causative role not detected by association analysis, we evaluated interaction between these CNVs. We did not detect nor an additive neither an epistatic effect. Finally, the most plausible explanation is that these not replicated CNVs were in fact false positives caused by a bias in the frequencies of the CNVs in the pooled samples.

The use of pools of samples, although presenting many advantages in terms of reduction of data analyzed and money spent, can lead to the detection of false positives. For common variants, a random distribution of the different genotypes could lead to enrichment in cases or controls of a particular variant and thus to false changes in hybridization signals. For rarer variants, a slight mistake in the preparation of the pool leading to the presence of a larger amount of one or more samples carrying the rare genotype could also lead to a false hybridization difference between cases and controls. The use of pools may also lead to a decreased power to detect variants in highly polymorphic regions of the genome. On these regions, the pool will represent an intermediate between the polymorphic and non-polymorphic states, resulting in smaller relative difference in intensity than a non-polymorphic single reference would yield (165).

An option that may overcome these limitations is to increase the number of samples of the pool. Another option would be the use of various control pools, which will help in getting rid of false positives caused by random bias in the selection of samples. Furthermore, by performing several hybridizations of the same control pool against different pools and eliminating recurrent CNVs from the analysis it should be possible to reduce errors due to random bias. In our case, since we were including FM cases with family history of FM and, at the time of pools design, we only had 300 samples available, we could not increase the number of cases included in each FM pool. For controls, we used one pool but we took into account results coming from the hybridization of the same pool of controls against pools of cases from other diseases, and those variants that recurrently appeared in different hybridizations (such as *SIRPB1*) performed in our laboratory were not considered for follow-up studies. Finally, interexperimental replication involved the same

conditions and/or an experimental alternate, such as 'dye-swap' of the two fluorochrome labels between the test and the reference samples.

Finally, in retrospect and based on our GWAS and aCGH findings, we should have used only females for pools design. Although the proportion of males and females was similar in FM and the controls pools, both with a major proportion of females, it would have been more interesting to only include females. This, not only for the different behavior of the variants in each gender, but also in order to explore the X chromosome structural variants, in particular those involving X linked neuroligin (*NLGN4*) (neurexins partner).

For the detection of CNV regions showing differential hybridization between cases and controls, we used the ADM2 algorithm. This algorithm poses different advantages in comparison with other methods. For instance, it uses a variable window size, in contrast to the Z score algorithm, that uses windows of fixed size, and it takes into account the quality score of each probe, not included in ADM1. However it presents some limitations: since it calculates the median of the LRR of the detected region, the regions appearing in the results have to be carefully checked, since in some cases positive probes are not consecutive and therefore the region as a whole cannot be immediately considered as differentially hybridized. Different scenarios are then possible: a large CNV region including a negative probe that persistently fails in different hybridizations would be considered as differentially hybridized, whereas a three probe region where one of the probes does not fulfil selection criteria would be less reliable and discarded for follow-up. Another limitation, that we observed most prominently in the case of 1 M array, is that it does not take into account those probes that are repeated, and sometimes a region in which only two non-repeated probes fulfil the selection criteria appears as including a much larger number of probes.

Finally, we have been able to replicate aCGH results for NRXN3_DEL. The genotyping of the initial subset of samples (over 300 FM cases and 300 controls) showed a statistically significant association between FM and the deletion. When we completed the genotyping of the entire cohort, the association remained statistically significant in females.The stronger association in the initial replication cohort could be explained by the fact that most of the samples included in this subset were the ones that had been included in the GWAS. They were therefore samples selected by the clinicians for having a clear and confirmed FM phenotype and low levels of comorbidities. For this reason, we explored the association of the CNV in the different clusters and found that, when considering samples with low levels of comorbidities (clusters 1 and 3), the association was replicated, and it was stronger if only considering females. *NRXN3* is a good candidate for FM as it is essential for neuronal development and for signal transmission. As we have mentioned before, several variants in *NRXN3* (both rare CNVs and SNPs) have been associated with different phenotypes, mainly neuropsychiatric disorders, including addictive behavior. Changes in *NRXN3* and therefore in signal transmission could explain the central nervous pain dysfunction that is underlying the disease.

Evaluation of possible functional consequences of NRXN3_DEL

Since the detected variant was located in an intronic region, we intended to evaluate its possible functional consequences. First, we assessed if the CNV could be in LD with a functional SNP, or if it could be tagged by a SNP, facilitating its detection. This was done by a Veracode genotyping assay. The Veracode experiment included 45 SNPs capturing the main functional SNPs in the gene and SNPs previously associated with disease, but it did not detect a proxy for the deletion. It could be useful to further saturate the region surrounding NRXN3_DEL, since the Veracode experiment, designed to capture most of the entire gene variability, only included 15 SNPs in a 200 Kb window from NRXN3_DEL. None of the previously associated SNPs with other disorders showed association with FM either.

In addition to assessing its possible link to other functional variants, we also evaluated possible consequences of NRXN3_DEL at the mRNA level. We identified two neuronal cell lines that were homozygous deleted and homozygous non-deleted for the variant. Performing RT-PCR experiments to test for mRNA isoforms, we observed a different ratio of the two possible *NRNX3* isoforms when amplifying cDNA from exon18 to exon21 (with or without exon20). This was confirmed and quantified by a specific Taqman RTqPCR. Furthermore, in accordance with this hypothesis, we found PTBP2 (SS4# inhibitory factors) binding sites in the CNV region, and these corresponded to high conserved regions. The possible correlation between exon20 skipping and NRXN3_DEL would have been of great interest, since exon 20 skipping in *NRXN1* is essential for its link with neuroligins: only neurexins isoforms without exon20 bind to neuroligins (242).

We then extended this analysis to a larger number of samples, as we had access to cell lines from HapMap samples with different NRXN3_DEL genotypes, but we did not see this correlation. A possible explanation for this could have been that the behavior of the variant could be different depending on the tissue, as HapMap cell lines are derived from lymphocytes, and the cell lines in which we observed the differences were neuronal. Therefore, we then performed RTqPCR on RNA from three human brain samples carrying each of the possible genotypes for the deletion, where again we did not see the effect. Thus, the differences observed in the two glioblastoma cell lines could be due to many other factors, such as the tumor stage (T98G cell line comes from a grade III whereas U87 comes from a grade IV glioblastoma multiforme) or gender (T98G came from a male and U87 from a female).

We cannot exclude a possible regulatory effect of the CNV region at other levels. In fact, we found binding sites for other splicing regulators (supplementary figure 4) among the CNV region. It could be possible that

the regulatory effect could affect alternative splicing at other sites: at ss#5 or in non-canonical sites of alternative splicing. In order to further explore these consequences, it would be of interest to clone the deletion and transfect it into a neuronal cell model and explore expression variants (by means of an expression array or RNAseq). However, the size of the deletion (8.9 kb) makes this difficult. Another possibility would be to transfect a minigene involving a few exons before and after the deletion (ranging from *NRXN3* exon18 to exon21) but we find the same issue regarding the size of the minigene, which would be larger than 500 kb, if all intronic regions are included.

A possible solution in order to explore in a comprehensive manner the putative functional consequences of NRNX3_DEL would consist in evaluating the different changes, at the mRNA level, in different samples (preferably neuronal) carrying the different NRXN3_DEL genotypes. For this, it would be necessary to have a minimal number of samples (at least 10 per genotype) and to evaluate the transcripts with next generation sequencing techniques (RNAseq). The limitation of this would be the power to detect differences related to the CNV and not to the interindividual variability or variability due to a variable cell type distribution in the sample. In order to minimize this, it would be necessary to ensure the same tissue for the RNA extraction, and, if possible, to extend it to single cell extraction (as the different isoforms could be linked, not only to the brain area, but also to cell type).

Pain transmission and genetic susceptibility to fibromyalgia

In summary, both GWAS and aCGH results point at a role for the CNS in FM genetic susceptibility. In fact, variants detected by both studies are linked: calcium transport appears as one of the main GWAS molecular functions and the neurexins-neuroligins complex formation is dependent in calcium (24), and SNTG2 interacts with neuroligins 3 and 4, which are partners of the neurexins (225). In the 14th world congress on pain (International association for the study of pain (IASP)) August 2012, Milan; http://www.iasp-pain.org/) evidences for specific neurophysiological alterations in FM patients were reported. Of particular interest to this study, functional and morphological impairment of small fibers have been reported in FM cases (243). Individuals presenting those electrophysiological changes would be the ideal models to explore the association and the functionality of our detected genetic variants (by using, for example, tissue from peripheral nerve) in order to try to establish a correlation with clinical severity, outcome and response to treatment. In this meeting, another study evaluating gene expression of FM individuals also detected changes in genes implicated in pain transmission, which supports our findings (244).

In spite of the difficulties encountered in the study of genetic factors of FM (clinical heterogeneity, lack of previously validated genetic factors, reduced availability of replication cohorts and non availability of target tissue) we have been able to detect variants that can shed a light on genetic factors determining FM

susceptibility. To our knowledge, only neurotransmitter related genes (including receptors, transporters and enzymes implicated in neurotransmitters metabolism) had been tested as FM susceptibility candidates. The possible role of synaptic structural molecules such as ANK3 or NRXN3 and molecules implicated in CNS development and functioning, such as MYTL1, open a new wide field of research in terms of aetiology and drug targets. One consideration that we have to take into account is that all of these molecules have been previously associated with neuropsychiatric disorders. If these synapse genes associations are confirmed in other FM cohorts, FM could be considered as neuropsychiatric disorder more than a rheumatological disease.

Some issues remain unsolved though: although the CNS appears as the best candidate target tissue for FM, the muscle could also have an active role in disease development. Furthermore, it is not clear whether there is only one type of FM or the disease is actually a mixture of phenotypes, which is somehow suggested by our cluster analysis. Finally, it seems that some of the detected variants have stronger association when considering only females, which would indicate that it would be important to perform studies in males and females separately, as the genetic factors could be gender specific.

In addition to the replication in other cohorts and the evaluation of functional consequences of the detected variants, other complementary approaches could help in the understanding of the disease. Next generation sequencing technologies could be implemented for the study of FM genetics. Considering our findings, a good approach could be targeted resequencing of genes encoding for synapse molecules. Furthermore, since chronic pathological pain and in particular fibromyalgia is influenced by environmental-social factors, epigenetic changes may play an important role in FM susceptibility and development. The study of methylation changes and other regulatory marks, their evolution in time and their correlation with FM clinical course, could be of great help in order to understand the pathophysiology. And natural history of FM In this sense, the possible role of changes in the chromosome X, as suggested by initial genome-wide study results, could fit together as the inactivating methylation of one of the X copies in females could perhaps play a relevant role in FM susceptibility.

CONCLUSIONS

1. Fibromialgia is an extremely complex clinical disorder. The addition of multiple small effect genetic factors seems to be underlying FM susceptibility. It is therefore necessary to study large cohorts including thousands of samples in order to detect these risk variants.

2. The definition of clinically homogeneous FM subgroups constitutes a key step for the identification of FM genetic susceptibility factors leading to a better understanding of its biological basis. In this sense, our work indicates that the evaluation of personal and family history of FM comorbidities can add important information to FM classification based on somatic symptoms.

3. Our GWAS results point towards a possible contribution of CNS genes to FM susceptibility: pathway analysis of top associated SNPs identified neurological disease and calcium channel pathway as overrepresented.

4. A nucleotide change in the *MYT1L* gene showed statistical significant association in FM samples with high levels of symptomatology and low levels of disease comorbidities. This confirms the key role of nervous transmission in FM etiopathogenesis and highlights the importance of identifying FM homogeneous subgroups for the detection of FM genetic susceptibility factors.

5. Replication analysis showed a stronger association when considering only female cases and controls. This enhances the importance of gender in FM ethiopathogenesis and could be pointing to the existence of a different genetic background for FM in males and females.

6. The inference of CNVs in mosaic state also supported a role for the CNS, with the detection of possible mosaic events in *SLC2A14*. Mosaicism analysis also identified a CNV in mosaic state in a gene implicated in muscle degeneration, suggesting thus a role for musculoskeletal system in FM ethiopathogenesis.

7. An intronic insertion/deletion polymorphism in the *NRXN3* gene (NRXN3_DEL) showed association to FM female cases with low levels of comorbidities. Since this molecule is essential in the development, maintenance and functioning of the synapse, this result constitutes an additional argument supporting a dysfunction in neural transmission in FM.

8. A Possible functional consequence of NRXN3_DEL may be affecting alternative splicing, since binding sites for splicing regulatory factors were detected in the region. Although a possible effect at splicing site 4 leading to a switch in the expression of two different isoforms was discarded, the deletion could be affecting other sites of alternative splicing (canonical or not).

9. If the proposed FM candidate genes were further validated in replication studies, this would constitute a change in the FM etiology concept, as several of these candidates are known neuropsychiatric disease associated genes (autism, addiction and mental disability). This would highlight a novel neurocognitive involvement in this disorder, currently considered to involve the musculoskeletal and affective systems and circuits.

RESUMEN

El dolor es un mecanismo fisiológico de defensa ante agentes externos. Existen diferentes tipos de dolor: el dolor agudo, el dolor inflamatorio y el dolor patológico.Tanto el dolor agudo como el inflamatorio tienen una función protectora: el primero en condiciones basales y el segundo para proteger regiones lesionadas. Sin embargo, el dolor patológico es una alteración en la transmisión del dolor. El desarrollo técnológico de las técnicas de imagen, ha permitido el estudio en profundidad del dolor patológico. Principalmente el uso de la resonancia magnética funcional junto a otras técnicas de medicina nuclear, han permitido visualizar el dolor. Alteraciones tanto a nivel del sistema nervioso periférico como del sistema nervioso central se han postulado como causantes de este tipo de disfunción en la transmisión del dolor. Una de las patologías más frecuentes caracterizadas por dolor patológico es la fibromialgia.

La fibromialgia (FM) es una enfermedad de etiología desconocida que se caracteriza por dolor crónico generalizado, junto a una amplia constelación de síntomas acompañantes. En un porcentaje muy elevado de casos, la fatiga crónica (entidad en si misma) es, junto al dolor, uno de los síntomas predominantes. Los pacientes pueden presentar, además, alteraciones del sueño, depresión, ansiedad, rigidez articular, parestesias en extremidades, cefalea, así como una hipersensibilidad al dolor, que se manifiesta por la aparición de sensación dolorosa a la presión en múltiples lugares de inserción osteotendinosa. Según el estudio EPISER, la fibromialgia definida según los criterios del *American College of Rheumatology*, es una patología frecuente en España, con una prevalencia del 2,4% de la población mayor de 20 años, lo que supone unos 700.000 pacientes de FM en el estado español. Afecta de forma predominante a mujeres, con una relación mujer: hombre de 21:1.

La FM es una enfermedad altamente incapacitante. Su impacto afecta tanto al paciente a nivel familiar y profesional como a su entorno laboral y personal. Además conlleva un gran impacto económico, en términos de gasto sanitario y de bajas laborales. Teniendo en cuenta la elevada prevalencia de la patología, constituye actualmente un problema sanitario importante. Por este motivo, se han desarrollado unidades multiciplinarias para el seguimiento y el tratamiento específico de los pacientes con el principal objetivo de mejorar la calidad asistencial de dichos pacientes.

Dado que no existen pruebas objetivas que permitan evaluar la FM, el diagnóstico se basa en una historia de síntomas y la exclusión de enfermedades somáticas que expliquen dichos síntomas. En 1990, el American College of Rheumatology, estableció unos criterios diagnósticos que definían la FM como la presencia de dolor generalizado de más de tres meses de evolución. Dicho dolor debía ser bilateral, afectar al tronco superior e inferior e incuir dolor axial junto a la presencia de dolor a la presión en 11 de 18 puntos específicos del cuerpo, denominados puntos dolorosos.

Posteriormente, han ido apareciendo objeciones a estos criterios diagnósticos. La baja tasa de exploración de los puntos dolorosos en la práctica clínica habitual, ha puesto en evidencia que el diagnóstico de FM acababa basándose más en la presencia de síntomas que en la exploración física. Por ello en 2010 sedesarrollaron unos nuevos criterios diagnósticos que incluían un mayor reconocimiento de los problemas cognitivos y los síntomas somáticos. Esta nueva definición de FM se basa en la construción de un índice el *Widespread Pain Index (WPI)* constituido a partir de la presencia o ausencia de diferentes síntomas.

Dentro del diagnóstico diferencial de FM entran todas aquellas patologías que puedan causar dolor generalizado: desde enfermedades tumorales a infección virales, pasando por enfermedades inflamatorias o autoinmunes. Para descartar la mayoría de estas patologías, una analítica general es normalmente suficiente. Además no hay que olvidar la existencia de rentistas o simuladores que buscan en el diagnóstico de FM el beneficio de FM en términos de bajas laborales o pensiones por invalidez.

Además del uso de criterios diagnósticos para el seguimiento de la enfermedad es muy común el uso de escalas de actividad. Dichas escalas son tanto específicas de FM (FIQ) como usadas en múltiples patologías para la monitorización de la calidad de vida del paciente (SF-36) o la evaluación de sintomatología acompañante.

Dado que se trata de una entidad de etiología desconocida, no existe, hasta el momento un tratamiento curativo. Únicamente existen tratamientos sintomáticos de limitado efecto terapéutico. Estos tratamientos combinan la terapia farmacológica con la física y su indicación no está claramente establecida. Los fármacos utilizados en FM son básicamente analgésicos así como antiepilépticos utilizados para el tratamiento del dolor neurpático. Finalmente, dado que los enfermos de FM tienen una mayor tasa de efectos secundarios, el establecer una pauta terapéutica adecuada representa un reto para los facultativos y pasa por realizar un tratamiento personalizado, prácticamente a medida.

La ausencia de signos objetivos que permitan la confirmación del diagnóstico así como el seguimiento de la enfermedad hacen que sea un tema controvertido. Incluso algunos facultativos ponen en duda la existencia de dicha patología. Esta incomprehensión, tanto por parte del colectivo médico como de la sociedad dificultan el diagnóstico y el tratamiento de la enfermedad con el consecuente sufrimiento por parte del paciente. Pero además conlleva más inconvenientes. En muchos casos es considerada como un cajón desastre donde van a parar, de forma incorrecta, otras entidades mal diagnosticadas.

La base etiopatogénica que explica este estado permanente de dolor es aún desconocida. Se han postulado diferentes hipótesis (inmunológica, neuro-vegetativa, etiología vírica…) que expliquen las causa de la FM. Hasta la fecha la teoría más plausible es la existencia de una disfunción en la transmisión del dolor con un desequilibrio entre las vías inhibitorias y activadoras del impulso doloroso. Alteraciones en los niveles de

diferentes neurotransmisores, así como en pruebas de resonancia magnética funcional o neurofisiología apoyan dicha hipótesis. La hipótesis más plausible es que la fibromialgia sea un proceso complejo causado por la confluencia de factores genéticos y ambientales.

Los factores genéticos pueden explicar en gran medida la variabilidad en la percepción del dolor. Los estudios familiares han mostrado una considerable agregación familiar en fibromialgia y cuadros relacionados, sugiriendo la importancia de los factores genéticos en el desarrollo de estos cuadros La FM tiene un componente genético como muestra la existencia de agregación familiar. Diferentes estudios han evaluado la prevalencia de FM en familiares de afectos estimando una mayor recurrencia que en la población geneeral. Dicha mayor prevalencia ha demostrado, además, ser mayor en familiares de sangre que en esposos de pacientes de FM, lo que va a favor de una mayor importancia de los factores genéticos que los factores ambientales en determinar la susceptibilidad a desarrollar la enfermedad. Numerosos estudios en gemelos se han desarrollado en este sentido y, aunque en algunos casos los resultados sean contradictorios, en general muestran una mayor concordancia en gemelos monozigóticos que dizigóticos. Es , pues, un argumento más a favor de la existencia de factores de susceptibilidad genética para FM.

Con el fin de evaluar dicho componente genetico, se han llevado a cabo numerosos estudios. Los estudios realizados hasta la fecha han sido básicamente estudios de genes candidato y no han sido capaces de establecer asociaciones genéticas. Esto se debe a que presentaban numerosas limitaciones: la mayoría fueron realizados en cohortes pequeñas, los marcadores evluados eran en la mayoría de los casos facotres de susceptibilidad genética para comorbilidades psiquiátricas de la enfermedad, no presentabans asociaciones estadísticamente significativas tras corregir por *multiple testing* e intentos de replicar las asociaciones encontradas mostraban resultados contradictorios. En particular, polimorfismos de base única de genes implicados en las vías serotoninérgicas, dopaminérgicas o catecolaminérgicas, pero hasta la fecha ningún estudio ha sido capaz de establecer un marcador genético para FM.

Es por todo ello que la identificación de las bases biológicas de la fibromialgia precisa de un diseño de investigación que integre aproximaciones epidemiológicas y genómicas en un gran número de muestras bien caracterizadas clínicamente.

La búsqueda de los factores de susceptibilidad genética a las enfermedades complejas se centra en la identificación de variantes del genoma que se encuentren con mayor frecuencia entre la población de enfermos en comparación con la población general, con la esperanza de que ellas mismas sean responsables de la enfermedad o bien que sean marcadores (debido al desequilibrio de ligamiento) para los verdaderos genes responsables. Los SNPs (polimorfismos de nucleótido único) representan la mayor fuente de variabilidad del genoma humano y constituyen una herramienta fundamental para realizar estudios de

asociación que permitan detectar los genes implicados en las enfermedades comunes que afectan a la población.

Tanto en Estados Unidos como en el Reino Unido existen grandes iniciativas financiadas por el National Institute of Health y el Welcome Trust, respectivamente, en el seno de las cuales se están genotipando más de 500.000 SNPs en millares de muestras de pacientes de distintas enfermedades. Los primeros resultados del Welcome Trust Case Control Consortium (WTCCC) publicados en los últimos años, presentaban datos para 14.000 pacientes afectos de siete enfermedades distintas (trastorno bipolar, diabetes mellitus tipo 1, diabetes mellitus tipo 2, artritis reumatoide, enfermedad de Crohn, hipertensión y enfermedad cardiovascular)1, además de otros estudios en cáncer, asma y otras enfermedades inflamatorias. Del mismo modo, y haciendo uso de las últimas tecnologías de secuenciación (*next generation sequencing*), el consorcio 1000*genomes* está produciendo datos de exoma y *whole genome sequencing* que vienen a completar el mapa de variablilidad del genoma humano.Estos trabajos demuestran de forma clara que es posible identificar factores genéticos implicados en enfermedades comunes, abriendo nuevas vías biológicas implicadas en estos procesos.

La reciente identificación de variantes de número de copia (CNVs) en el genoma humano ha abierto un nuevo campo de estudio en la búsqueda de bases genéticas de las enfermedades. En algunas patologías complejas ya se ha identificado una relación entre la susceptibilidad a padecer la enfermedad y variantes de número de copia de ciertos genes, como por ejemplo en lupus eritematoso sistémico (genes *FCGR3B32* y *C4A*), psoriasis (*LCE3C* y *LCE3B*) o susceptibilidad a la infección por HIV (gen CCL3L1)4,5. Por ello es importante tener en cuenta este tipo de variabilidad al investigar las causas genéticas de otras enfermedades complejas.

Tanto para evluar SNPs como CNVs existen diferentes tecnologías. En el caso de los SNPs, además de la metodología implementada, el número de marcardores que se quiere evaluar determina el tipo de tecnología a usar: desde el genotipado de un único SNP mediante Taqman hasta la realización de un estudio de genoma completo que, en la actualidad puede evaluar más de 2.5 millones de SNPs. En los úlitmos años, además de las tecnologías de genotipado, el desarrollo de la secuenciación de última generación ha permitido su uso más extendido. La progresiva rapidez así como el abaratamiento de costes, hace que problablemente, en los próximos años, la secuenciación pase a sustituir, en la mayor parte de casos, a la genotipación.

Para realizar un barrido a nivel de todo el genoma en variantes en el número de copia , una de las tecnologías más utilizada es la hibridación genómica comparada basada en arrays, en la que la muestra problema y la muestra control son hibridadas de forma competitiva en el mismo array. En la implementación de dicha tecnología se pueden usar muestras únicas o pooles de muestras. El uso de pooles tiene como

objetivo el diluir la variabilidad interindividual y resaltar las diferencias debidas a enfermedad. Asimismo, conlleva una reducción en el número de datos generados, así como en los costes.Tras la detección de regiones presentando hibridaciñon diferencial entre casos y controles, dichas regiones deben ser validadas mediante diferentes técnicas como PCR , FISH o PCR cuantitativa en tiempo real.

La realización de estudios de genoma completo, tanto a nivel de CNVs como de SNPs necesita del uso de una gran cantidad de muestras (del orden de millares) para tener el suficiente poder estadístico que permita detectar variantes de bajo efecto. En el caso de las enfermedades complejas se han postulado diferentes hipótesis que expliquen el componente genético. Se cree que la suma e interacción de un gran número de variables de bajo riesgo serían las responsables de determinar susceptibilidad a desarrollar dicho tipo de patologías. Para alcanzar un tamaño muestral adecuado en muchos casos es necesario la integración de diferentes cohortes. Esto supone una limitación añadida: la heterogeniedad interpoblacional que puede añadir sesgos en el análisis y la disminución en la homogeneidad del fenotipo y en la exactitud en la valoración clínica. Todos estos factores deberán tenerse en cuenta a la hora de realizar un análisis de factores de susceptibilidad genética para una enfermedad com la fibromialgia

Objetivos de la tesis

Dado que la fibromialgia es una enfermedad de origen desconocido, compleja, frecuente y altamente incapacitante, con la presente tesis se ha pretendido estudiar e identificar variantes del genoma (polimorifsmo de base única *single nucleotide polymorphisms* –SNPs- y variantes en el número de copia – CNVs) asociadas a FM, con el objetivo de profundizar en la etiología de la enfermedad. Para ello, se han llevado a cabo tres grandes aproximaciones:

1. Identificación de subgrupos clínicos homogéneos de fibromialgia mediante un análisis de clusters :
    a) Construcción de dimensiones de variables clínicas para la posterior
    b) Identificación de subfenotipos de FM
2. Realización de un estudio de genoma completo (genome wide association study (GWAS)) para el análisis directo de SNPs así como la detección de regiones variables en número de copia asociadas a enfermedad, y su potencial presencia en mosaicismo. Este objetivo incluye los siguientes subobjetivos:

    e) Estudio de asociación de los datos del GWAS,
    f) Identificación de CNVs,

g) Replicacion de los SNPs con un mayor nivel de asociación
h) Aplicación del análisis de cluster en los resultados del GWAS

3. Realización de experimentos de hibridación genómica comparada mediante arrays (aCGH) con el fin de identificar regiones variables en el número de copia asociadas a fibromialgia:

d) Análsis y validación de experimentos de aCGH
e) Aplicación del análisis de cluster en los resultados del aCGH
f) Evaluación de las posibles consecuencias funcionales de las variatnes en número de copia asociadas

Para llevar a cabo dichos estudios, hemos dispuesto de una gran cohorte de casos de fibromialgia, muy bien caracterizados clínicamente.

Se incluyeron en el studio 1510 casos independientes de FM que cumplían los criterios diagnósticos para FM del *American College of Rheumatology* de 1990. Estos casos fueron recogidos en un estudio multicéntrico cuyo principal objetivo fue la creación del banco español de datos clínicos y ADN de fibromialgia y síndrome de fatiga crónica. Cinco unidades de FM de cinco hospitales españoles (Hospital del Mar, Barcelona, Jordi Carbonell; Hospital Clínic i Provincial, Barcelona, Antonio Collado; Hospital de la Vall d'Hebrón, Barcelona, Jose Alegre; Instituto General de Rehabilitación de Madrid, Madrid, Javier Rivera; and Hospital General de Guadalajara, Guadalajara, Javier Vidal) participaron en la recogida de muestras. Un total de 1,510 pacientes de FM fueron seleccionados por reumatólogos para ser posteriormente evluados por otro grupo de médicos entrenados para la evaluación de este tipo de pacientes.

La recogida de datos se realizó mediante un protocolo estánadard de cuestionarios y exploración física que incluía: datos sociodemográficos, historia personal y familiar de enfermedad, tiempo de evolución de la enfermedad, presencia de síntomas del amplio espectro de afectación de la FM y tratamientos. Las medidas centrales de actividad de la enfermedad fueron evaluadas mediante escalas validadas al español. 1,000 control samples came from the National DNA Bank of Salamanca).

Asimismo se incluyeron 1000 muestras de individuos españoles con bajos niveles tanto de dolor como de fatiga según las respuestas a un cuestionario.

Identificación de subgrupos de fibromialgia mediante un análisis de clusters

Dada la heterogeneidad de la patología, se ha realizado, en primer lugar, un análisis de clusters. En un primer tiempo, se han construido dimensiones de variables (en base a su similitud) que en un segundo tiempo han servido para identificar subgrupos de pacientes. El análisis se ha llevado a cabo, inicialmente, en un subset de 559 casos de FM para ser reproducido posteriormente en un segundo grupo de 887 casos.

El estudio de 48 variables clínicas (sintomatología, antecedentes personales de psicopatología y patología osteoarticular, antecedentes familiares y escalas de medición de severidad de la enfermedad) en 1446 individuos no relacionados diagnosticados de FM, cumpliendo los criterios ACR de 1990 y seleccionados en el marco de la creación del banco de ADN de fibromialgia, ha dado lugar a la construcción de tres grandes dimensiones de variables: sintomatología, antecedentes (personales y familiares) y escalas de medición de la actividad de la enfermedad. Únicamente las dos primeras, por ser más robustas al incluir un mayor número de variables y de mayor peso, han sido en un segundo tiempo utilizadas para la identificación de subgrupos de FM. Tres grupos de pacientes han sido identificados: FM con baja sintomatología y niveles bajos de comorbilidad, FM con elevada sintomatología y elevados niveles de comorbilidad y FM elevada sintomatología pero niveles bajos de comorbilidad. Las escalas clínicas que no han sido utilizadas para la construcción de dichos subgrupos han permitido evaluar que los subgrupos detectados eran clínicamente diferentes. En particular, los niveles de actividad de la enfermedad han sido diferentes de forma estadísticamente significativa en el grupo de FM con bajos niveles tanto de sintomatología como de comorbilidades.

Estudio de asociación de genoma completo

Para el análisis de los SNPs, se ha realizado un estudio de genoma completo (GWAS) mediante un array con más de 1 millón de marcadores, el Illumina 1M duo. Se han hibridado un total de 321 muestras de fibromialgia. Para realizar el estudio de asociación, se han utilizado los datos resultantes de la hibridación de 200 muestras control con el array 610 Quad de Illumina. Dado que casos y controles se han genotipado con plataformas diferentes, se han realizado controles de calidad (*Quality Control* -QC-) por separado en cada cohorte, a nivel de muestras y de SNPs, con el software PLINK. Para las muestras se ha comprobado: el origen, el sexo, el parentesco, la heterogeneidad y el *genotyping calling*. A nivel de los SNPs se ha evaluado el *genotyping calling*, el equilibrio de Hardy Weinberg y la *minimum allele frequency*. Una vez filtrados los

datos de casos y controles, se ha realizado el estudio de asociación, mediante el software PLINK, de 505 454 SNPs en 300 casos y 203 controles.

No se han detectado diferencias poblacionales significativas entre casos y controles, como muestra el valor del *genomic inflation* (λ=1.013). La asociación alélica no ha identificado SNPs con niveles de asociación inferiores al umbral establecido para un estudio de genoma completo, aunque 77 SNPs presentaron asociación con pvalores inferiores a 10-4.

Se ha procedido entonces a la evaluación de dichos SNPs mediante los programas Ingenuity Systems Pathway analysis (IPA) y GeneSet analysis Toolkit v2. Se han identificado como principales pathways enfermedades reproductivas y enfermedades neurológicas. El transporte del calcio también ha aparecido como un proceso biológico destacado.

Con el fin de realizar *fine-mapping*, es decir, obtener una visión de alta resolución de una región asociada para aumentar la posibilidad de que un SNP causal sea identificado directamente, se ha realizado imputación con el software Impute-v2, usando como referencia tanto el panel de CEU de Hapmap3 como datos de individuos CEU del 1000 genomes Project. Para implementar dicho software ha sido necesario transformar los datos al formato PLINK mediante el software GTOOL. Más de 800000 de SNPs han sido imputados de los cuales más de cuatro millones han sido considerados por su alta calidad en la imputación y por presentar una mínimum allele frequency superior a 0.05. Finalmente, se ha procedido al estudio de asociación de los datos imputados mediante el software SNPTEST, usando el método de asociación *frequentist* que tiene en cuenta la incerteza de los genotipos.

De los 77 SNPs mostrando mayores niveles de asociación junto a los datos de imputación de las regiones cercanas a dichos SNPs, han sido seleccionados 21 SNPs para su genotipación en el resto de la cohorte. Se han detectado diferentes SNPs asociados. Tras el QC, 20 SNPs en 968 FM y 937 controles han sido considerados para estudio de asociación. La asociación detectada en el GWAS no ha sido replicada de forma estadísticamente significativa ($p<0.0023$ tras corrección de Bonferroni). 4 SNPs han presentado asociación nominal o tendencia a la asociación que ha dado lugar una asociación del orden de 10^{-4} al considerar las muestras genotipadas en el GWAS y en la réplica de forma conjunta: rs12556003 (*MCF2*), rs9381682 (*ANK3*) and rs11127292 *(MYT1L*) y rs9381682 (intergénico).

Dado que los cuatro SNPs con asociación nominal en el análsisi conjunto, se encontraban en regiones no codificantes, se han evaluado las posibles consecuencias funcionales de dichos SNPs. Mediante el uso de diferentes softwares, se han detectado elevados niveles de conservación para dos de estos SNPs. Asimismo, el análisis mediante el software Genvar, ha identificado una posible correlación de los genotipos de rs11127292 y rs11685526 (SNP qu está en desequilibrio de ligamiento con rs11127292) con la expresión en

linfocitos del gen *SNTG2.* Estudios previos, han mostrado variantes de dicho gen abasociadas con autismo e intentos de autolisis en depresión mayor.

Posteriormente, se ha llevado a cabo la evaluación del comportamiento de estos 4 polimorfismos en los tres subgrupos de FM detectados mediante el análisis de clusters. Únicamente rs11127292 ha mostrado un comportamiento diferente, de forma que, al considerar únicamente el cluster 3 de pacientes (elevada sintomatología con bajos niveles de comorbilidad) la asociación detectada en el GWAS se replica en el resto de la cohorte, con una mayor significación estadística de la asociación en el estudio conjunto (asociación alélica $p=4.28X10^{-5}$, OR 0.58(0.44-0.75).

El polimorfismo rs11127292 es una variante intrónica del gen *MYT1L*, que codifica una proteína zinc finger postmitótica implicada en la diferenciación neuronal. Variantes en dicho gen han sido asociadas a diferentes trastornos neuropsiquiátricos. Se ha evaluado, asimismo, la posible existencia de variantes funcionales en desequilibrio de ligamiento con rs11127292, identificándose únicamente como ligadas las variantes intrónicas rs6719219 y rs11685526.

También se ha procedido a la detección de CNVs mediantes los datos generados por el array de SNPs, en concreto mediante el análisis de la bi allele frequency (BAF) y el logaritmo de la ratio entre intensidades de cada uno de los alelos posibles para cada SNP (log ratio). Para ello se ha utilizado el software PennCNV, detectándose más de 18000 SNPs. Se ha procedido entonces a la selección de los CNVs para validar teniendo en cuenta los siguientes criterios: CNVs presentes en más del 5% de las muestras, localizados en genes y no siendo elementos previamente relacionados con intervariabilidad poblacional. Dos CNVs han cumplido dichos criterios.

Se ha procedido al seguimiento de uno de ellos, localizado en el gen *ACACA*. Tras la identificación de los puntos de rotura mediante experimentos de PCR y posterior PCR secuenciación , previo subclonaje del producto de PCR, se ha diseñado una PCR multiplex. Con dicho diseño se han genotipado 209 FM y 211 controles sin identificar una asociación estadísticamente significativa que validase los hallazgos del análisis mediante PennCNV.

También se ha procedido al filtrado de las regiones detectadas por PennCNV con el objetivo de identificar CNVs grandes y poco frecuentes asociados a FM. No se ha detectado ninguna región no presente en población general que apareciera de forma recurrente en las muestras de FM.

Del mismo modo, se ha evaluado la existencia de CNV en mosaicismo que pudieran estar asociados a FM. Mediante el software Mosaic Alteration Detection-MAD se han identificado dos CNVs no presentes en población control. Una ganancia en mosaicismo en el gen *OVOS2* y otra en el gen *SLC2A14.* No se ha

realizado la validación y la replicación de estos CNV en posible mosaicismo porque no se disponía de otros tejidos para poder evaluar cambios en el número de copias entre los diferentes tejidos de un mismo individuo.

Estudio de la implicación de variantes en el número de copia mediante aCGH

Para el estudio de los variantes en número de copia (CNV), se han llevado a cabo experimentos de hibridación genómica comparada por array(aCGH), mediante el array 400K de Agilent, hibridando tres pooles de muestras de pacientes afectos de fibromialgia (FM) con un pool de controles. Los pooles de pacientes incluían casos de FM con antecedentes familiares para la enfermedad y las siguientes características: pool FM (individuos sin fatiga), pool FM con fatiga y pool FM de inicio precoz (edad de inicio de la enfermedad inferior o igual a 20 años).Mediante el análisis informático de los resultados de dichas hibridaciones, se han identificado diferentes regiones mostrando diferencias de intensidad entre casos y controles. Para dicho análisis se ha utilizado el algoritmo ADM2 del software genomic Workbench de Agilent, estableciendo como criterios de selección la presencia de al menos tres sondas con un logRatio superior a 0.3 o inferior a -0.3 tanto en las hibridaciones directas como en las hibridaciones *dye-swap*.

Se han realizado también hibridaciones con el array 1 Million de Agilent, de forma complementaria. Esta vez únicamente se llevaron a cabo hibridaciones directas (no hubo dye swap).

En las hibridaciones del array 400K, se han detectado 7 regiones que mostraban hibridación diferencial entre casos y controles. Dos de ellas han sido descartadas por aparecer de forma sistemática en todas las hibridaciones realizadas en otros estudios que se llevaban a cabo en el laboratorio. Además se ha incluido en el seguimiento una región ,de forma complementaria,detectada por el algortimo ADM1 (menos astringente), en el gen *MYO5B*, por constituir un buen candidato para fibromialgia. En los resultados de las hibridaciones con el array 1 Million, no se ha seleccionado ninguna región para el seguimiento por ser la regiones poco consistentes (presencia de sondas repetidas, ganancias o pérdidas según el pool de FM…).

Así pues se ha realizado el seguimiento de las regiones localizadas en los genes *GALTNL6*, *WWOX*,*MYO5B*,*PTPRD* y *NRXN3*.Los puntos de rotura de cinco de las seis regiones seleccionadas para validación,estaban disponibles en las base datos públicas por lo que se han podido diseñar directamente experimentos de genotipación mediante PCR multiplex. Para la validación del CNV localizado en una región

intrónica de *WWOX,* ha sido necesario identificar los puntos de rotura mediante una PCR y posterior PCR de secuenciación.

Para la validación y replicación de las regiones detectadas mediante aCGH,se han realizado experimentos de PCR multiplex. Se han genotipado una parte de la cohorte (300 casos y 300 controles) y en función de los resultados se ha porseguido con el genotipado de más muestras o se ha descartado la región. Finalmente, en cinco de ellas no se ha encontrado asociación con FM, mientras que en la otra, NRXN3_DEL, se ha confirmado el resultado del experimento de aCGH.

NRXN3_DEL es una indel de 8.8 kb situada en el segundo intrón del gen neurexina 3 (*NRXN3*). *NRXN3* codifica a un proteína transmembrana sitúada en las neuronas presinápticas siendo esencial para el desarrollo y funcionamiento de la sinapsis. Las neurexinas se unen a las neuroliginas (proteínas trans membrana potsinápticas) y el complejo neurexina-neuoligina puede promover la fomración *de novo* de sinapsis y la diferenciación de receptores postsinápticos Dicha unión está estrechamente regulada por mecanismos de *alternative splicing* en ambas familias de moléculas. Tanto SNPs como CNVs en los diferentes genes neurexina (en seres humanos esxiten tres) y neuroligina se han asociado a trastornos neuropsiquiátricos y a adicción a diferentes substancias.

El genotipado de 359 casos de FM y 378 controles ha mostrado una asociación estadísticamente significativa de FM con la deleción 359 (asociación genotípica, modelo recesivo (genotipo riesgo: homozigoto delecionado) p=0.0037, OR (95%CI) = 1.74 (1.19-2.54); asociación alélica p=0.12 Test de convirtiéndose en una tendencia (asociación genotípica, modelo dominante, (genotipos de riesgo: homozigoto delecionado) p=0.064 OR 1.18 (0.99-1.40, asociación alélica, Test de Fisher p=0.07, OR (95%CI) = 1.12 (0.98-1.27)). Al considerar únicamente mujeres tanto en casos como en controles, la asociación sí que ha sido estadísticamente significativa (asociación genotípica, modelo recesivo (genotipo riesgo: homozigoto delecionado) p=0.021,OR 1.46(1.05-2.04), asociación alélica Test de Fisher p=0.015 OR (95%CI) = 1.22(1.03-1.43))

Asimismo, al aplicar el análisis de clusters, se ha detectado una asociación estadísticamente significativa al considerar los clusters 1 y 3 (FM con bajos niveles de comorbilidad) (asociación genotípica, modelo dominante p=0.009, OR (95%CI)=1.28(1.06-1.54);asociación alélica test de Fisher p=0.019, OR(95%CI)=1.17(1.021-1.34)).

Se ha evaluado la posible existencia de interacción entre NRXN3_DEL y los otros CNVs evaluados, así como los 4 SNPs mostrando una asociación nominal en el GWAS, sin evidenciarse interacción alguna.

Al tratarse de un CNV situado en una región intrónica, para evaluar las posibles consecuencias funcionales de dicha variante, se ha diseñado un experimento Veracode con el fin de analizar 48 polimorfismos de base única (SNPs) en la región y así poder establecer el desequilibrio de ligamiento existente. Se han seleccionado un total de 45 tagSNPs, SNPs funcionales y SNPs situados en sitios de splicing alternativo, a los que se han añadido 3 SNPs especialmente diseñados para genotipar el CNV. Ningún SNP ha mostrado asociación con FM ni ha demostrado estar en desequilibrio de ligamiento con el CNV como para poder ser su *proxy.* Dicho ensayo sí que ha demostrado su utilidad para genotipar el CNV ya que la concordancia entre la PCR multiplex y el Veracode ha sido del 100% con lo se ha podido genotipar el total de la cohorte de una forma más rápida y precisa, mediante el análisis de los SNPs.

Por otra parte, se han realizado experimentos con líneas celulares con el fin de evaluar la posible correlación entre el genotipo para NRXN3_DEL y los tránscritos resultantes. Se han seleccionado las líneas de glioblastoma U87 y T98G por provenir de tejido neuronal y presentar diferentes genotipos para NRXN3_DEL (U87 homozigota delecionada y T98G homozigota no delecionada) y se ha procedido a experimentos de PCR amplificando cDNA (provineinte de la extracción de ARN de ambas líneas celulares) de los exones 19 a 21. Han amplificado las dos isoformas posibles a dicho nivel (con y sin el exon20) y presentando una diferente proporción de cada una en cada línea celular. Dichas diferencias de expresión han sido identificadas mediante PCR cuantitativa a tiempo real (ensayos Taqman) confirmando una mayor proporción de la isoforma sin el exón 20 en la línea celular U87.

Al constatar en la bibliografía que existe una posible relación entre el splicing alternativo en el sitio canónico 4 y la presencia de secuencias de unión para inhibidores del splicing, hemos evaluado *in silico* la posible presencia de dichos motivos en la región NRXN3_DEL, confirmando la presencia de sitios de unión de splicing. Se han detectado secuencias para la unión de dichos factores en la región NRXN3_DEL. Además muchas de estas secuencias presentaban una elevada tasa de conservación entre especies.

Finalmente, para confirmar la relación entre el CNV y cambios en el splicing alternativo, evaluamos los tránscritos en líneas de HapMap que presentaban difenretes genotipos para la deleción. No encontramos relación y para descartar que el efecto no fuera dependiente de tejido, también hemos evaluado muestras de cerebro humano (vermix) portadoras de diferentes genotipos de NRNX3_DEL, descartando una relación directa.

En la presente tesis, se ha llevado a cabo un estudio exhaustivo de los posibles factores genéticos implicados en la susceptibilidad a desarrollar fibromialgia. Dada la heterogeneidad de dicha patología, en un primer tiempo hemos realizado un análisis de clusters de variables clínicas que nos ha permitido identificar tres subgrupos de FM.

El análisis de clusters de variables clínicas de FM ha resultado en la identificación de tres subgrupos de FM: FM con bajos niveles de sintomatología y de comorbilidades, FM con elevados niveles tanto de sintomatología como de comorbilidades y FM con elevados niveles de sintomatología pero bajos niveles de comorbilidades. Los subgrupos identificados han resultado ser ,además, diferentes a nivel de direntes escalas de medición de actividad clínica. El análisis de cluster, además de diferenciar grupos clínicamente homogéneos de FM, ha mostrado la importancia de los antecedentes personales y familiares, para la correcta clasificación de los pacientes en cada uno de los grupos. Este hecho apoya la nueva clasificación de FM en la que se subraya la importancia de la sintomatología a la hora de evaluar la enfermedad. Finalmente, dicho estudio ha sido aplicado al análisis de variantes genéticas.

Se ha realizado un estudio de genoma completo para el análisis de la posible contribución de SNPs en susceptiblidad a desarrollar FM. A pesar de que el estudio inicial no ha detectado ningún polimorfismo con un nivel de asociación estadísticamente significativo según el umbral establecido para este tipo de estudios, el análisis *in silico* de los variantes con una mayor asociación, ha mostrado una sobrerepresentación de enfermedades neurológicas. Dicho hallazgo iría en concordancia con las últimas hipótesis en las que se postula que la FM es debida a una disfunción en el sistema nervioso central y concretamente, en la transmisión del impulso doloroso. Además, el estudio de replicación en el que 3 de los 4 SNPs con mayor asociación (aunque nominal) están localizados en genes implicados en el sistema nervioso (*MCF2*,*ANK3*,*MYT1L*) lo que viene a reforzar esta idea. Finalmente, la asociación de rs11127292 con casos de FM con baja comorbilidad y elevada sintomatología viene a confirmar la posible implicación del sistema nervioso en el desarrollo de esta patología.

El proto oncogen *MCF2* se expresa en diferentes tejidos y está implicado, entre otras funciones, en la ovogénesis, en la proliferación dendrítica y en la apoptosis en el sistema nervioso central. Variantes en el gen *MCF2* se han asociado con autismo y esquizofrenia.

ANK3 (ankyrina 3) pertenece a una familia de proteínas que juegan un papel fundamental en funciones tales como motilidad cellular, activación y proliferación de dominios especializados de membrana. ANK3 tiene sus principales funciones en el cerebro, implicada en el transporte a lo largo del axon y en la polaridad dendrítica. Variantes en *ANK3* se han asociado con esquizofrenia, enfermedad bipolar y autismo.

rs11227292 ha mostrado asociación nominal en el estudio conjunto de toda la cohorte. Al considerar únicamente mujeres con niveles bajos de comorbilidad (clusters 1 y 3), la asociación ha sido ser mayor tanto en el GWAS como en la cohorte de replica y la significación estadística de la asociación del análsis conjunto ha resultado superior. rs11127292 (chr2:2008950-2008950; Hg18) pertenece al tercer intrón del gen *MYT1L* (myelin transcriptor factor 1-like).Variantes en el gne *MYT1L* , tanto CNVs como SNPs, se han asociado con esquizofrenia, depresión mayor y autismo.

El estudio de variantes en el número de copia mediante aCGH también va en la misma dirección, detectando asociación en las muestras con bajos niveles de comorbilidad de una variante intrónica en el gen neurexina 3, que codifica por una molécula transmembrana presináptica esencial para el desarrollo y la función de la sinapsis. Si bien parece que el ratio entre isoformas con y sin exón 20 no parece estar relacionada con la deleción, la presencia de motivos de unión de factores reguladores del splicing,en la región afectada por el CNV, podría indicar que NRXN3_DEL puede asociarse a otras alteraciones en el splicing. Los cambios en el ratio de las dos isoformas en U87 y T98G pudiera deberse a otros facotres como el estadio de la enfermedad (del glioblastoma de origen) y el que provienen de individuos de sexo diferente. Para poder evaluar de forma correcta posibles consecuencias de NRXN3_DEL a nivel de splicing, sería necesario realizar experimentos de RNAseq en diferentes muestras de tejido cerebral provinientes de individuos con genotipos diferentes para la deleción.

El estudio que hemos realizado presenta varias limitaciones. A nivel del GWAs el número reducido de muestras del que disponíamos, hace que no tuviéramos suficiente poder estadístico para detectar vairantes de bajo efecto. El análisis de CNVs a partir de los datos del GWAS también ha mostrado ser poco eficiente para detectar CNVs de pequeño tamaño. Y para la validación de los posibles CNVs en mosaico hubiera sido necesario disponer de diferentes tejidos de un mismo individuo.Finalmente, el uso de pooles en los experimentos de hibridación genómica comparada puede explicar el que no se hayan podido validar cinco de los CNVs detectados. Cualquier pequeño error al cuantificar las muestras para preparar los pooles podría haber dado lugar a la sobrerepresentación de una muestra en concreto. También hay que tener en cuenta que, para variantes frecuentes, y con la elección de las muestras que conformar el pool, se puede introducir un sesgo aleatorio que dé lugar a un falso positivo.

En esta tesis,tanto los SNPs com el CNV detectados como posibles asociaciones genéticas a FM, pertenecen a genes que no sólo se caracterizan por tener su principal función en el sistema nervioso central. Además, dichos genes han sido previamente asociados a enfermedades neuropsiquiátricas . Si se confirman estas asociaciones en otras cohortes, el concepto de fibromialgia como enfermedad reumatológica daría un giro

para pasar a situarse junto a enfermedades neurocognitivas. Asimismo, tanto los SNPs que han presentado una asociación nominal en el estudio conjunto (datos GWAS junto a réplica), uno de ellos presentando asociación en fibromialgia con bajos niveles de comorbilidad, como la deleción del gen neurixina han presentado una mayor asociación al condirerar unícamente mujeres. Dicho hallazgo puede apuntar a la existencia de diferentes factores de susceptibilidad genética para FM según el sexo.

En resumen, el presente estudio sugiere la implicación de variantes, tanto SNPs como CNVs implicados en el sistema nervioso en la etiopatogenia de FM. También pone de manifiesto la importancia en identificar subgrupos homogéneos de la enfermedad para el desarrollo exitoso de estudios genéticos. Dada la complejidad de la patología serán necesarias un mayor número de cohortes de mayor tamaño pero siempre y cuando la clasificación fenotípica sea exhaustiva y permita identificar diferentes subgrupos que puedan tener un diferente *background* genético.

Las principales conclusiones a las que se ha llegado en la presente tesis son:

1 La fibromialgia es una entidad extremadamente compleja. La susceptibilidad genética a padecer la enfermedad parece resultar de la suma de diferentes factores genéticos de bajo efecto. Es por ello indispensable el uso de grandes cohortes que incluyan miles de muestras con el fin de poder detectar dichas variantes.

2 La identificación de subgrupos clínicamente homogéneos de la enfermedad constituye un paso indispensable para la identificación de factores de susceptibilidad genética para fibromialgia. En este sentido, la presente tesis apunta a que la inclusión de comorbilidades tanto personales como familiares puede aportar información complementaria de gran utilidad para la clasificación de la FM basada en síntomas somáticos.

3 Los resultados del GWAS indican un posible contribución del sistema nervioso central en el desarrollo de fibromialgia: las enfermedades neurológicas aparecen como sobrerepresentadas en el estudio de *pathways* realizado en los SNPs que presentaban mayor asociación.

4 Un SNP en el gen *MYT1L* ha presentado asociación estadísticamente significativa con fibromialgia con niveles bajos de comorbilidad, poniendo de manifiesto tanto la posible implicación del sistema nervioso central en la enfermedad, como la importancia de identificar subgrupos clínicamente homogéneos para la detección de factores de susceptibilidad genética para FM.

5 El estudio de replicación del GWAS ha mostrado mayor asociación al considerar únicamente mujeres. Este hecho subraya la importancia del género en la etiopatogenia de la fibromialgia y sugiere la posible existencia de diferentes factores de susceptibilidad genética para cada uno de los géneros.

6 La posible existencia del CNV *SLC2A14*. en mosaicismo en muestras de fibromialgia constituiría una nueva evidencia de la posible implicación del sistema nervioso en la enfermedad. El otro posible CNV en mosaico detectado, implicado en degeneración muscular, apuntaría al rol de otros mecanismos a nivel osteomuscular.

7 Un inserción-deleción intrónica en el gen neurexina 3 ha mostrado asociación en mujeres con fibromialgia, y en particular, en aquellas con bajos niveles de comorbilidad. Dado que la molecula es esencial para el desarrollo, mantenimiento y funcionamiento de la sinapsis, sería un nuevo argumento a favor de la implicación del sistema nervioso en la enfermedad.

8 La confirmación de las variantes detectadas en nuevas cohortes de fibromialgia supondría un giro conceptual de la enfermedad hacia una visión más neurocognitiva que osteomuscular.

REFERENCES

1. Basbaum AI, Bautista DM, Scherrer G, Julius D. Cellular and molecular mechanisms of pain. Cell. 2009 Oct 16;139(2):267-84.
2. Woolf CJ. What is this thing called pain? J Clin Invest. 2010 Nov;120(11):3742-4.
3. Guillot X, Semerano L, Decker P, Falgarone G, Boissier MC. Pain and immunity. Joint Bone Spine. 2011 May;79(3):228-36.
4. Woolf CJ. Central sensitization: implications for the diagnosis and treatment of pain. Pain. 2010 Mar;152(3 Suppl):S2-15.
5. Nussbaum J, Xu Q, Payne TJ, Ma JZ, Huang W, Gelernter J, et al. Significant association of the neurexin-1 gene (NRXN1) with nicotine dependence in European- and African-American smokers. Hum Mol Genet. 2008 Jun 1;17(11):1569-77.
6. Kuner R. Central mechanisms of pathological pain. Nat Med. 2010 Nov;16(11):1258-66.
7. Millan MJ. Descending control of pain. Prog Neurobiol. 2002 Apr;66(6):355-474.
8. Levinson JN, El-Husseini A. New players tip the scales in the balance between excitatory and inhibitory synapses. Mol Pain. 2005;1:12.
9. Villemure C, Bushnell MC. Cognitive modulation of pain: how do attention and emotion influence pain processing? Pain. 2002 Feb;95(3):195-9.
10. Villemure C, Schweinhardt P. Supraspinal pain processing: distinct roles of emotion and attention. Neuroscientist. 2010 Jun;16(3):276-84.
11. Schweinhardt P, Bushnell MC. Pain imaging in health and disease--how far have we come? J Clin Invest. 2010 Nov;120(11):3788-97.
12. Ren K, Dubner R. Descending modulation in persistent pain: an update. Pain. 2002 Nov;100(1-2):1-6.
13. Nielsen LA, Henriksson KG. Pathophysiological mechanisms in chronic musculoskeletal pain (fibromyalgia): the role of central and peripheral sensitization and pain disinhibition. Best Pract Res Clin Rheumatol. 2007 Jun;21(3):465-80.
14. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res (Hoboken). 2010 May;62(5):600-10.
15. Rozental R, Giaume C, Spray DC. Gap junctions in the nervous system. Brain Res Brain Res Rev. 2000 Apr;32(1):11-5.
16. Cowan WM, Kandel,E.R. . A brief history of synapses and synaptic transmission. In: *Cowan WM, Südhof, T. C & Stevens, C.F.*, editor. Synapses. Baltimore: The Jonhs Hopkins University Press; 2003.
17. Sudhof TC. Neuroligins and neurexins link synaptic function to cognitive disease. Nature. 2008 Oct 16;455(7215):903-11.
18. Dean C, Dresbach T. Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function. Trends Neurosci. 2006 Jan;29(1):21-9.
19. Shapiro L, Love J, Colman DR. Adhesion molecules in the nervous system: structural insights into function and diversity. Annu Rev Neurosci. 2007;30:451-74.
20. Craig AM, Kang Y. Neurexin-neuroligin signaling in synapse development. Curr Opin Neurobiol. 2007 Feb;17(1):43-52.
21. Craig AM, Graf ER, Linhoff MW. How to build a central synapse: clues from cell culture. Trends Neurosci. 2006 Jan;29(1):8-20.
22. Ullrich B, Ushkaryov YA, Sudhof TC. Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. Neuron. 1995 Mar;14(3):497-507.
23. Missler M, Fernandez-Chacon R, Sudhof TC. The making of neurexins. J Neurochem. 1998 Oct;71(4):1339-47.
24. Missler M, Zhang W, Rohlmann A, Kattenstroth G, Hammer RE, Gottmann K, et al. Alpha-neurexins couple Ca2+ channels to synaptic vesicle exocytosis. Nature. 2003 Jun 26;423(6943):939-48.
25. Sons MS, Busche N, Strenzke N, Moser T, Ernsberger U, Mooren FC, et al. alpha-Neurexins are required for efficient transmitter release and synaptic homeostasis at the mouse neuromuscular junction. Neuroscience. 2006;138(2):433-46.

26. Ichtchenko K, Hata Y, Nguyen T, Ullrich B, Missler M, Moomaw C, et al. Neuroligin 1: a splice site-specific ligand for beta-neurexins. Cell. 1995 May 5;81(3):435-43.
27. Ushkaryov YA, Petrenko AG, Geppert M, Sudhof TC. Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin. Science. 1992 Jul 3;257(5066):50-6.
28. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008 Apr 25;320(5875):539-43.
29. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet. 2003 May;34(1):27-9.
30. Lawson-Yuen A, Saldivar JS, Sommer S, Picker J. Familial deletion within NLGN4 associated with autism and Tourette syndrome. Eur J Hum Genet. 2008 May;16(5):614-8.
31. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature. 2007 Jan 11;445(7124):168-76.
32. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet. 2007 Jan 1;16(1):24-35.
33. Novak G, Boukhadra J, Shaikh SA, Kennedy JL, Le Foll B. Association of a polymorphism in the NRXN3 gene with the degree of smoking in schizophrenia: a preliminary study. World J Biol Psychiatry. 2009;10(4 Pt 3):929-35.
34. Stoltenberg SF, Lehmann MK, Christ CC, Hersrud SL, Davies GE. Associations among types of impulsivity, substance use problems and Neurexin-3 polymorphisms. Drug Alcohol Depend. 2011 Jun 13.
35. Lachman HM, Fann CS, Bartzis M, Evgrafov OV, Rosenthal RN, Nunes EV, et al. Genomewide suggestive linkage of opioid dependence to chromosome 14q. Hum Mol Genet. 2007 Jun 1;16(11):1327-34.
36. Hishimoto A, Liu QR, Drgon T, Pletnikova O, Walther D, Zhu XG, et al. Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms. Hum Mol Genet. 2007 Dec 1;16(23):2880-91.
37. Heard-Costa NL, Zillikens MC, Monda KL, Johansson A, Harris TB, Fu M, et al. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. PLoS Genet. 2009 Jun;5(6):e1000539.
38. Ching MS, Shen Y, Tan WH, Jeste SS, Morrow EM, Chen X, et al. Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. Am J Med Genet B Neuropsychiatr Genet. 2010 Jun 5;153B(4):937-47.
39. Gauthier J, Siddiqui TJ, Huashan P, Yokomaku D, Hamdan FF, Champagne N, et al. Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. Hum Genet. 2011 Oct;130(4):563-73.
40. Kim HG, Kishikawa S, Higgins AW, Seong IS, Donovan DJ, Shen Y, et al. Disruption of neurexin 1 associated with autism spectrum disorder. Am J Hum Genet. 2008 Jan;82(1):199-207.
41. Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, Curran S, et al. Rare Deletions at the Neurexin 3 Locus in Autism Spectrum Disorder. Am J Hum Genet. 2011 Dec 28.
42. Wisniowiecka-Kowalnik B, Nesteruk M, Peters SU, Xia Z, Cooper ML, Savage S, et al. Intragenic rearrangements in NRXN1 in three families with autism spectrum disorder, developmental delay, and speech delay. Am J Med Genet B Neuropsychiatr Genet. 2010 Jul;153B(5):983-93.
43. Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, et al. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. Hum Mol Genet. 2008 Feb 1;17(3):458-65.
44. Fitzcharles MA, Yunus MB. The clinical concept of fibromyalgia as a changing paradigm in the past 20 years. Pain Res Treat. 2012;2012:184835.
45. Burkham J, Harris E. D. Fibromyalgia: A chronic Pain Syndrome. In: Harris ED, Budd R., Genovese M.C., Firestein G. S.,sargent J.S., Sledge C.B., editor. Kelley's textbook of Rheumatology. 7 ed. Pennsylvania: Elsevier; 2005. p. 522-36.
46. Yunus MB, Khan MA, Rawlings KK, Green JR, Olson JM, Shah S. Genetic linkage analysis of multicase families with fibromyalgia syndrome. J Rheumatol. 1999 Feb;26(2):408-12.

47. Ballina F. Fibromialgia. In: Reumatología SEd, editor. Manual SER de las Enfermedades Reumáticas. 4 ed. Madrid: Editorial Medica panamericana; 2004. p. 112-5.
48. Rivera J AC, Ballina FJ, Carbonell J, Carmona L, Castel B, Collado A, Esteve JJ, Martínez FG, Tornero J, Vallejo MA, Vidal J. Consensus document on fibromyalgia of the Spanish Society of Rheumatology Reumatología Clínica. 2006;2S1:55-66.
49. Osorio CD, Gallinaro AL, Lorenzi-Filho G, Lage LV. Sleep quality in patients with fibromyalgia using the Pittsburgh Sleep Quality Index. J Rheumatol. 2006 Sep;33(9):1863-5.
50. Roizenblatt S, Moldofsky H, Benedito-Silva AA, Tufik S. Alpha sleep characteristics in fibromyalgia. Arthritis Rheum. 2001 Jan;44(1):222-30.
51. Hamilton NA, Affleck G, Tennen H, Karlson C, Luxton D, Preacher KJ, et al. Fibromyalgia: the role of sleep in affect and in negative event reactivity and recovery. Health Psychol. 2008 Jul;27(4):490-7.
52. Katz RS, Heard AR, Mills M, Leavitt F. The prevalence and clinical impact of reported cognitive difficulties (fibrofog) in patients with rheumatic disease with and without fibromyalgia. J Clin Rheumatol. 2004 Apr;10(2):53-8.
53. Dick B, Eccleston C, Crombez G. Attentional functioning in fibromyalgia, rheumatoid arthritis, and musculoskeletal pain patients. Arthritis Rheum. 2002 Dec 15;47(6):639-44.
54. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum. 2008 Jan;58(1):26-35.
55. Branco JC, Bannwarth B, Failde I, Abello Carbonell J, Blotman F, Spaeth M, et al. Prevalence of fibromyalgia: a survey in five European countries. Semin Arthritis Rheum. 2010 Jun;39(6):448-53.
56. Reumatología SEd. Estudio EPISER. Prevalencia e impacto de las enfermedades reumáticas en la población adulta española. In: Reumatología SEd, editor. Madrid; 2001.
57. Mourao AF, Blyth FM, Branco JC. Generalised musculoskeletal pain syndromes. Best Pract Res Clin Rheumatol. 2010 Dec;24(6):829-40.
58. White KP, Thompson J. Fibromyalgia syndrome in an Amish community: a controlled study to determine disease and symptom prevalence. J Rheumatol. 2003 Aug;30(8):1835-40.
59. Nuesch E, Hauser W, Bernardy K, Barth J, Juni P. Comparative efficacy of pharmacological and non-pharmacological interventions in fibromyalgia syndrome: network meta-analysis. Ann Rheum Dis. 2012 Jun 27.
60. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 Criteria for the Classification of Fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum. 1990 Feb;33(2):160-72.
61. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Hauser W, Katz RS, et al. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: a modification of the ACR Preliminary Diagnostic Criteria for Fibromyalgia. J Rheumatol. 2011 Jun;38(6):1113-22.
62. Goldenberg DL, Burckhardt C, Crofford L. Management of fibromyalgia syndrome. JAMA. 2004 Nov 17;292(19):2388-95.
63. Dreyer L, Kendall S, Danneskiold-Samsoe B, Bartels EM, Bliddal H. Mortality in a cohort of Danish patients with fibromyalgia: increased frequency of suicide. Arthritis Rheum. 2010 Oct;62(10):3101-8.
64. Macfarlane GJ, Jones GT, Knekt P, Aromaa A, McBeth J, Mikkelsson M, et al. Is the report of widespread body pain associated with long-term increased mortality? Data from the Mini-Finland Health Survey. Rheumatology (Oxford). 2007 May;46(5):805-7.
65. McBeth J, Symmons DP, Silman AJ, Allison T, Webb R, Brammah T, et al. Musculoskeletal pain is associated with a long-term increased risk of cancer and cardiovascular-related mortality. Rheumatology (Oxford). 2009 Jan;48(1):74-7.
66. Bennett RM, Jones J, Turk DC, Russell IJ, Matallana L. An internet survey of 2,596 people with fibromyalgia. BMC Musculoskelet Disord. 2007;8:27.
67. Bernard AL, Prince A, Edsall P. Quality of life issues for fibromyalgia patients. Arthritis Care Res. 2000 Feb;13(1):42-50.

68. Clauw DJ, Arnold LM, McCarberg BH. The science of fibromyalgia. Mayo Clin Proc. 2011 Sep;86(9):907-11.
69. Silverman S, Dukes EM, Johnston SS, Brandenburg NA, Sadosky A, Huse DM. The economic burden of fibromyalgia: comparative analysis with rheumatoid arthritis. Curr Med Res Opin. 2009 Apr;25(4):829-40.
70. Sumpton JE, Moulin DE. Fibromyalgia: presentation and management with a focus on pharmacological treatment. Pain Res Manag. 2008 Nov-Dec;13(6):477-83.
71. Castel A, Fontova R, Montull S, Perinan R, Poveda MJ, Miralles I, et al. The efficacy of a multidisciplinary fibromyalgia treatment adapted for women with low educational levels: A randomised, controlled trial. Arthritis Care Res (Hoboken). 2012 Aug 16.
72. Carville SF, Arendt-Nielsen S, Bliddal H, Blotman F, Branco JC, Buskila D, et al. EULAR evidence-based recommendations for the management of fibromyalgia syndrome. Ann Rheum Dis. 2008 Apr;67(4):536-41.
73. Zhou M, Zhou D, He L. A randomized trial of tai chi for fibromyalgia. N Engl J Med. 2010 Dec 2;363(23):2265; author reply 6-7.
74. Kuchinad A, Schweinhardt P, Seminowicz DA, Wood PB, Chizh BA, Bushnell MC. Accelerated brain gray matter loss in fibromyalgia patients: premature aging of the brain? J Neurosci. 2007 Apr 11;27(15):4004-7.
75. Gracely RH, Ambrose KR. Neuroimaging of fibromyalgia. Best Pract Res Clin Rheumatol. 2011 Apr;25(2):271-84.
76. Dadabhoy D, Crofford LJ, Spaeth M, Russell IJ, Clauw DJ. Biology and therapy of fibromyalgia. Evidence-based biomarkers for fibromyalgia syndrome. Arthritis Res Ther. 2008;10(4):211.
77. Gibson SJ, Littlejohn GO, Gorman MM, Helme RD, Granges G. Altered heat pain thresholds and cerebral event-related potentials following painful CO2 laser stimulation in subjects with fibromyalgia syndrome. Pain. 1994 Aug;58(2):185-93.
78. Diers M, Schley MT, Rance M, Yilmaz P, Lauer L, Rukwied R, et al. Differential central pain processing following repetitive intramuscular proton/prostaglandin E(2) injections in female fibromyalgia patients and healthy controls. Eur J Pain. 2011 Aug;15(7):716-23.
79. Wood PB, Patterson JC, 2nd, Sunderland JJ, Tainter KH, Glabus MF, Lilien DL. Reduced presynaptic dopamine activity in fibromyalgia syndrome demonstrated with positron emission tomography: a pilot study. J Pain. 2007 Jan;8(1):51-8.
80. Harris RE, Clauw DJ, Scott DJ, McLean SA, Gracely RH, Zubieta JK. Decreased central mu-opioid receptor availability in fibromyalgia. J Neurosci. 2007 Sep 12;27(37):10000-6.
81. Fayed N, Garcia-Campayo J, Magallon R, Andres-Bergareche H, Luciano JV, Andres E, et al. Localized 1H-NMR spectroscopy in patients with fibromyalgia: a controlled study of changes in cerebral glutamate/glutamine, inositol, choline, and N-acetylaspartate. Arthritis Res Ther. 2010;12(4):R134.
82. Julien N, Goffaux P, Arsenault P, Marchand S. Widespread pain in fibromyalgia is related to a deficit of endogenous pain inhibition. Pain. 2005 Mar;114(1-2):295-302.
83. Legangneux E, Mora JJ, Spreux-Varoquaux O, Thorin I, Herrou M, Alvado G, et al. Cerebrospinal fluid biogenic amine metabolites, plasma-rich platelet serotonin and [3H]imipramine reuptake in the primary fibromyalgia syndrome. Rheumatology (Oxford). 2001 Mar;40(3):290-6.
84. Di Franco M, Iannuccelli C, Valesini G. Neuroendocrine immunology of fibromyalgia. Ann N Y Acad Sci. 2010 Apr;1193:84-90.
85. Larson AA, Giovengo SL, Russell IJ, Michalek JE. Changes in the concentrations of amino acids in the cerebrospinal fluid that correlate with pain in patients with fibromyalgia: implications for nitric oxide pathways. Pain. 2000 Aug;87(2):201-11.
86. Seidel MF, Herguijuela M, Forkert R, Otten U. Nerve growth factor in rheumatic diseases. Semin Arthritis Rheum. 2010 Oct;40(2):109-26.
87. Staud R, Nagel S, Robinson ME, Price DD. Enhanced central pain processing of fibromyalgia patients is maintained by muscle afferent input: a randomized, double-blind, placebo-controlled study. Pain. 2009 Sep;145(1-2):96-104.
88. Martinez-Lavin M. Biology and therapy of fibromyalgia. Stress, the stress response system, and fibromyalgia. Arthritis Res Ther. 2007;9(4):216.

89. Moldofsky H, Lue FA. The relationship of alpha and delta EEG frequencies to pain and mood in 'fibrositis' patients treated with chlorpromazine and L-tryptophan. Electroencephalogr Clin Neurophysiol. 1980 Oct;50(1-2):71-80.
90. Lentz MJ, Landis CA, Rothermel J, Shaver JL. Effects of selective slow wave sleep disruption on musculoskeletal pain and fatigue in middle aged women. J Rheumatol. 1999 Jul;26(7):1586-92.
91. Bennett RM, Clark SR, Campbell SM, Burckhardt CS. Low levels of somatomedin C in patients with the fibromyalgia syndrome. A possible link between sleep and muscle pain. Arthritis Rheum. 1992 Oct;35(10):1113-6.
92. Bartels EM, Danneskiold-Samsoe B. Histological abnormalities in muscle from patients with certain types of fibrositis. Lancet. 1986 Apr 5;1(8484):755-7.
93. Dantzer R, O'Connor JC, Freund GG, Johnson RW, Kelley KW. From inflammation to sickness and depression: when the immune system subjugates the brain. Nat Rev Neurosci. 2008 Jan;9(1):46-56.
94. Uceyler N, Hauser W, Sommer C. Systematic review with meta-analysis: cytokines in fibromyalgia syndrome. BMC Musculoskelet Disord. 2011;12:245.
95. Walker EA, Keegan D, Gardner G, Sullivan M, Bernstein D, Katon WJ. Psychosocial factors in fibromyalgia compared with rheumatoid arthritis: II. Sexual, physical, and emotional abuse and neglect. Psychosom Med. 1997 Nov-Dec;59(6):572-7.
96. Al-Allaf AW, Dunbar KL, Hallum NS, Nosratzadeh B, Templeton KD, Pullar T. A case-control study examining the role of physical trauma in the onset of fibromyalgia syndrome. Rheumatology (Oxford). 2002 Apr;41(4):450-3.
97. Atherton K, Berry DJ, Parsons T, Macfarlane GJ, Power C, Hypponen E. Vitamin D and chronic widespread pain in a white middle-aged British population: evidence from a cross-sectional population survey. Ann Rheum Dis. 2009 Jun;68(6):817-22.
98. Buskila D, Atzeni F, Sarzi-Puttini P. Etiology of fibromyalgia: the possible role of infection and vaccination. Autoimmun Rev. 2008 Oct;8(1):41-3.
99. Ablin J, Neumann L, Buskila D. Pathogenesis of fibromyalgia - a review. Joint Bone Spine. 2008 May;75(3):273-9.
100. Unwin C, Blatchley N, Coker W, Ferry S, Hotopf M, Hull L, et al. Health of UK servicemen who served in Persian Gulf War. Lancet. 1999 Jan 16;353(9148):169-78.
101. Hotopf M, David A, Hull L, Ismail K, Unwin C, Wessely S. Role of vaccinations as risk factors for ill health in veterans of the Gulf war: cross sectional study. BMJ. 2000 May 20;320(7246):1363-7.
102. Norbury TA, MacGregor AJ, Urwin J, Spector TD, McMahon SB. Heritability of responses to painful stimuli in women: a classical twin study. Brain. 2007 Nov;130(Pt 11):3041-9.
103. Mogil JS. Pain genetics: past, present and future. Trends Genet. 2012 Jun;28(6):258-66.
104. Kim H, Ramsay E, Lee H, Wahl S, Dionne RA. Genome-wide association study of acute post-surgical pain in humans. Pharmacogenomics. 2009 Feb;10(2):171-9.
105. Smith SB, Maixner DW, Fillingim RB, Slade G, Gracely RH, Ambrose K, et al. Large candidate gene association study reveals genetic risk factors and therapeutic targets for fibromyalgia. Arthritis Rheum. 2012 Feb;64(2):584-93.
106. Pellegrino MJ, Waylonis GW, Sommer A. Familial occurrence of primary fibromyalgia. Arch Phys Med Rehabil. 1989 Jan;70(1):61-3.
107. Buskila D, Neumann L, Hazanov I, Carmi R. Familial aggregation in the fibromyalgia syndrome. Semin Arthritis Rheum. 1996 Dec;26(3):605-11.
108. Buskila D, Neumann L. Fibromyalgia syndrome (FM) and nonarticular tenderness in relatives of patients with FM. J Rheumatol. 1997 May;24(5):941-4.
109. Arnold LM, Hudson JI, Hess EV, Ware AE, Fritz DA, Auchenbach MB, et al. Family study of fibromyalgia. Arthritis Rheum. 2004 Mar;50(3):944-52.
110. Mikkelsson M, Kaprio J, Salminen JJ, Pulkkinen L, Rose RJ. Widespread pain among 11-year-old Finnish twin pairs. Arthritis Rheum. 2001 Feb;44(2):481-5.
111. Kato K, Sullivan PF, Evengard B, Pedersen NL. Importance of genetic influences on chronic widespread pain. Arthritis Rheum. 2006 May;54(5):1682-6.

112. Burda CD, Cox FR, Osborne P. Histocompatability antigens in the fibrositis (fibromyalgia) syndrome. Clin Exp Rheumatol. 1986 Oct-Dec;4(4):355-8.
113. Horven S, Stiles TC, Holst A, Moen T. HLA antigens in primary fibromyalgia syndrome. J Rheumatol. 1992 Aug;19(8):1269-70.
114. Offenbaecher M, Bondy B, de Jonge S, Glatzeder K, Kruger M, Schoeps P, et al. Possible association of fibromyalgia with a polymorphism in the serotonin transporter gene regulatory region. Arthritis Rheum. 1999 Nov;42(11):2482-8.
115. Cohen H, Buskila D, Neumann L, Ebstein RP. Confirmation of an association between fibromyalgia and serotonin transporter promoter region (5- HTTLPR) polymorphism, and relationship to anxiety-related personality traits. Arthritis Rheum. 2002 Mar;46(3):845-7.
116. Gursoy S. Absence of association of the serotonin transporter gene polymorphism with the mentally healthy subset of fibromyalgia patients. Clin Rheumatol. 2002 Jun;21(3):194-7.
117. Bondy B, Spaeth M, Offenbaecher M, Glatzeder K, Stratz T, Schwarz M, et al. The T102C polymorphism of the 5-HT2A-receptor gene in fibromyalgia. Neurobiol Dis. 1999 Oct;6(5):433-9.
118. Gursoy S, Erdal E, Herken H, Madenci E, Alasehirli B. Association of T102C polymorphism of the 5-HT2A receptor gene with psychiatric status in fibromyalgia syndrome. Rheumatol Int. 2001 Oct;21(2):58-61.
119. Zubieta JK, Heitzeg MM, Smith YR, Bueller JA, Xu K, Xu Y, et al. COMT val158met genotype affects mu-opioid neurotransmitter responses to a pain stressor. Science. 2003 Feb 21;299(5610):1240-3.
120. Vargas-Alarcon G, Fragoso JM, Cruz-Robles D, Vargas A, Lao-Villadoniga JI, Garcia-Fructuoso F, et al. Catechol-O-methyltransferase gene haplotypes in Mexican and Spanish patients with fibromyalgia. Arthritis Res Ther. 2007;9(5):R110.
121. Tander B, Gunes S, Boke O, Alayli G, Kara N, Bagci H, et al. Polymorphisms of the serotonin-2A receptor and catechol-O-methyltransferase genes: a study on fibromyalgia susceptibility. Rheumatol Int. 2008 May;28(7):685-91.
122. Lee YH, Choi SJ, Ji JD, Song GG. Candidate gene studies of fibromyalgia: a systematic review and meta-analysis. Rheumatol Int. 2012 Feb;32(2):417-26.
123. Martinez-Jauand M, Sitges C, Rodriguez V, Picornell A, Ramon M, Buskila D, et al. Pain sensitivity in fibromyalgia is associated with catechol-O-methyltransferase (COMT) gene. Eur J Pain. 2012 Apr 24.
124. Vargas-Alarcon G, Fragoso JM, Cruz-Robles D, Vargas A, Martinez A, Lao-Villadoniga JI, et al. Association of adrenergic receptor gene polymorphisms with different fibromyalgia syndrome domains. Arthritis Rheum. 2009 Jul;60(7):2169-73.
125. Potvin S, Larouche A, Normand E, de Souza JB, Gaumond I, Marchand S, et al. No relationship between the ins del polymorphism of the serotonin transporter promoter and pain perception in fibromyalgia patients and healthy controls. Eur J Pain. 2010 Aug;14(7):742-6.
126. Gursoy S, Erdal E, Herken H, Madenci E, Alasehirli B, Erdal N. Significance of catechol-O-methyltransferase gene polymorphism in fibromyalgia syndrome. Rheumatol Int. 2003 May;23(3):104-7.
127. Ablin JN, Bar-Shira A, Yaron M, Orr-Urtreger A. Candidate-gene approach in fibromyalgia syndrome: association analysis of the genes encoding substance P receptor, dopamine transporter and alpha1-antitrypsin. Clin Exp Rheumatol. 2009 Sep-Oct;27(5 Suppl 56):S33-8.
128. Buskila D, Cohen H, Neumann L, Ebstein RP. An association between fibromyalgia and the dopamine D4 receptor exon III repeat polymorphism and relationship to novelty seeking personality traits. Mol Psychiatry. 2004 Aug;9(8):730-1.
129. Alasehirli B, Demiryurek S, Arica E, Gursoy S, Demiryurek AT. No evidence for an association between the Glu298Asp polymorphism of the endothelial nitric oxide synthase gene and fibromyalgia syndrome. Rheumatol Int. 2007 Jan;27(3):275-80.
130. Frank B, Niesler B, Bondy B, Spath M, Pongratz DE, Ackenheil M, et al. Mutational analysis of serotonin receptor genes: HTR3A and HTR3B in fibromyalgia patients. Clin Rheumatol. 2004 Aug;23(4):338-44.
131. Su SY, Chen JJ, Lai CC, Chen CM, Tsai FJ. The association between fibromyalgia and polymorphism of monoamine oxidase A and interleukin-4. Clin Rheumatol. 2007 Jan;26(1):12-6.

132. Gursoy S, Erdal E, Sezgin M, Barlas IO, Aydeniz A, Alasehirli B, et al. Which genotype of MAO gene that the patients have are likely to be most susceptible to the symptoms of fibromyalgia? Rheumatol Int. 2008 Feb;28(4):307-11.
133. Vargas-Alarcon G, Alvarez-Leon E, Fragoso JM, Vargas A, Martinez A, Vallejo M, et al. A SCN9A gene-encoded dorsal root ganglia sodium channel polymorphism associated with severe fibromyalgia. BMC Musculoskelet Disord. 2012;13:23.
134. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921.
135. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001 Feb 16;291(5507):1304-51.
136. Consortium TGP. A map of human genome variation from population-scale sequencing. Nature. 2010 Oct 28;467(7319):1061-73.
137. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011;13(1):101.
138. Hapmap. The International HapMap Project. Nature. 2003 Dec 18;426(6968):789-96.
139. Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC, Lambert SM, et al. The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. J Biomol Tech. 2005 Dec;16(4):398-406.
140. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005 Apr 15;308(5720):385-9.
141. Amos CI. Successful design and conduct of genome-wide association studies. Hum Mol Genet. 2007 Oct 15;16 Spec No. 2:R220-5.
142. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010 Jul;11(7):499-511.
143. Consortium TWTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007 Jun 7;447(7145):661-78.
144. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009 Apr 9;458(7239):719-24.
145. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, et al. Fine-scale structural variation of the human genome. Nat Genet. 2005 Jul;37(7):727-32.
146. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010 Apr 1;464(7289):704-12.
147. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006 Nov 23;444(7118):444-54.
148. Estivill X, Armengol L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. PLoS Genet. 2007 Oct;3(10):1787-99.
149. Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nat Genet. 2009 Jul;41(7):849-53.
150. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell. 2012 Mar 16;148(6):1223-41.
151. Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask BJ, et al. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell. 1991 Jul 26;66(2):219-32.
152. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science. 2005 Mar 4;307(5714):1434-40.
153. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet. 2006 Sep;79(3):439-48.
154. Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, et al. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. Am J Hum Genet. 2007 Jun;80(6):1037-54.

155. de Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, Dannhauser EN, et al. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. Nat Genet. 2009 Feb;41(2):211-5.
156. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature. 2010 Apr 1;464(7289):713-20.
157. Lupski JR, Stankiewicz P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet. 2005 Dec;1(6):e49.
158. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science. 2007 Oct 19;318(5849):420-6.
159. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004 Sep;36(9):949-51.
160. Ju YS, Hong D, Kim S, Park SS, Lee S, Park H, et al. Reference-unbiased copy number variant analysis using CGH microarrays. Nucleic Acids Res. 2010 Nov;38(20):e190.
161. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007 Nov;17(11):1665-74.
162. Winchester L, Yau C, Ragoussis J. Comparing CNV detection methods for SNP arrays. Brief Funct Genomic Proteomic. 2009 Sep;8(5):353-66.
163. Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, et al. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet. 2010 Jun;42(6):489-91.
164. Koboldt DC, Larson DE, Chen K, Ding L, Wilson RK. Massively parallel sequencing approaches for characterization of structural variation. Methods Mol Biol. 2012;838:369-84.
165. Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, et al. Challenges and standards in integrating surveys of structural variation. Nat Genet. 2007 Jul;39(7 Suppl):S7-15.
166. Florijn RJ, Bonden LA, Vrolijk H, Wiegant J, Vaandrager JW, Baas F, et al. High-resolution DNA Fiber-FISH for genomic DNA mapping and colour bar-coding of large genes. Hum Mol Genet. 1995 May;4(5):831-6.
167. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 2002 Jun 15;30(12):e57.
168. Flores C, Maca-Meyer N, Gonzalez AM, Oefner PJ, Shen P, Perez JA, et al. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. Eur J Hum Genet. 2004 Oct;12(10):855-63.
169. Laayouni H, Calafell F, Bertranpetit J. A genome-wide survey does not show the genetic distinctiveness of Basques. Hum Genet. 2010 Apr;127(4):455-8.
170. Pino-Yanes M, Corrales A, Basaldua S, Hernandez A, Guerra L, Villar J, et al. North African influences and potential bias in case-control association studies in the Spanish population. PLoS One. 2011;6(3):e18389.
171. Julia A, Ballina J, Canete JD, Balsa A, Tornero-Molina J, Naranjo A, et al. Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility. Arthritis Rheum. 2008 Aug;58(8):2275-86.
172. Aktas A, Walsh D, Rybicki L. Symptom clusters: myth or reality? Palliat Med. 2010 Jun;24(4):373-85.
173. de Souza JB, Goffaux P, Julien N, Potvin S, Charest J, Marchand S. Fibromyalgia subgroups: profiling distinct subgroups using the Fibromyalgia Impact Questionnaire. A preliminary study. Rheumatol Int. 2009 Mar;29(5):509-15.
174. Hurtig IM, Raak RI, Kendall SA, Gerdle B, Wahren LK. Quantitative sensory testing in fibromyalgia patients and in healthy subjects: identification of subgroups. Clin J Pain. 2001 Dec;17(4):316-22.
175. Giesecke T, Williams DA, Harris RE, Cupps TR, Tian X, Tian TX, et al. Subgrouping of fibromyalgia patients on the basis of pressure-pain thresholds and psychological factors. Arthritis Rheum. 2003 Oct;48(10):2916-22.
176. Loevinger BL, Shirtcliff EA, Muller D, Alonso C, Coe CL. Delineating psychological and biomedical profiles in a heterogeneous fibromyalgia population using cluster analysis. Clin Rheumatol. 2012 Dec 27.

177. Bennett RM, Russell J, Cappelleri JC, Bushmakin AG, Zlateva G, Sadosky A. Identification of symptom and functional domains that fibromyalgia patients would like to see improved: a cluster analysis. BMC Musculoskelet Disord. 2010;11:134.
178. Cuesta-Vargas A, Luciano JV, Penarrubia-Maria MT, Garcia-Campayo J, Fernandez-Vergel R, Arroyo-Morales M, et al. Clinical dimensions of fibromyalgia symptoms and development of a combined index of severity: The CODI index. Qual Life Res. 2012 Feb 26.
179. Rutledge DN, Mouttapa M, Wood PB. Symptom clusters in fibromyalgia: potential utility in patient assessment and treatment evaluation. Nurs Res. 2009 Sep-Oct;58(5):359-67.
180. Wilson HD, Robinson JP, Turk DC. Toward the identification of symptom patterns in people with fibromyalgia. Arthritis Rheum. 2009 Apr 15;61(4):527-34.
181. Okifuji A, Turk DC, Sinclair JD, Starz TW, Marcus DA. A standardized manual tender point survey. I. Development and determination of a threshold point for the identification of positive tender points in fibromyalgia syndrome. J Rheumatol. 1997 Feb;24(2):377-83.
182. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983 Jun;67(6):361-70.
183. Herrero MJ, Blanch J, Peri JM, De Pablo J, Pintor L, Bulbena A. A validation study of the hospital anxiety and depression scale (HADS) in a Spanish population. Gen Hosp Psychiatry. 2003 Jul-Aug;25(4):277-83.
184. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989 May;28(2):193-213.
185. Burckhardt CS, Clark SR, Bennett RM. The fibromyalgia impact questionnaire: development and validation. J Rheumatol. 1991 May;18(5):728-33.
186. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol. 1989 Oct;46(10):1121-3.
187. Benito-Leon J, Martinez-Martin P, Frades B, Martinez-Gines ML, de Andres C, Meca-Lallana JE, et al. Impact of fatigue in multiple sclerosis: the Fatigue Impact Scale for Daily Use (D-FIS). Mult Scler. 2007 Jun;13(5):645-51.
188. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 Jun;30(6):473-83.
189. Alonso J, Prieto L, Anto JM. [The Spanish version of the SF-36 Health Survey (the SF-36 health questionnaire): an instrument for measuring clinical results]. Med Clin (Barc). 1995 May 27;104(20):771-6.
190. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. Ann Intern Med. 1994 Dec 15;121(12):953-9.
191. Fraley C R, A.E. How many clusters? Which clustering method? . The Computer Journal. 1998;41(8):578-88.
192. Gower. A general coefficient of similarity and some of its properties. Biometrics. 1971;27:857-74.
193. Hartigan JAWMA. Algorithm AS 136: a K-means clustering algorithm. *Appl Stat*. 1978;28:100-8.
194. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559-75.
195. Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ, et al. WGAViewer: software for genomic annotation of whole genome association studies. Genome Res. 2008 Apr;18(4):640-3.
196. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263-5.
197. Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, et al. Patterns of cis regulatory variation in diverse human populations. PLoS Genet. 2012;8(4):e1002639.
198. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet. 2011;7(2):e1002003.
199. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009 Sep 4;325(5945):1246-50.

200. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012 Sep;22(9):1790-7.
201. Zhang D, Qian Y, Akula N, Alliey-Rodriguez N, Tang J, Gershon ES, et al. Accuracy of CNV Detection from GWAS Data. PLoS One. 2011;6(1):e14511.
202. Gonzalez JR, Rodriguez-Santiago B, Caceres A, Pique-Regi R, Rothman N, Chanock SJ, et al. A fast and accurate method to detect allelic genomic imbalances underlying mosaic rearrangements using SNP array data. BMC Bioinformatics. 2011;12:166.
203. Bierut LJ. Genetic vulnerability and susceptibility to substance dependence. Neuron. 2011 Feb 24;69(4):618-27.
204. Occhi G, Rampazzo A, Beffagna G, Antonio Danieli G. Identification and characterization of heart-specific splicing of human neurexin 3 mRNA (NRXN3). Biochem Biophys Res Commun. 2002 Oct 18;298(1):151-5.
205. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001 May 1;29(9):e45.
206. Ladd AN, Cooper TA. Finding signals that regulate alternative splicing in the post-genomic era. Genome Biol. 2002 Oct 23;3(11):reviews0008.
207. Park H, Kim JI, Ju YS, Gokcumen O, Mills RE, Kim S, et al. Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. Nat Genet. 2010 May;42(5):400-5.
208. Resnick M, Segall A, G GR, Lupowitz Z, Zisapel N. Alternative splicing of neurexins: a role for neuronal polypyrimidine tract binding protein. Neurosci Lett. 2008 Jul 18;439(3):235-40.
209. Kashikar-Zuck S, Lynch AM, Slater S, Graham TB, Swain NF, Noll RB. Family factors, emotional functioning, and functional impairment in juvenile fibromyalgia syndrome. Arthritis Rheum. 2008 Oct 15;59(10):1392-8.
210. Kindler LL, Jones KD, Perrin N, Bennett RM. Risk factors predicting the development of widespread pain from chronic back or neck pain. J Pain. 2010 Dec;11(12):1320-8.
211. Gaston-Johansson F, Fall-Dickson JM, Bakos AB, Kennedy MJ. Fatigue, pain, and depression in pre-autotransplant breast cancer patients. Cancer Pract. 1999 Sep-Oct;7(5):240-7.
212. Rehm SE, Koroschetz J, Gockel U, Brosz M, Freynhagen R, Tolle TR, et al. A cross-sectional survey of 3035 patients with fibromyalgia: subgroups of patients with typical comorbidities and sensory symptom profiles. Rheumatology (Oxford). 2010 Jun;49(6):1146-52.
213. Piton A, Gauthier J, Hamdan FF, Lafreniere RG, Yang Y, Henrion E, et al. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. Mol Psychiatry. 2011 Aug;16(8):867-80.
214. Schulze TG, Detera-Wadleigh SD, Akula N, Gupta A, Kassem L, Steele J, et al. Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. Mol Psychiatry. 2009 May;14(5):487-91.
215. Bi C, Wu J, Jiang T, Liu Q, Cai W, Yu P, et al. Mutations of ANK3 identified by exome sequencing are associated with Autism susceptibility. Hum Mutat. 2012 Dec;33(12):1635-8.
216. Romm E, Nielsen JA, Kim JG, Hudson LD. Myt1 family recruits histone deacetylase to regulate neural transcription. J Neurochem. 2005 Jun;93(6):1444-53.
217. Kim JG, Armstrong RC, v Agoston D, Robinsky A, Wiese C, Nagle J, et al. Myelin transcription factor 1 (Myt1) of the oligodendrocyte lineage, along with a closely related CCHC zinc finger, is expressed in developing neurons in the mammalian central nervous system. J Neurosci Res. 1997 Oct 15;50(2):272-90.
218. Yoo AS, Sun AX, Li L, Shcheglovitov A, Portmann T, Li Y, et al. MicroRNA-mediated conversion of human fibroblasts to neurons. Nature. 2011 Aug 11;476(7359):228-31.
219. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. Nature. 2010 Feb 25;463(7284):1035-41.
220. Pfisterer U, Kirkeby A, Torper O, Wood J, Nelander J, Dufour A, et al. Direct conversion of human fibroblasts to dopaminergic neurons. Proc Natl Acad Sci U S A. 2011 Jun 21;108(25):10343-8.

221. Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, Strengman E, Sabatti C, Geurts van Kessel A, et al. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. Am J Hum Genet. 2008 Oct;83(4):504-10.
222. Li W, Wang X, Zhao J, Lin J, Song XQ, Yang Y, et al. Association study of myelin transcription factor 1-like polymorphisms with schizophrenia in Han Chinese population. Genes Brain Behav. 2012 Feb;11(1):87-93.
223. Meyer KJ, Axelsen MS, Sheffield VC, Patil SR, Wassink TH. Germline mosaic transmission of a novel duplication of PXDN and MYT1L to two male half-siblings with autism. Psychiatr Genet. 2012 Jun;22(3):137-40.
224. Wang T, Zeng Z, Li T, Liu J, Li J, Li Y, et al. Common SNPs in myelin transcription factor 1-like (MYT1L): association with major depressive disorder in the Chinese Han population. PLoS One. 2010;5(10):e13662.
225. Yamakawa H, Oyama S, Mitsuhashi H, Sasagawa N, Uchino S, Kohsaka S, et al. Neuroligins 3 and 4X interact with syntrophin-gamma2, and the interactions are affected by autism-related mutations. Biochem Biophys Res Commun. 2007 Mar 30;355(1):41-6.
226. Perlis RH, Ruderfer D, Maussion G, Chambert K, Gallagher P, Turecki G, et al. Bipolar Disorder and a History of Suicide Attempts With a Duplication in 5HTR1A. Am J Psychiatry. 2012 Nov 1;169(11):1213-4.
227. Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, et al. Complement factor H variant increases the risk of age-related macular degeneration. Science. 2005 Apr 15;308(5720):419-21.
228. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009 Aug 6;460(7256):748-52.
229. Aparicio VA, Ortega FB, Carbonell-Baeza A, Femia P, Tercedor P, Ruiz JR, et al. Are there gender differences in quality of life and symptomatology between fibromyalgia patients? Am J Mens Health. 2012 Jul;6(4):314-9.
230. Castro-Sanchez AM, Mataran-Penarrocha GA, Lopez-Rodriguez MM, Lara-Palomo IC, Arendt-Nielsen L, Fernandez-de-Las-Penas C. Gender Differences in Pain Severity, Disability, Depression, and Widespread Pressure Pain Sensitivity in Patients with Fibromyalgia Syndrome Without Comorbid Conditions. Pain Med. 2012 Nov 21.
231. Mogil JS. Sex differences in pain and pain inhibition: multiple explanations of a controversial phenomenon. Nat Rev Neurosci. 2012 Nov;13(12):859-66.
232. Shulman JM, Chipendo P, Chibnik LB, Aubin C, Tran D, Keenan BT, et al. Functional screening of Alzheimer pathology genome-wide association signals in Drosophila. Am J Hum Genet. 2011 Feb 11;88(2):232-8.
233. Wang W, Yu JT, Zhang W, Cui WZ, Wu ZC, Zhang Q, et al. Genetic association of SLC2A14 polymorphism with Alzheimer's disease in a Han Chinese population. J Mol Neurosci. 2012 Jul;47(3):481-4.
234. Franke B, Vasquez AA, Johansson S, Hoogman M, Romanos J, Boreatti-Hummer A, et al. Multicenter analysis of the SLC6A3/DAT1 VNTR haplotype in persistent ADHD suggests differential involvement of the gene in childhood and persistent ADHD. Neuropsychopharmacology. 2010 Feb;35(3):656-64.
235. Minogue BM, Richardson SM, Zeef LA, Freemont AJ, Hoyland JA. Characterization of the human nucleus pulposus cell phenotype and evaluation of novel marker gene expression to define adult stem cell differentiation. Arthritis Rheum. 2010 Dec;62(12):3695-705.
236. Lewandowska U, Zelazowski M, Seta K, Byczewska M, Pluciennik E, Bednarek AK. WWOX, the tumour suppressor gene affected in multiple cancers. J Physiol Pharmacol. 2009 May;60 Suppl 1:47-56.
237. Sze CI, Su M, Pugazhenthi S, Jambal P, Hsu LJ, Heath J, et al. Down-regulation of WW domain-containing oxidoreductase induces Tau phosphorylation in vitro. A potential role in Alzheimer's disease. J Biol Chem. 2004 Jul 16;279(29):30498-506.
238. Li MY, Lai FJ, Hsu LJ, Lo CP, Cheng CL, Lin SR, et al. Dramatic co-activation of WWOX/WOX1 with CREB and NF-kappaB in delayed loss of small dorsal root ganglion neurons upon sciatic nerve transection in rats. PLoS One. 2009;4(11):e7820.
239. Takahashi H, Katayama K, Sohya K, Miyamoto H, Prasad T, Matsumoto Y, et al. Selective control of inhibitory synapse development by Slitrk3-PTPdelta trans-synaptic interaction. Nat Neurosci. 2012 Mar;15(3):389-98, S1-2.

240. Lise MF, Wong TP, Trinh A, Hines RM, Liu L, Kang R, et al. Involvement of myosin Vb in glutamate receptor trafficking. J Biol Chem. 2006 Feb 10;281(6):3669-78.
241. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, et al. Whole-genome association study of bipolar disorder. Mol Psychiatry. 2008 Jun;13(6):558-69.
242. Boucard AA, Chubykin AA, Comoletti D, Taylor P, Sudhof TC. A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins. Neuron. 2005 Oct 20;48(2):229-36.
243. Üçeyler N, Ann-Kathrin Kahn, Daniel Zeller, Susanne Kewenig, Sarah Kittel-Schneider, Department of Psychiatry, University of Karlheinz Reiners, Claudia Sommer., editor. Functional and morphological impairment of small nerve fibers in fibromaylgia syndrome
The 14th international association for the study of pain (IASP) world congress on pain; 2012; Milan Department of Neurology, University of Würzburg.
244. Saligan L, Majors B., Lukkahatai N., Hyung-Suk K., Walitt, B., editor. Gene expression differences between fibromyalgia and pain-free controls
14th international association for the study of pain (IASP) world congress on pain; 2012; Milan.

ABBREVIATIONS

5-HT	5-hydroxytryptamine (serotonin)
aCGH	Array comparative genomic hybridization
ACR	American College of Rheumatology
ACTH	Adrenocorticotrophic hormone
ANA	Antinuclear antibodies
BAC	Bacterial artificial chromosome
BAF	B allele Frequency
CBT	Cognitive Behavioral Therapy
CI	Confidence Interval
CFS	Chronic Fatigue Syndrome
CNS	Central Nervous System
CNV	Copy Number Variant
CSF	Cerebrospinal Fluid
CWP	Chronic Widespread Pain
DA	Dopamine
EULAR	European League Against Rheumatism
FIQ	Fibromyalgia Impact Questionnaire
eQTL	Expression quantitative trait loci
FISH	Fluorescence in situ hybridization
fMRI	Functional Magnetic Resonance Imaging
FM	Fibromyalgia
FoSTeS	Fork stalling and template switching
GABA	γ-amino-butyric acid
GO	Gene Ontology
GWAS	Genome Wide Association Study
HPA	Hypothalamic-Pituitary- Adrenal axis
HWE	Hardy Weinberg Equilibrium

IBD	Identity by descent
IBS	Identity by state
IL	Interleukin
INDEL	Insertion-Deletion
IPA	Ingenuity Systems Pathway analysis
Kb	Kilo base
L1	LINE-1
LD	Linkage disequilibrium
LB	Lysogeny broth
LNS	Laminin Like domain
LOH	Loss of heterozygosity
LRR	LogRratio: logged ratio of observed probe intensity to expected intensity
LRT	Likelihood Ratio Test
MAD	Mosaic Alteration Algorithm
MAF	Minimum Allele Frequency
MLPA	Multiplex Ligation Probe Amplification
MMBIR	Microhomology-mediated break-induced replication
MRS	Magnetic Resonance Spectroscopy
NAHR	Nonallelic homologuous recombination
NE	Noradrenalin
Ng	Nanogram
NHEJ	Nonhomologous end-joining
NGS	Next Generation Sequencing
NMDA	N-Methyl-D-aspartate
NRXN	Neurexin
NSAIDs	Non-steroidal anti-inflammatory drugs
ORF	Open reading frame

PCR	Polymerase Chain reaction
PET	Positron Emission Tomography
PI	Principal Investigator
PTBP2	Polypirimidine tract binding protein 2
QC	Quality control
RPM	Revolutions per minute
RT	Reverse Transcription
RTqPCR	Real Time Quantitative Polymerase Chain Reaction
SNP	Single Nucleotide Polymorphism
SNRIs	Serotonin Noradrenaline Reuptake Inhibitors
SPECT	Single-Photon Emission Computed Tomography
SSRIs	Serotonin Reuptake Inhibitors
TCA	Triciclic Antidepressants
TF	Transcription factor
TNF	Tumoral necrosis factor
UPD	Uniparental Disomy
WTCCC	Welcome trust Case Control Consortium

ANNEXES

SUPPLEMENTARY INFORMATION

Supplementary table1: Multinomial analysis of rs9381682 across FM clusters. Interaction pvalue=0.40

	Cl3 (N %)	**Cl2 (N %)**	**OR(95% CI)**	**Cl1 (N %)**	**OR(95% CI)**
AA	7 1.07	6 1.91	1	2 0.82	1
AG	117 17.89	69 21.97	0.69 (0.22 - 2.13)	51 20.9	1.52 (0.31 - 7.59)
GG	530 81.04	239 76.11	0.53 (0.17 - 1.58)	191 78.28	1.26 (0.26 - 6.12)
Trend	654	314	0.75 (0.56 - 1.01)	244	0.88 (0.63 - 1.22)
Pvalue Trend			0.0586		0.4401

Supplementary table2: Multinomial analysis of rs10821659 across FM clusters. Interaction pvalue=0.39

	Cl3 (N %)	**Cl2 (N %)**	**OR(95% CI)**	**Cl1 (N %)**	**OR(95% CI)**
AA	86 13.11	40 12.82	1	35 14.23	1
AG	312 47.56	161 51.6	1.11 (0.73 - 1.69)	108 43.9	0.85 (0.54 - 1.33)
GG	258 39.33	111 35.58	0.93 (0.6 - 1.43)	103 41.87	0.98 (0.62 - 1.55)
Trend	656	312	0.93 (0.76 - 1.13)	246	1.03 (0.83 - 1.28)
Pvalue Trend			0.4564		0.7769

Supplementary table3: Multinomial analysis of rs12556003 across FM clusters. Interaction pvalue=0.193

	Cl3 (N %)	**Cl2 (N %)**	**OR(95% CI)**	**Cl1 (N %)**	**OR(95% CI)**
CC	3 0.48	2 0.71	1	1 0.43	1
CT	78 12.54	34 12.01	0.65 (0.1 - 4.09)	39 16.88	1.5 (0.15 - 14.91)
TT	541 86.98	247 87.28	0.68 (0.11 - 4.12)	191 82.68	1.06 (0.11 - 10.25)
Trend	622	283	1.01 (0.68 - 1.5)	231	0.74 (0.5 - 1.1)
Pvalue Trend			0.9757		0.1368

Supplementary table4: Multinomial analysis of rs11127292 across FM clusters. Interaction pvalue=0.43

	Cl3 (N %)	**Cl2 (N %)**	**OR(95% CI)**	**Cl1 (N %)**	**OR(95% CI)**
AA	3 0.46	2 0.64	1	2 0.82	1
AG	91 13.85	57 18.15	0.94 (0.15 - 5.8)	37 15.1	0.61 (0.1 - 3.8)
GG	563 85.69	255 81.21	0.68 (0.11 - 4.09)	206 84.08	0.55 (0.09 - 3.31)
Trend	657	314	0.74 (0.53 - 1.03)	245	0.87 (0.59 - 1.28)
Pvalue Trend			0.078		0.4815

Supplementary table 5: aCGH results corresponding to the *GNG1* region detected by PennCNV algorithm

PROBE	CHR	START	END	GENE	FM vs C	FM vs C_DS	FM_FC vs C	FM_FC vs C_DS	FM_EARLY vs C	FM_EARLY vs C_DS
A_16_P18028824	chr7	93165160	93165219	GNG1	0.51826024	-0.27536365	0.6019959	-0.27371195	0.43737462	-0.069126844
A_16_P01756731	chr7	93166102	93166161	GNG1	0.4588328	-0.22160305	0.3580741	-0.37493905	0.26458326	-0.2593865
A_16_P01756732	chr7	93166604	93166663	GNG1	0.48650372	-0.47130048	0.45676237	-0.47091928	0.37744698	-0.24169426
A_16_P38112884	chr7	93168559	93168618	GNG1	0.55047625	-0.4335387	0.4814904	-0.25831863	0.25360194	-0.31467307

Supplementary table 6: Rare large CNV events detected by PennCNV in FM samples

Chr	SNP_CNV_posini	SNP_CNV_posend	nSNP	Size	Sample
chr10	1074360	1861976	439	659364	sample.0034_01
chr11	30757347	30860076	39	102730	sample.0064_04
chr11	99792469	99896260	34	103792	sample.0027_02
chr12	23197970	23653255	157	455286	sample.0122_01
chr12	27302867	27667171	201	364305	sample.0029_04
chr13	54041755	54371733	88	329979	sample.0062_05
chr16	35026333	35143083	17	116751	sample.0122_01
chr16	76545661	77220411	420	663690	sample.0005_04
chr17	995559	1103046	39	107488	sample.0515_01
chr18	1036304	1192623	48	156320	sample.0509_01
chr19	24253122	24365871	21	112750	sample.0118_01
chr2	49069064	49742180	304	673117	sample.0515_01
chr22	19490584	19795780	212	305197	sample.0398_01
chr22	47974831	48370692	332	395862	sample.0051_04
chr3	86219308	86344171	45	124864	sample.0509_01
chr4	3873060	3993797	12	120738	sample.0518_01
chr4	53831985	54026463	60	194479	sample.0003_04
chr4	188692088	190189143	563	1320903	sample.0398_01
chr4	189865993	190154509	115	288517	sample.0082_02
chr5	45948724	46435031	68	441249	sample.0082_02
chr5	46174214	46435031	38	260818	sample.0067_05
chr6	8537674	8662286	28	124613	sample.0051_04
chr6	57661845	58138983	81	347513	sample.0476_01
chr7	69312391	69449148	30	136758	sample.0008_04
chr7	84404435	84734722	93	330288	sample.0118_01
chr9	128946715	129322660	134	340251	sample.0046_01

Supplementary tables 7-11: Single probe 400k-CNV aCGH results, as detected by ADM2 algorithm (except *MYO5B* detected by ADM1). Genomic location is based on build Hg 18. DS: Dye swap

Supplementary table 7: Single probe 400k-CNV aCGH results in *GALNTL6* region

PROBE	CHR	START	END	GENE	FM vs C	FM vs C_DS	FM_FC vs C	FM_FC vs C_DS	FM_EARLY vs C	FM_EARLY vs C_DS
A_16_P37014308	chr4	173660973	173661032	GALNTL6	0.08529253	-0.19402722	-0.0986076	-0.045194484	-0.096624814	-0.0307554
A_16_P16956845	chr4	173661281	173661340	GALNTL6	-0.199432	-0.20926546	0.09380534	-0.21207479	-0.17027839	-0.3254448
A_16_P37014311	chr4	173661991	173662050	GALNTL6	0.32443044	0.0037429724	0.351901	-0.576656	-1.4055429	1.5115635
A_16_P16956850	chr4	173662633	173662692	GALNTL6	-0.5817815	0.18040426	0.5002587	-0.36148244	-0.6886951	0.9229039
A_16_P37014314	chr4	173663054	173663113	GALNTL6	0.25556508	0.40459538	0.63884676	-0.39548576	-0.95690316	1.1741178
A_16_P16956854	chr4	173663563	173663619	GALNTL6	0.36496556	-0.1120656	0.46575466	-1.1264752	-0.99932873	1.4243106
A_16_P16956855	chr4	173665579	173665638	GALNTL6	-0.6296597	0.3143576	0.3123617	-0.6497465	-0.93122363	1.1777378
A_16_P16956857	chr4	173666072	173666131	GALNTL6	-0.2220633	0.09262541	0.7364325	-0.35405615	-1.1265422	1.0924625
A_16_P16956857	chr4	173666072	173666131	GALNTL6	0.07001504	0.2788554	0.8776765	-0.49385715	-1.0031898	1.3683885
A_16_P16956857	chr4	173666072	173666131	GALNTL6	-0.1100783	0.72634137	0.85614103	-0.7350299	-0.97015876	1.1772723
A_16_P16956858	chr4	173672958	173673017	GALNTL6	0.26191145	0.2547935	0.13965371	0.08119407	-0.0016348136	0.1393855

Supplementary table 8: Single probe 400K aCGH results in WWOX_INDEL region.

PROBE	CHR	START	END	GENE	FM vs C	FM vs C_DS	FM_FC vs C	FM_FC vs C_DS	FM_EARLY vs C	FM_EARLY C_DS
A_16_P03181915	chr16	76928627	76928686	WWOX	0.28158298	0.19629586	0.18516344	0.06635989	0.2887281	0.32507694
A_16_P40692347	chr16	76929598	76929657	WWOX	-0.12481168	-0.065518945	-0.28566095	0.041399505	-0.6650886	0.7119267
A_16_P40692349	chr16	76930251	76930310	WWOX	-0.15443718	0.42234224	0.36055475	-0.1266613	-0.19667232	0.9261816
A_16_P20529478	chr16	76931247	76931301	WWOX	-0.19234885	0.21160926	-0.11883988	-0.009393483	-0.51001096	0.5417443
A_16_P20529479	chr16	76932880	76932939	WWOX	0.06862109	-0.06111997	0.23052749	0.21321464	-0.294821	0.3995713
A_16_P20529481	chr16	76933534	76933593	WWOX	-0.056140203	0.3544674	0.19905257	0.20053083	-0.48757106	0.62653726
A_16_P03181924	chr16	76934755	76934812	WWOX	-0.29839915	-0.016645182	-0.15770896	-0.02818546	-0.68682563	0.58352226
A_18_P12522425	chr16	76935880	76935939	WWOX	-0.34534696	0.07977769	-0.38944054	-0.008749906	-0.60195565	0.30090234
A_18_P12525015	chr16	76937156	76937215	WWOX	-0.20381327	0.4568865	-0.24411707	0.08652396	-0.22669499	0.38370594
A_16_P03181929	chr16	76938723	76938782	WWOX	-0.07459494	0.22284941	0.059813984	0.1010824	-0.607083	0.62241185
A_16_P20529495	chr16	76940218	76940277	WWOX	-0.02925669	0.28461063	0.07680479	0.15250087	-0.057597097	0.5593411
A_16_P20529500	chr16	76941574	76941633	WWOX	-0.15378556	0.11600846	-0.15076958	0.1771808	-0.66758573	0.7325019
A_16_P03181935	chr16	76942679	76942738	WWOX	-0.032101676	-0.0053606257	-0.004461236	0.5783381	-0.14082606	-0.08295628

Supplementary table 9: aCGH 400K array results in PTPRD_INDEL region.

PROBE	CHR	START	END	GENE	FM Vs C	FM vs C_DS	FM_FC vs C	FM_FC vs C_DS	FM_EARLY vs C	FM_EARLY vs C_DS
A_18_P16745535	chr9	10394237	10394296	PTPRD	0.13007563	0.2824293	0.082413584	0.30320624	0.16762814	0.2109006
A_16_P18550515	chr9	10394569	10394628	PTPRD	-2.3631587	2.5230289	-0.5547822	0.30764696	0.17761183	-0.90652484
A_18_P16747922	chr9	10394791	10394850	PTPRD	-1.874148	1.8231975	0.061165757	0.7986101	-0.09189645	-0.12869057
A_16_P38647342	chr9	10395062	10395121	PTPRD	-2.240932	1.1376731	-0.40710583	0.13865355	-0.07161968	-0.26055107
A_16_P18550517	chr9	10395198	10395257	PTPRD	0.28663224	0.06954047	0.34778836	0.23918259	0.21191937	0.11030927

Supplementary table 10: aCGH 400K array results in MYO5B_INDEL region.

PROBE	CHR	START	END	GENE	FM Vs C	FM vs C_DS	FM_FC vs C	FM_FC vs C_DS	FM_EARLY vs C	FM_EARLY vs C_DS
A_16_P20858028	chr18	45949290	45949349	MYO5B	0.5636287	-0.24847256	0.010896813	0.23036605	0.15394087	0.14406267
A_16_P20858030	chr18	45950437	45950496	MYO5B	0.23850645	-0.3400926	-0.3137207	0.023292942	-0.08800473	-0.22030301
A_16_P41036292	chr18	45950746	45950805	MYO5B	0.47462562	-0.42723644	-0.020762814	0.0070652533	0.29897565	0.037801277
A_16_P20858033	chr18	45951188	45951247	MYO5B	0.37659758	-0.51448077	-0.5253001	0.04363984	-0.07216798	-0.18815546
A_18_P12878496	chr18	45951578	45951637	MYO5B	0.18794645	-0.48702845	-0.32932615	-0.0467543	-0.073677145	-0.114734754
A_18_P12881286	chr18	45951913	45951972	MYO5B	0.3969381	-0.21023215	-0.02321451	0.25585207	0.1881237	0.1372318
A_16_P20858037	chr18	45952242	45952301	MYO5B	0.36879095	-0.5066229	-0.20882441	-0.06194123	-0.0381851	-0.010018041

Supplementary table 11: aCGH 400K array results in NRXN3_INDEL region.

PROBE	START	END	GENE	FM vs C	FM vs C_DS	FM_FC vs C	FM_FC vs C_DS	FM_EARLY vs C	FM_EARLY vs C_DS
A_16_P02953411	79176037	79176096	NRXN3	-0,4795623	0,30427477	-0,49750778	0,5259584	-0,46278116	0,6449666
A_18_P12071355	79176628	79176687	NRXN3	-0,2441538	0,15020816	-0,22433914	0,14568233	-0,7520391	0,55166185
A_16_P20110531	79176951	79177010	NRXN3	-0,4254829	0,03661296	-0,52878815	0,2859342	-0,67901367	0,52906233
A_16_P40255621	79177710	79177758	NRXN3	0,02314774	0,37165424	-0,3042572	0,45597467	-0,2656493	0,49884096
A_16_P40255622	79178939	79178994	NRXN3	-0,0351931	-0,061897863	-0,05913251	0,66612834	-0,06728733	0,87978923
A_18_P12070550	79179266	79179325	NRXN3	-0,1666244	-0,090419054	0,09677741	0,44182938	-0,40539366	0,42340028
A_16_P20110538	79180204	79180263	NRXN3	-0,5853588	0,22122027	-0,6862277	0,1692285	-0,8400456	0,4334101
A_18_P12073341	79180648	79180707	NRXN3	-0,2555005	0,14799342	-0,67632663	0,38491964	-0,7174981	0,3723865
A_16_P20110541	79181083	79181140	NRXN3	-0,3968149	0,21465287	-0,34479716	0,43062505	-0,64381367	0,50692374
A_16_P02953420	79181946	79182005	NRXN3	-0,5366972	0,26065788	-0,5997461	0,2993942	-0,7734734	0,4831312
A_16_P20110544	79182226	79182285	NRXN3	-0,2520727	0,18726353	-0,32880846	0,49681765	-0,465008	0,58169186
A_18_P12071299	79182450	79182509	NRXN3	-0,6277766	0,24885204	-0,67979443	0,47922066	-0,7775375	0,55087715
A_16_P20110546	79182711	79182764	NRXN3	-0,342092	0,5742588	-0,58788574	0,43994248	-0,52316344	0,5121261
A_16_P02953422	79182878	79182937	NRXN3	-0,498412	0,1590231	-0,5726793	0,4331416	-0,6051793	0,6970408
A_18_P12071370	79183154	79183213	NRXN3	-0,3276374	0,17555709	-0,4320735	0,5652103	-0,54288846	0,6509929
A_16_P40255638	79183384	79183443	NRXN3	-0,0560496	0,16019842	-0,36573297	0,4973633	-0,3526407	0,7155004
A_18_P12067956	79183688	79183747	NRXN3	-0,1216632	0,60281694	-0,24096733	0,70596945	-0,42653054	0,95046717
A_18_P12069437	79184222	79184281	NRXN3	-0,1939572	0,26349366	-0,44449794	0,5588709	-0,5027476	0,66544336

aCGH results in NRXN3_INDEL genomic region

Supplementary table 12: Multinomial analysis of NRXN3_DEL across FM clusters. Interaction pvalue=0.046

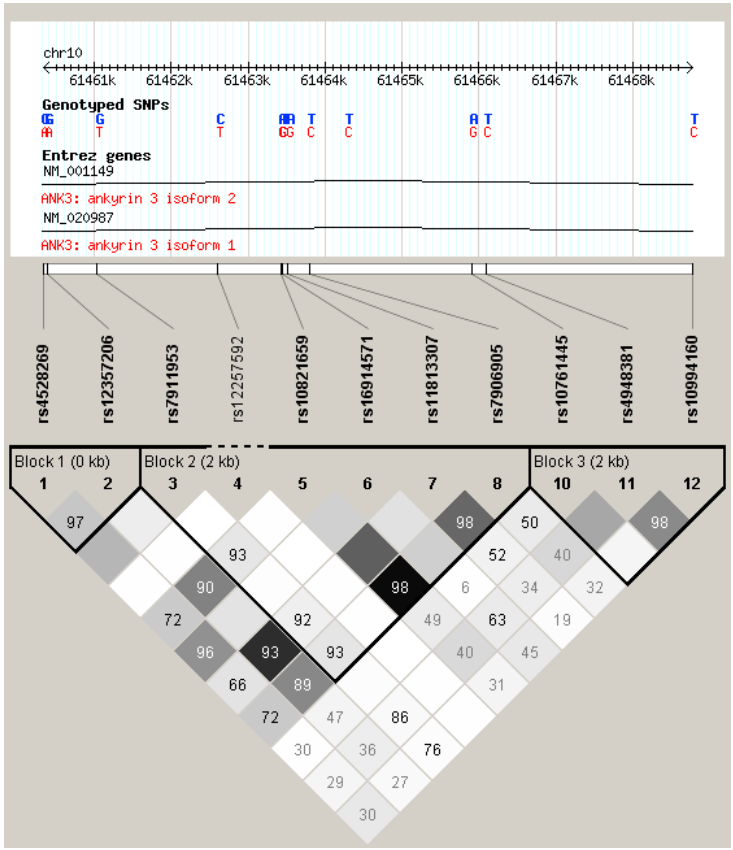
	Cl3 (N %)	**Cl2 (N %)**	**OR(95% CI)**	**Cl1 (N %)**	**OR(95% CI)**
DelDel	104 14.25	49 14.37	1	50 18.05	1
DelNodel	357 48.9	149 43.7	0.89 (0.6 - 1.31)	132 47.65	0.77 (0.52 - 1.14)
NodelNodel	269 36.85	143 41.94	1.13 (0.76 - 1.68)	95 34.3	0.73 (0.49 - 1.11)
Trend	730	341	1.11 (0.92 - 1.34)	277	0.88 (0.72 - 1.07)
Pvalue Trend			0.27		0.193

Supplementary table 13:*NRXN3* Veracode association analysis results

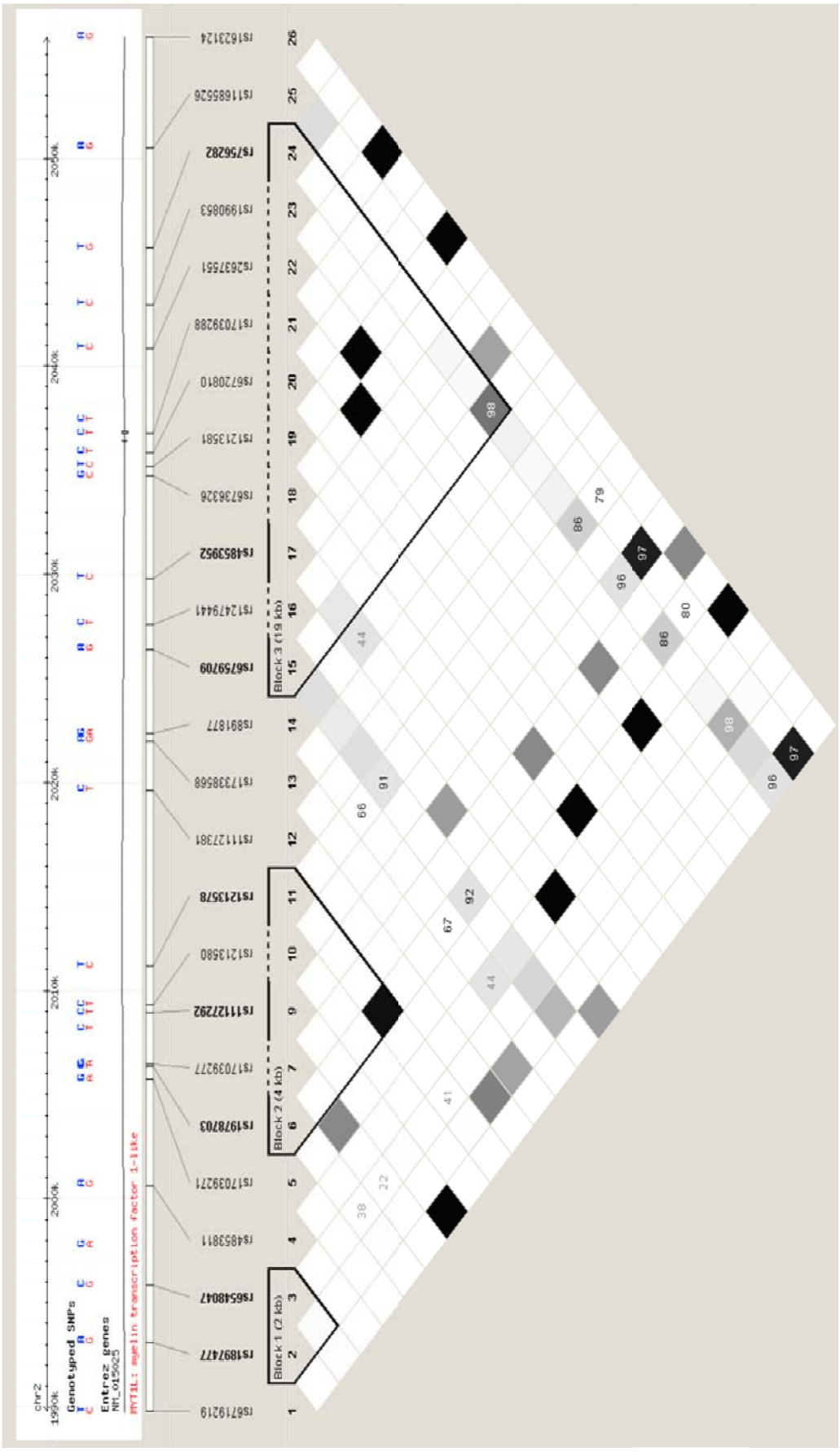
CHR	SNP	POSITION	A1	F_A	F_U	A2	CHISQ	P	OR	SE	L95	U95
14	rs10146997	79014914	G	0.2	0.1688	A	5.219	0.02234	1.231	0.09093	1.03	1.471
14	rs2293839	79387212	A	0.3052	0.3341	G	3.109	0.07787	0.8752	0.0756	0.7547	1.015
14	rs1159039	79361345	G	0.3542	0.3252	A	3.024	0.08204	1.138	0.07436	0.9837	1.317
14	rs8021767	79263758	A	0.1897	0.1671	G	2.821	0.09306	1.167	0.09203	0.9744	1.398
14	rs11848580	79038540	G	0.4039	0.3757	A	2.685	0.1013	1.126	0.07221	0.977	1.297
14	rs1030127	79231192	G	0.1355	0.1174	A	2.395	0.1217	1.179	0.1062	0.957	1.451
14	rs8018724	79318557	A	0.3792	0.4037	C	2.028	0.1544	0.9023	0.07221	0.7832	1.039
14	rs7153625	79119014	A	0.1716	0.1532	G	2.021	0.1551	1.145	0.09557	0.9498	1.381
14	rs178377	79247778	G	0.1492	0.1326	A	1.849	0.1739	1.148	0.1013	0.9409	1.4
14	rs12323794	78631351	A	0.3729	0.3542	G	1.224	0.2686	1.084	0.07322	0.9394	1.252
14	rs221415	79115621	G	0.1768	0.1917	A	1.188	0.2757	0.9056	0.09097	0.7577	1.082
14	rs10145867	79339926	A	0.08592	0.07551	G	1.181	0.2772	1.151	0.1294	0.8931	1.483
14	rs932265	79360177	A	0.1406	0.1522	G	0.8605	0.3536	0.9117	0.09973	0.7498	1.108
14	rs6574495	78669026	A	0.4019	0.387	G	0.7495	0.3866	1.064	0.07204	0.9242	1.226
14	rs11627269	78784324	A	0.3252	0.3115	G	0.6906	0.4059	1.065	0.07558	0.9182	1.235
14	rs2202175	78625403	C	0.3929	0.3813	A	0.4598	0.4977	1.05	0.07235	0.9114	1.21
14	rs221473	79174368	T	0.1877	0.1968	A	0.4249	0.5145	0.9434	0.0894	0.7918	1.124
14	rs2196443	78946968	C	0.2052	0.1962	A	0.4046	0.5247	1.057	0.0879	0.8901	1.256
14	rs31431	78510788	A	0.1161	0.1095	C	0.3518	0.5531	1.068	0.1113	0.8589	1.329
14	rs1424850	79067756	A	0.2888	0.28	T	0.3024	0.5824	1.044	0.07807	0.8958	1.216
14	rs17108457	78520530	T	0.1111	0.1058	A	0.233	0.6293	1.056	0.1132	0.846	1.319
14	rs760288	79376682	G	0.3097	0.3175	A	0.228	0.633	0.9644	0.07591	0.8311	1.119
14	rs8019381	79390335	A	0.1512	0.157	G	0.2075	0.6488	0.9565	0.09761	0.79	1.158
14	rs221449	79193929	A	0.2607	0.2676	G	0.1927	0.6607	0.9653	0.08033	0.8247	1.13
14	rs2543576	79141981	A	0.4465	0.4518	G	0.09477	0.7582	0.9784	0.07079	0.8517	1.124
14	rs8022725	79129222	C	0.3606	0.3555	G	0.09186	0.7618	1.023	0.07344	0.8854	1.181
14	rs9323679	79421543	G	0.1432	0.1457	A	0.03866	0.8441	0.9805	0.1002	0.8057	1.193
14	rs2202167	78569377	A	0.3769	0.3744	C	0.0221	0.8818	1.011	0.07279	0.8765	1.166
14	rs17836266	79203779	G	0.1539	0.1521	A	0.0212	0.8842	1.014	0.09803	0.8371	1.229
14	rs994010	79419355	G	0.2387	0.2408	A	0.01905	0.8902	0.9887	0.08248	0.8411	1.162
14	rs5014481	79282614	A	0.4232	0.4256	C	0.01853	0.8917	0.9903	0.07126	0.8613	1.139
14	rs10130593	79211814	G	0.3271	0.3288	C	0.01031	0.9191	0.9924	0.075	0.8567	1.15
14	rs2215840	79339820	G	0.3529	0.352	A	0.003132	0.9554	1.004	0.0737	0.8691	1.16
14	rs221497	79158232	A	0.1065	0.106	G	0.002117	0.9633	1.005	0.1143	0.8035	1.258

Supplementary table 13 :FM Clusters proportions among GWAS and replication subsets

	Cluster1	**Cluster2**	**Cluster3**
GWAS (291)	66(22.7%)	29(10%)	196 (67.3%)
Replication(1135)	217(19.1%)	328(28.9%)	590 (52%)
Joint (1398)	283(20.2%)	357(25.6%)	758 (54.2)

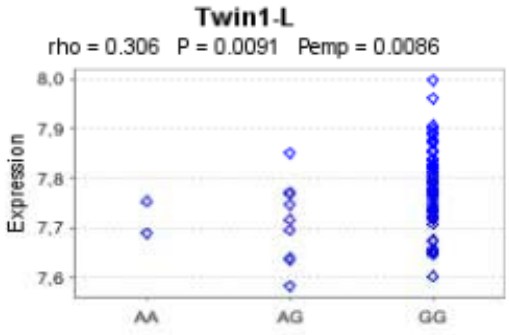


Suppementary figure 1: Linkage desequilibrium plot of rs10821659 D' LD score is shown inside the boxes.

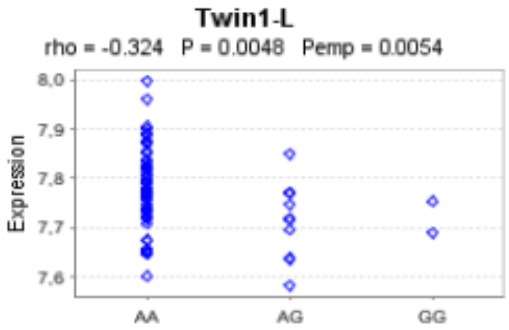


Supplementary figure 2: Linkage desequilibrium plot of rs11127292 D' LD score is shown inside the boxes.

a) rs11127292 genotypes and *SNTG2* expression (as assessed by one probe) in twins' lymphoblastyoid cell lines



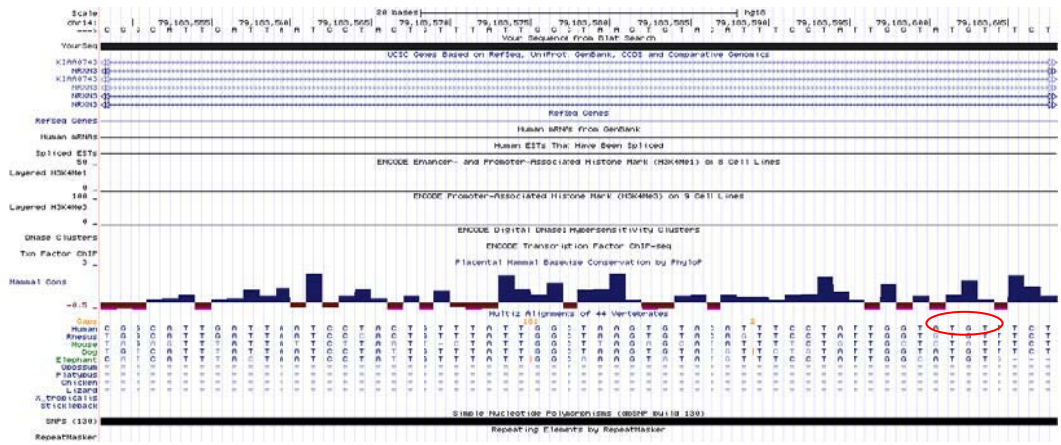
b) rs11685526 genotypes and *SNTG2* expression (as assessed by one probe) in twins' lymphoblastyoid cell lines



Supplementary figure 3: SNP-probe association plots for rs11127292 and rs11685526 .Expression levels for one expression array probe are plotted against SNP genotypes.

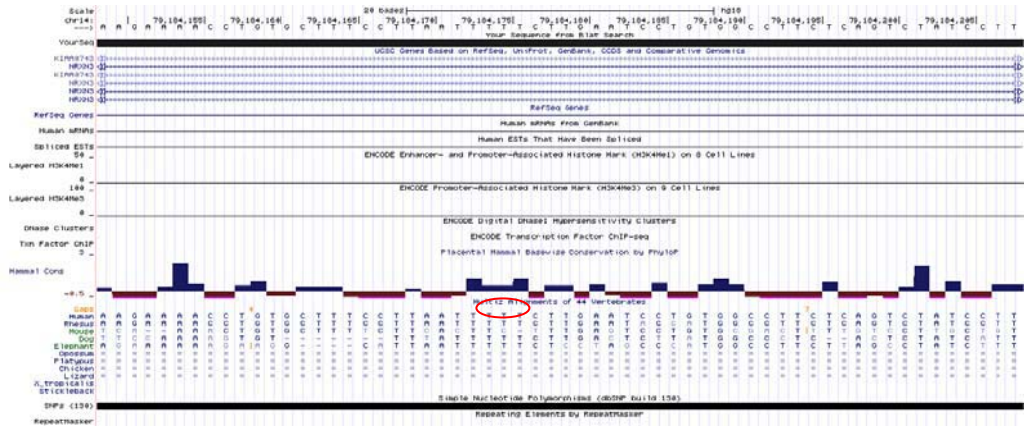
ATGTTT Splicing enhancer

```
1   CGGCATTGATTAATCCTACTGTTTATTGGCTAAGTGTACATTTCCTATTGGTATGTTTCT 8160
2   ------------------------------------------------------ATGTTT-- 6
                                                           ******
```



TTT Splicing silencer

```
1   AAGAAAACCTGTGCTTTCCTTAATTTTTCTTGAATCCTGTGGCCTTCTCAGTCTATCCTT 8760
5   -------------------------TTT-------------------------------- 3
                              ***
```



Supplementary figure 4: Splicing factors binding sites detected in NRXN3_DEL.Matchging sequence appears in yellow. Red circle represents the consensus sequence in UCSC genome browser image including level of conservation among species.

GLOSSARY

Paraneoplastic syndrome: tumours of the lung are particularly likely to produce diffuse neurologic syndromes, through the formation of antibodies against central nervous system structures. The relatively sudden onset (in older patients) of fatigue, anorexia and weight loss has been inappropriately designated as FM.

Polymialgia rheumatica: Inflammatory disorder that causes muscle pain and stiffness, mainly in shoulders, neck, upper arms and hips. Since it mainly affects people over 65 years old, blood test shows an elevation of erythrocyte sedimentation rate and anaemia and it has a very good response to glucocorticoid therapy, it can be easily differentiated from FM.

Polymiositis: Inflammatory disease of the muscle characterized by muscle weakness (out of proportion of pain). Abnormal high levels of creatine kinase and presence of specific autoantibodies (Anti-synthetasa) in blood test as well the presence of pathologic specific findings in electromyography and muscle biopsy, enable a right diagnosis.

Propioceptor: A sensory receptor, found chiefly in muscles, tendons, joints, and the inner ear, that detects the motion or position of the body or a limb by responding to stimuli arising within the organism.

Reflex sympathetic dystrophy: Pain, swelling, and vasomotor dysfunction of an extremity of variable course and unknown cause. This condition is often the result of trauma or surgery. It is also known as complex regional pain syndrome (CRPS) and may occur spontaneously. Physical findings (glazed and swollen skin and vasomotor changes) can rapidly separate it from FM, but when these signs are minimal, the differential diagnosis is necessary.

Seronegative spondyloarthropathy: Inflammatory rheumatic disease which mainly affects axial skeleton. Autoantibodies in serum are negative and for this reason, in initial steps of the disease in which bones erosions are not visible in X-ray, it can be misdiagnosed as FM. An accurate anamnesis (asking for presence of psoriasis, uveitis as well as characteristic features of the disease such as heels pain) as well as other more sensitive imaging tests (such as Magnetic Resonance Imaging) will help in the diagnosis.

Somatoform disorders: Group of disorders that have physical symptoms that are not explained by physical alterations.

Systemic lupus erythematosus (SLE): Autoimmune disorder characterized by the presence of autoantibodies against DNA and nucleus. Clinically it presents with a wide spectrum of symptoms affecting different organs (skin, joints, muscles, kidneys, lungs, brain) from very mild to a very severe disease. When only the muskeloskeletal system is affected, since the most frequent symptoms are arthralgias (pain in the joints without swollening) and it mainly affects young women, FM should be discarded. Blood test with rheumatologic serologies will aid to the differencial diagnosis. However, in the case of SLE, secondary FM is quite frequent.

SCALES

As reviewed in this thesis, clinical scales are currently used in fibromyalgia management. We have included the questionnaires of three of the most widely used:

-SF-36 which is a general scale to evaluate health state

-FIQ: which is used to assess FM severity

-PSQI: which is used to evaluate sleep disturbances associated to the disease

SF-36 QUESTIONNAIRE

Name:_____ **Ref. Dr:**_____ **Date:** _____

ID#: _____ **Age:** _____ **Gender: M / F**

Please answer the 36 questions of the **Health Survey** completely, honestly, and without interruptions.

GENERAL HEALTH:
In general, would you say your health is:

◯ Excellent ◯ Very Good ◯ Good ◯ Fair ◯ Poor

Compared to one year ago, how would you rate your health in general now?

◯ Much better now than one year ago
◯ Somewhat better now than one year ago
◯ About the same
◯ Somewhat worse now than one year ago
◯ Much worse than one year ago

LIMITATIONS OF ACTIVITIES:
The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports.
◯ Yes, Limited a lot ◯ Yes, Limited a Little ◯ No, Not Limited at all

Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf
◯ Yes, Limited a Lot ◯ Yes, Limited a Little ◯ No, Not Limited at all

Lifting or carrying groceries
◯ Yes, Limited a Lot ◯ Yes, Limited a Little ◯ No, Not Limited at all

Climbing several flights of stairs
◯ Yes, Limited a Lot ◯ Yes, Limited a Little ◯ No, Not Limited at all

Climbing one flight of stairs
◯ Yes, Limited a Lot ◯ Yes, Limited a Little ◯ No, Not Limited at all

Bending, kneeling, or stooping
◯ Yes, Limited a Lot ◯ Yes, Limited a Little ◯ No, Not Limited at all

Walking more than a mile
◯ Yes, Limited a Lot ◯ Yes, Limited a Little ◯ No, Not Limited at all

Walking several blocks
◯ Yes, Limited a Lot ◯ Yes, Limited a Little ◯ No, Not Limited at all

Walking one block
◯ Yes, Limited a Lot ◯ Yes, Limited a Little ◯ No, Not Limited at all

Bathing or dressing yourself

◯Yes, Limited a Lot ◯Yes, Limited a Little ◯No, Not Limited at all

PHYSICAL HEALTH PROBLEMS:

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

Cut down the amount of time you spent on work or other activities

◯Yes ◯No

Accomplished less than you would like

◯Yes ◯No

Were limited in the kind of work or other activities

◯Yes ◯No

Had difficulty performing the work or other activities (for example, it took extra effort)

◯Yes ◯No

EMOTIONAL HEALTH PROBLEMS:

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

Cut down the amount of time you spent on work or other activities

◯Yes ◯No

Accomplished less than you would like

◯Yes ◯No

Didn't do work or other activities as carefully as usual

◯Yes ◯No

SOCIAL ACTIVITIES:

Emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

◯Not at all ◯Slightly ◯Moderately ◯Severe ◯Very Severe

PAIN:

How much bodily pain have you had during the past 4 weeks?

◯None ◯Very Mild ◯Mild ◯Moderate ◯Severe ◯Very Severe

During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

◯Not at all ◯A little bit ◯Moderately ◯Quite a bit ◯Extremely

ENERGY AND EMOTIONS:
These questions are about how you feel and how things have been with you during the last 4 weeks. For each question, please give the answer that comes closest to the way you have been feeling.

Did you feel full of pep?
- ○ All of the time
- ○ Most of the time
- ○ A good Bit of the Time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

Have you been a very nervous person?
- ○ All of the time
- ○ Most of the time
- ○ A good Bit of the Time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

Have you felt so down in the dumps that nothing could cheer you up?
- ○ All of the time
- ○ Most of the time
- ○ A good Bit of the Time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

Have you felt calm and peaceful?
- ○ All of the time
- ○ Most of the time
- ○ A good Bit of the Time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

Did you have a lot of energy?
- ○ All of the time
- ○ Most of the time
- ○ A good Bit of the Time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

Have you felt downhearted and blue?

- ○ All of the time
- ○ Most of the time
- ○ A good Bit of the Time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

Did you feel worn out?

- ○ All of the time
- ○ Most of the time
- ○ A good Bit of the Time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

Have you been a happy person?

- ○ All of the time
- ○ Most of the time
- ○ A good Bit of the Time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

Did you feel tired?

- ○ All of the time
- ○ Most of the time
- ○ A good Bit of the Time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

SOCIAL ACTIVITIES:
During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?

- ○ All of the time
- ○ Most of the time
- ○ Some of the time
- ○ A little bit of the time
- ○ None of the Time

GENERAL HEALTH:
How true or false is each of the following statements for you?

I seem to get sick a little easier than other people
Definitely true | Mostly true | Don't know | Mostly false | Definitely false

I am as healthy as anybody I know
Definitely true | Mostly true | Don't know | Mostly false | Definitely false

I expect my health to get worse
Definitely true | Mostly true | Don't know | Mostly false | Definitely false

My health is excellent
Definitely true | Mostly true | Don't know | Mostly false | Definitely false

FIBROMYALGIA IMPACT QUESTIONNAIRE (FIQ)

Last name: **First name:** **Age :** **Todays date :**

Duration of FM symptoms (years) : **Years since diagnosis of FM :**

Question 1
Directions: For questions "a" through "k", please check the number that best describes how you did overall for the *past week*. If you don't normally do something that is asked, place an 'X' in the 'Not Applicable' box.

Were you able to:	Always	Most	Occasionally	Never	Not Applicable
a. Do shopping?	☐0	☐1	☐2	☐3	☐4
b. Do laundry with a washer and dryer?	☐0	☐1	☐2	☐3	☐4
c. Prepare meals?	☐0	☐1	☐2	☐3	☐4
d. Wash dishes / cooking utensils by hand?	☐0	☐1	☐2	☐3	☐4
e. Vacuum a rug?	☐0	☐1	☐2	☐3	☐4
f. Make beds?	☐0	☐1	☐2	☐3	☐4
g. Walk several blocks?	☐0	☐1	☐2	☐3	☐4
h. Visit friends or relatives?	☐0	☐1	☐2	☐3	☐4
i. Do yard work?	☐0	☐1	☐2	☐3	☐4
j. Drive a car?	☐0	☐1	☐2	☐3	☐4
k. Climb stairs?	☐0	☐1	☐2	☐3	☐4
Sub-total scores *(for internal use only)*	☐	☐	☐	☐	☐
Total score *(for internal use only)*	☐				

2. Of the 7 days in the past week, how many days did you feel good?

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7

Score ☐

3. How many days last week did you miss work, including housework, because of fibromyalgia?

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7

Score ☐

(Continued)

(Continuation)

Directions: For the remaining items, mark the point on the line that best indicates how you felt overall for the past week.

4. When you worked how much did pain or other symptoms of your fibromyalgia interfere with your ability to do your work, including housework?

No problem with work — Great difficulty with work

5. How bad has your pain been?

No pain — Very severe pain

6. How tired have you been?

No tiredness — Very tired

7. How have you felt when you get up in the morning?

Awoke well rested — Awoke very tired

8. How bad has your stiffness been?

No stiffness — Very stiff

9. How nervous or anxious have you felt?

Not anxious — Very anxious

10. How depressed or blue have you felt?

Not depressed — Very depressed

(for internal use only)

Score

Score

Score

Score

Score

Score

Score

Sub-total

FIQ TOTAL

Subject's Initials_____ID#_____Date_____Time_____AM PM

PITTSBURGH SLEEP QUALITY INDEX

INSTRUCTIONS:
The following questions relate to your usual sleep habits during the past month only. Your answers should indicate the most accurate reply for the majority of days and nights in the past month. Please answer all questions.

1. During the past month, what time have you usually gone to bed at night?

   BED TIME _____

2. During the past month, how long (in minutes) has it usually taken you to fall asleep each night?

   NUMBER OF MINUTES _____

3. During the past month, what time have you usually gotten up in the morning?

   GETTING UP TIME _____

4. During the past month, how many hours of actual sleep did you get at night? (This may be different than the number of hours you spent in bed.)

   HOURS OF SLEEP PER NIGHT _____

For each of the remaining questions, check the one best response. Please answer all questions.

5. During the past month, how often have you had trouble sleeping because you . . .

a) Cannot get to sleep within 30 minutes

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

b) Wake up in the middle of the night or early morning

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

c) Have to get up to use the bathroom

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

d) Cannot breathe comfortably

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

e) Cough or snore loudly

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

f) Feel too cold

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

g) Feel too hot

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

h) Had bad dreams

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

i) Have pain

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

j) Other reason(s), please describe_____

_____

How often during the past month have you had trouble sleeping because of this?

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

6. During the past month, how would you rate your sleep quality overall?

Very good _____

Fairly good _____

Fairly bad _____

Very bad _____

7. During the past month, how often have you taken medicine to help you sleep (prescribed or "over the counter")?

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

8. During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

9. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?

No problem at all _____

Only a very slight problem _____

Somewhat of a problem _____

A very big problem _____

10. Do you have a bed partner or room mate?

No bed partner or room mate _____

Partner/room mate in other room _____

Partner in same room, but not same bed _____

Partner in same bed _____

If you have a room mate or bed partner, ask him/her how often in the past month you have had . . .

a) Loud snoring

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

b) Long pauses between breaths while asleep

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

c) Legs twitching or jerking while you sleep

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

d) Episodes of disorientation or confusion during sleep

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

e) Other restlessness while you sleep; please describe_____

_____

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ: Psychiatry Research, 28:193-213, 1989.

PUBLICATIONS

Work from this thesis will give rise to the following publications:

Cluster analysis of clinical data identifies fibromyalgia subgroups.
Docampo E, Escaramis G,Rabionet R, Carbonell J, Rivera J, Alegre J,Vidal J, Estivill X, Collado A.
The Clinical Journal of Pain. *Under Review*

Copy number variants analysis in fibromyalgia
Docampo E, Rabionet R, Escaramis G, Villatoro S, Puig A, Collado A, Carbonell J, Rivera J, Alegre J, Vidal J, Estivill X. *Under preparation*

Publications related to this thesis:

Screening for the presence of FMR1 premutation alleles in women with fibromyalgia.
Rodriguez-Revenga L, Madrigal I, Blanch-Rubió J, Elurbe D, Docampo E, Collado A, Vidal J, Carbonell J, Estivill X, Mila M.
Gene. 2012 Oct 27. doi:pii: S0378-1119(12)01280-2. 10.1016/j.gene.2012.10.016. [Epub ahead of print] PMID:23111161

Association of Neurexin 3 polymorphims with nicotine addiction.
Docampo E, Ribasés M, Gratacòs M, Bruguera E, Cabezas C, Nieva G, Puente D, Argimon-Pallàs JM, Casas M, Rabionet R and Estivill X.
Genes Brain Behav. 2012 Jun 21;9999(999A). doi:10.1111/j.1601-183X.2012.00815.x. [Epub ahead of print].PMID:22716474

Additional publications during thesis period:

Deletion of LCE3C and LCE3B is a susceptibility factor for psoriatic arthritis: a study in Spanish and Italian populations and meta-analysis.
Docampo E, Giardina E, Riveira-Muñoz E, de Cid R, Escaramís G, Perricone C, Fernández-Sueiro JL, Maymó J, González-Gay MA, Blanco FJ, Hüffmeier U, Lisbona MP, Martín J, Carracedo A, Reis A, Rabionet R, Novelli G, Estivill X.
Arthritis Rheum. 2011 Jul;63(7):1860-5. doi: 10.1002/art.30340. PMID:21400479

Deletion of the late cornified envelope genes, LCE3C and LCE3B, is associated with rheumatoid arthritis.
Docampo E, Rabionet R, Riveira-Muñoz E, Escaramís G, Julià A, Marsal S, Martín JE, González-Gay MA, Balsa A, Raya E, Martín J, Estivill X.
Arthritis Rheum. 2010 May;62(5):1246-51.PMID:20213803

ACKNOWLEDGEMENTS

Gràcies et al.

Tota aquesta història va començar gràcies al Jordi Monfort que em va proposar fer recerca. Moltes gràcies per la primera empenta!

Gracias a los colaboradores clínicos (Jordi Carbonell y Javier Rivera en particular) y sobretodo a Antonio Collado por su implicación en el proyecto.

Gràcies al Xavier per l'oportunitat que em va donar de poder formar part del laboratori, sobretot tenint en compte que quan vaig arribar no sabia quina era la diferència entre un eppendorf i un falcon…. Moltes gràcies per l'aposta que vas fer amb mi! Ha estat un privilegi poder aprendre tantes coses amb tots voslatres!

A la Kelly per tot el que m'has ensenyat en aquests cinc anys amb la paciència de començar de zero per fer qualsevol explicació. Moltes gràcies per la teva implicació i suport. Sempre seràs la meva mami científica!

Moltes gràcies a tots els components/excomponents (Txema-Maya-Nina) del laboratori-Cegen (Carles-Pili) però sobretot… Al Lluis que, en el inicio de los tiempos, em va tenir "en acogida" uns mesets. Als consells de la Yolanda i de la Eva. Al suport del Manel, el Cristian, el Sergi i la Bruni (siempre nos quedará Toronto!). A la Carre i la Sílvia per fer els arrays d'Illumina com ningú (eso es arte y lo demás son tonterías!). A Santa Geòrgia per fer més o menys terrenal l'estadística i aguantar-me (encara que faci servir el SPSS).A la Mònica Gratacós que va aguantar estoïcament el meu bombardeig de PCAs. A la Laia pel suport informàtic, anímic i volleiballistic… Gràcies per sempre estar disposats a ajudar!

A les microteam pels seus coneixements, entusiasme i solidaritat: gràcies Sílvia por el recuento final de exones de la neurexina, a Mónica Báñez por tu apoyo e implicación a pesar de mis limitaciones (RNAsa nooo!). A l'Eulàlia pels seus consells (amb la ilu que em feia transfectar…) i a l' Eli per les seves classes per a dummies de qPCR i cultius cel·lulars.

Aquesta tesi no hauria estat possible sense el suport incondicional del tripartit Marta-AnnaPuig-Birgit. Muchas gracias Birgit siempre dispuesta a ayudar en TODO en cualquier momento (y por los Mozart también…). A la Morell per ser eficient, solidària y una fuente de sabiduría escato-espiritual. I a la Puig i el seu robot (i els nostres cafés futbolístics…claro…).

A l'Ester Saus que sempre ha estat quan necessitava ajuda (científicament o per trobar el DNI) aportant el equilibri zen a la situació…Y a Susa!!! Mi hermanita científica con la que hemos compartido este camino. Gracias por estar siempre en todo y hasta el final (concretamente hasta la edición de esta tesis). Mil gracias!

Y gracias a Maria-Ainhoa-Sabina, Calvin-Delia, Elisa-Ana-David-Laura-Cristina-Raul-Mar-Agus-marid@s-hij@s et al, Montse-Isidre-Pil, por seguir ahí!

Y, ante todo, gracias a mi madre (además de por el dibujo de la portada) y a mi hermano Ángel que siempre me han apoyado en todo, todos estos años, en cualquier situación y ante cualquier adversidad. Esta tesis también es vuestra!

Y como no, gracias a Cristian por su apoyo incondicional, paciencia y comprensión durante toooooooooda la tesis.

Pues eso, que agradecida y emocionada…No hubiera sido posible sin vosotros…

MIL GRACIAS A TODOS!!!!!!!!!!