CHAPTER 3

SUBJECTS AND METHODS

Land and people

The word Punjab is formed from two Persian words ‘Panj’ meaning five and ‘ab’ which means water (Singh et al. 2008). It is a land of five rivers namely; Sutlej, Beas, Ravi, Chenab and Jehlum, that flew through it during pre independence time. It is one of the 29 states in India. It is located in the North Western region of the country. Geographically, during pre-independence era, it was extended upto Kabul and Lahore, which are now in Afghanistan and Pakistan. However, post independence; it was divided into Eastern Punjab, the one lying within the Indian territories, leaving the other part in Pakistan known as the Western Punjab. In 1956, Eastern or Indian Punjab was further fragmented on the basis of lingual groups, thus forming two new states; Himachal Pradesh and Haryana; hence Punjab was shrunk tremendously from its original size. At present, Punjab shares its international borders with Pakistan on West, regional borders with Jammu and Kashmir on North, with Himachal Pradesh on Northeast, Haryana on East and Rajasthan on South direction. Punjab is divided into 22 districts. The major city and capital Chandigarh is the cultural and educational hub of Punjab, which is also the capital city for its neighboring state; Haryana. Punjab may be divided into four subcategories of districts on the basis of river which irrigated them such as: Majha region, comprised of Amritsar, Pathankot, Gurdaspur and Tarantaran, Doaba region comprising of Jalandhar, Hoshiarpur, Shaheed Bhagat Singh Nagar (Nawanshahr), Kapurthala and Phagwara districts and Malwa region comprising of Ludhiana, Patiala, Barnala, Bathinda, Sangrur, Moga, Roopnagar, Ferozpur, Fazilka, Mansa and Mohali districts.

According to the 2011 census, Punjab has a population of 2,77,43,338 which raised to 2,96,73,462 in 2015, thus accounting for the 2.30 percent of the national population. It is subjected to an annual growth rate of 1.4 percent. The
average male to female ratio was recorded to be 895. Punjab has attained about 75 percent literacy rate, where highest values were recorded in Hoshiarpur district (85 percent). While 37.5 percent of the population is living in cities, the majority of people (62.5 percent) still live in a rural setup. Among cities, Ludhiana is recorded to be the most populated city. Sikh is the majority ethnic group here with 58 percent, Hindu religion is the second largest group present here (38 percent) along with Muslim and Christian population being found in minority. The Jains and Buddhists are also present in negligible fractions. The majority groups are further divided into sub groups such as Jat Sikhs, Brahmins, Banias, Khatris, Ramgarhias, Ramdasis, Gujjars and Rajputs. The majority of population speaks Punjabi followed by Hindi and other languages.

Punjab is an agriculture based state due to its fertile alluvial lands. The major crops grown in Punjab include wheat, rice sugarcane, maize, barley and millet. Owing to the massive boom in economy following green and white revolution, Punjab produces 1 percent of world’s rice and 2 percent of wheat and cotton. Punjab had a GDP of 3.17 in 2015 with 16 percent contribution by agriculture sector. It is one of the richest states in India. Recently, industrial sector is also booming in Punjab, with leading producers of hosiery, textile and steel industries.

Agriculture is one of the largest industries in Punjab. Other major industries include; manufacturing of scientific instruments, electrical goods, machine tools, textiles, sewing machines, sports goods, tourism, fertilizers, bicycles and the processing of sugar and pine oil. Punjab and its people are famous all over the world for its culture, food, Punjabi folk dance, folk songs, festivals etc.

The people of Punjab are considered to be a progeny of a mixture of many proto and post harappan invaders who entered this land from west and became original settlers of this area. It has been inhabited by Harappans, Proto-Dravidians and Indo-Aryans. The people of Punjab today are called Punjabis and their principal language is called Punjabi. The main categories of the population of Punjab region are Jat Sikhs, Khatris, Banias, Ramgarhias, Ramdasis, Rajputs, Lohars, Brahmins, Kambojs and scheduled castes. They are distributed all over the state. In 1947, with the dissolution of British India, the region was
partitioned into India and Pakistan. The Punjab region of India and Pakistan has a historical and cultural link to Indo-Aryan heritage identity as well as partially to the Dravidic indigenous communities. As a result of numerous invasions, many ethnic groups and religions make up the cultural heritage of Punjab. The area has been ruled by many invaders and there is a long line of succeeding rulers of the area, of which Chandragupta Maurya and Ashoka stand out as the most renowned. Muslim emperors ruled Punjab for approximately 1000 years. The period was especially notable for the emergence of Guru Nanak Dev Ji (1469–1539), the first Guru and the founder of Sikhism. At the formation of Dal Khalsa in 1748 at Amritsar, the Punjab was divided into 36 areas and 12 separate Sikh principalities, called misls. From this point onward, the beginning of a Punjabi Sikh Empire emerged. At the time of partition in 1947, the province was split into East and West Punjab. East Punjab (about 35 percent) became part of India, while West Punjab (65 percent) became part of Pakistan.

Subjects and Methods

Subjects

The present study was carried out during October, 2012 to May, 2016. This cross-sectional study involved blood samples of 493 consenting patients who were suffering from chronic musculoskeletal pain and attended orthopedic outpatient departments (OPDs) and pain clinics of Government Medical College and Hospital, Patiala, Dayanand Medical College and Hospital (DMCH), Ludhiana, Orthonova Hospital, Jalandhar and Doctor Hardas Hospital and Advanced Research Centre, Amritsar. These hospitals are tertiary health care providers and cater to the referral patients of Punjab. Total 1147 subjects were screened initially and amongst them, 764 subjects were found eligible after preliminary exclusion criteria as shown in figure 2. These subjects were tested for musculoskeletal disorders by using Nordic Musculoskeletal Questionnaire (NMQ) (Kuorinka et al. 1987). Subjects having congenital analgesia or suffering from any type of pain other than musculoskeletal and/or if its duration was less than three months were also excluded. On this basis, out of 764 subjects, 114 subjects were excluded because they suffered from trauma, injury or surgical pain, cancer pain, neuropathic pain, phantom limb pain, migraine or headache.
and congenital analgesia. 157 subjects were excluded because of having neurological or psychiatric disorders other than depression, endocrinal disorders, post stroke pain, grieved with recent bereavement, were taking hormone replacement therapy, or other calcitropic, corticosteroidal, heparin and anticonvulsant drugs, women with unusual gynecological history, unclear menopause status, irregular cycles or premature menopause before the age of 40 years, or subjects were having multiple disorders such as, complicated hypertension, cerebrovascular infarcts or angina. The final representative data comprised of 493 patients suffering from pain because of various musculoskeletal disorders (Figure 2).

These patients were further categorized according to the severity of pain by using Numeric rating scale (NRS), an 11 point numeric scale for deducing pain intensity (Ferreira-Valente et al. 2011). Patients were categorized on the basis of pain severity into three categories i.e. subjects with mild, moderate and severe pain. In this way, 158 subjects were having mild pain, 171 were having moderate pain and 164 subjects were having severe pain. All these subjects were evaluated for co-existing sleep disturbances. Pattern and quality of sleep in the subjects was evaluated using Pittsburgh Sleep Quality Index (PSQI) (Buysse et al. 1989). All patients gave their written consent prior to participation and the study was approved by ethical committee of the institute.
Figure 2. Flow diagram showing split up of Study participants, NMQ: Nordic Musculoskeletal Questionnaire, NRS: Numeric Rating Scale
Assessment of musculoskeletal disorders and associated pain

All the subjects having pain were assessed for musculoskeletal disorders by Nordic musculoskeletal questionnaire (NMQ) (Kuorinka et al. 1987). NMQ contains 40 forced choice questions based on body map indicating nine anatomical regions of the body on which respondents mark the place of localization of the pain. NMQ has additional 25 questions which provide detailed account of other concomitants such as frequency of pain or discomfort, any accident eliciting the pain, duration of the problem and pain in the last seven days. Further these subjects were assessed for intensity of pain using Numeric rating scale (NRS), which is an 11 point numeric scale. In comparison to other commonly used pain intensity measuring instruments, NRS has been reported to be the most responsive (Ferreira-Valente et al. 2011), easy to use and has higher compliance rate (Hjermstad et al. 2011). Patients were categorized on the basis of pain severity into three categories i.e. subjects with mild, moderate and severe pain. Performas for NMQ and NRS are given as annexures I and IV respectively after summary.

Assessment of depression, cognition and sleep quality

All these subjects were evaluated for co-existing depression, neurocognitive impairment and sleep disturbances. The assessment of depression was done using Patient Health Questionnaire-9 (PHQ-9), which is a self-administered version of the Primary Care Evaluation of Mental Disorders (Prime-MD) (Spitzer et al. 1999). It reflects the criteria of Diagnostic and Statistical Manual for Mental Disorders-fourth version (DSM-IV). The assessment of depression by PHQ-9 is in good agreement with other tests administered by health care professionals (Kroenke et al. 2001, Wittkampf et al. 2007). Its reliability to detect depression has been validated in India (Kochhar et al. 2007). PHQ-9 score ≥10 has been reported to provide 88 percent sensitivity and specificity for diagnosing depression (Kroenke et al. 2001).

Patients were assessed for cognitive impairment by using the standard version of Mini Mental State Examination (MMSE) (Folstein et al. 1975). It comprises 30 point questionnaire which is used to screen arithmetic, language use, memory and orientation skills. Its validity to detect cognitive decline has
been confirmed, whereby MMSE score of <23 has sensitivity of 81.3 percent and specificity of 60.2 percent (Tsolaki et al. 2000).

Pattern and quality of sleep in the subjects was evaluated using Pittsburgh Sleep Quality Index (PSQI). It is comprised of nineteen questions generating seven components of sleep such as subjective sleep quality, sleep duration, sleep latency, habitual sleep efficiency, use of sleep medications, sleep disturbances and daytime dysfunction during the last month. The sum of scores from these components generates global score, which if, ≤5 or >5 indicates good sleep or poor sleep respectively. A global PSQI >5 has been confirmed to deliver 89.6 percent sensitivity and 86.5 percent specificity in differentiating poor sleep from good sleep (Buysse et al. 1989). The proformas for PHQ-9, PSQI and MMSE are appended as annexure V, VI and VII respectively after summery.

Variables

Socioeconomic status of the subjects was calculated according to the updated version of Kuppuswamy and Pareekh scale (Gadhave and Nagarkar 2015). Its categorization was done based on per capita per month income in rupees. Accordingly, subjects having per month income of ≤10,000 were considered as low income group, 10,000-50,000 as middle income group and >50,000 as high income group. Physical activity was determined on the basis that whether subjects were doing atleast 30 minutes of aerobic exercise/walk every day and accordingly were considered active otherwise sedentary. Information regarding lipid levels, duration of pain and statin use was noted down from their medical records. Subjects who had not been tested for lipid levels or if tested 3 months prior to the participation in this study, were tested for their complete lipid profiles. Information regarding marital status, education level, working status, smoking and drinking alcohol was recorded by interviewing them. Height and weight were measured and body mass index (BMI) was calculated according to Quetelet equation (BMI=weight in kilograms/height in meters squared). According to WHO expert consultation (2004) on BMI of Asian Indians, BMI was categorized as normal weight (<23 kg.m⁻²), overweight (≥23-29.9 kg.m⁻²) and obese (>30 kg.m⁻²). Systolic and diastolic blood pressure was noted down as a mean of two tests conducted after an
interval of 3 minutes in sitting position after 15 minutes of rest. The performa for various risk variables is appended as annexure III.

Sample size calculation and Power of the Study

Sample size calculation

In order to determine the number of participants needed to detect pertinent effects within groups without type I and type II errors, a formula given by Lawanga and Lemesho (1991) was used for calculating the sample size for estimating the population proportion with relative precision. The formula is

\[ \text{Sample size} (n) = \left( z_{1-\alpha} \right)^2 * P(1-P) / \epsilon^2 \]

Where \( z_{1-\alpha} \) is confidence level at 95%, whose standard value is 1.96. 'P' is estimated prevalence of musculoskeletal pain in the previous studies and '\( \epsilon \)' is allowable error (relative precision). In the present study the sample size was calculated on the basis of 16.66% prevalence of musculoskeletal pain in north India as reported by Dhillon et al. 2016 and relative precision of 3.5% with 95% confidence interval was used.

Sample size required = 3.84 (16.66*83.43)/3.5*3.5

= 427

The present study involved 493 subjects and this number is sufficient to give pertinent results and efficient to reject the correct null.

Power of the study

Statistical power of the present study was calculated as a priori power analysis according to the method given by Cohen (1988) with a software G* Power (Erdfelder et al. 1996). In this, sample size N is computed, having required power level (1-\( \beta \)), with the pre specified significance level \( \alpha \). The power (1-\( \beta \)) calculated for the present study is the complement of \( \beta \) which denotes that whether type II or \( \beta \) error probability of falsely retaining an incorrect null (\( H_0 \)) is present or not. For calculating the power of the study in G* Power following input values were used for the three parameters required.

1. Significance level (type I error probability) = 0.05
2. The size of the sample in each group used for the test i.e. $P_1$ and $P_2 = 0.5$ each.

3. The population effect size i.e. $1-\beta = 80%$

![Figure 3. A priori analysis showing statistical power of study with $\alpha = 0.5$ and Cohen's $D = 0.5$](image)

By putting these values, the sample size of 493 subjects in the present study delivered 81% statistical power to discriminate efficiently between $H_0$ and $H_1$, based on the priori analysis with effect size of 0.5 (Cohen's $d$), $\alpha = 0.05$ (two tailed) calculated by G*power, a software developed by the Institute of Experimental Psychology, University of Düsseldorf, Germany.

**Collection of blood samples**

For the current study various orthopedic hospitals and clinics were visited across Punjab. 3ml of intravenous blood was drawn by the paramedic staff of the hospitals and stored in Ethylenediaminetetraacetic acid (EDTA) spray coated tubes, kept in the ice box and transported to the molecular biology laboratory of the department. Information regarding risk variables was noted.
down and a written consent of the subjects was obtained prior to the participation in study. The performa for written consent is appended as annexure II.

**Selected genes and their relevance**

COMT and APOE genes were selected as potential candidates for musculoskeletal pain in the present study because in the scientific literature, both of them were found to be associated with neuro-processing, neuro-signaling and neuronal growth (Surguladze *et al.* 2012). The role and relevance of these genes in relation to musculoskeletal pain remains to be explored in the population of Punjab sofar.

**Catechol-0-methyltransferase**

Catechol-o-methyltransferase (COMT) is a major mammalian enzyme responsible for metabolic degradation of catecholamines by catalyzing the transfer of a methyl group from the S-adenosyl-methionine (SAM) to a hydroxyl group on catechol nucleus thus destabilizing the structure. It was first discovered by Julius Axelrod in 1957. Catechol-o-methyltransferase is coded by COMT gene, located on chromosome 22q11.21 in humans and contains 6 exons (Salminen *et al.* 1990).

COMT is the key regulator of dopaminergic system as it degrades the catecholamines such as epinephrine, norepinephrine and dopamine, hence modulates the final outcome of pain, cognitive and stress related behaviors. COMT gene produces two distinct proteins i.e. soluble COMT (S-COMT) and membrane bound COMT (MB-COMT) using single promoter. They have their distinct role in the synaptic transmission, and cortical neurons. S-COMT is found in nucleus and plasma membranes. S-COMT has a greater efficacy of degrading dopamine than MB-COMT as concluded in study kinetics of the two isoforms (Lotta *et al.* 1995). MB-COMT is found abundantly in cortical neurons and dendrites. MB-COMT is capable of inactivating synaptic and extra synaptic dopamine on pre and post synaptic neurons. Moreover, COMT inhibition induces apoptosis of neurons and the cytotoxic mechanism which is dependent on Val/Met genotypes of COMT single nucleotide polymorphism (SNP) rs4680 (Chen *et al.* 2011).
Apolipoprotein E

Apolipoprotein E or ApoE is a type of apolipoprotein found in peripheral and brain tissues. The gene coding for ApoE is located on chromosome 19q13.2 with 4 exons. APOE is a tri-allelic gene comprising of ApoE-2, ApoE-3 and ApoE-4 alleles as defined by preliminary iso-electric focusing studies (Utermann et al. 1980). It is the key transporter of cholesterol, thus regulating lipid metabolism through endogenous, exogenous and reverse cholesterol transport mechanisms. It was first examined by Havekes et al. (1980) in cultured fibroblasts, where suppression of cholesterol occurred on inhibiting the APOE activity.

APOE is observed to play key role in regulating inflammation, stress, cognition, depression, muscle degeneration, apoptosis, nitric oxide synthesis and vitamin K mediated bone growth and neuronal signaling. These processes indirectly, modulate the final outcome of pain perception. In a murine study, APOE gene double knockout mice, who were fed on high fat diet showed increased levels of oxidized low density lipoprotein (LDL) levels in tendons, thus accelerating its degradation and causing increased pain (Grewal et al. 2014). Polymorphisms within APOE gene have a differential effect on the same processes, thus suggesting association of genotypic variation within APOE gene on musculoskeletal pain.

Selection of SNPs and genotyping

The present study involved 7 SNPs of the two genes i.e. COMT and APOE. Four SNPs of COMT gene were; 3’-UTR G/A (rs165599), Exon 3 C/T (rs4633), Exon 4 C/G (rs4818), Exon 4 G/A (rs4680) and three APOE SNPs 5’-UTR G/C (rs440446), Exon4 C/T (rs429358), Exon4 T/C (rs7412) were selected and genotyped (Figure 3,4). The selection of SNPs for genotyping was based on the published reports of public database (http://www.ncbi.nlm.gov/SNP) and criteria was; (1) from previous association studies demonstrating their association with musculoskeletal pain in other populations (2) validated by multiple, independent submissions to the refSNP cluster and (3) validated by frequency or genotype data: minor allele frequency >0.05. DNA was extracted using salting out method from 3ml of fresh whole blood (Miller et al. 1988). PCR was performed with in 25μl reaction volume with 50ng of DNA, 12.5μl of Taq
polymerase mastermix, 2.5μl each of 10X forward and reverse primers and 5.5μl of water. In case of APOE rs7412 and rs429358 SNPs, 2μl of DMSO was added while reducing the water content to 3.5μl. The information regarding primer sequence and the domains of COMT and APOE genes have been summarized in table 1 and 2 respectively. These SNPs were analyzed with polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method. After amplification the amplicons were diluted in 1:20 in nuclease free water and used further for restriction digestion.

![Figure 4. SELECTED SNPs WITHIN COMT GENE](image-url)
The electrophoresis and voltage conditions of amplified products are given in table 3. The detailed information regarding the PCR conditions, size of the digested products of the selected SNPs and the enzymes used for respective SNPs are given in table 4 and 5. The digested products were electrophoresed on agarose gel having concentration from 2.75 to 3.5 percent. The information regarding electrophoresis, gel concentration and voltage conditions is given in table 6. The genotyping was performed blinded to the clinical and disease status in order to avoid bias.

**DNA Extraction**

DNA was extracted from fresh whole blood using salting out method given by Miller *et al.* (1988).

**Method**

1. 3ml of fresh blood sample was transferred into 15ml falcon tubes.
2. RBC lysis working buffer was added to make the final volume up to 15ml.
3. The suspension was mixed gently (15-20 minutes) by inverting the tubes several times with the help of a rocker or by hand until it became translucent.

4. The samples were kept in freezer for about 10 min then centrifuged at 3000 rpm for 10 min and supernatant was discarded.

5. 7.5ml of nucleic lysis working buffer was added in each tube alongwith 0.4ml of 10% sodium dodecyl sulfate (SDS) and 25μl of proteinase K was added.

6. After that samples were vortexed carefully and incubated overnight at 37° C.

7. Next day, 2 ml of 6M saturated NaCl solution was added and tubes were vigorously shaken for 15 seconds. Then tubes were centrifuged at 3000 rpm for 25 min.

8. The supernatant was transferred in 50ml tubes.

9. Added absolute ethanol in 50ml tubes to make the final volume up to 30ml.

10. The tubes were gently rotated 10-15 times for spooling/ precipitation of the DNA.

11. DNA was gently transferred into labeled eppendorfs with a pipette.

12. Added 1ml of 70% chilled ethanol. The samples were carefully vortexed for 3-5 min and again centrifuged for 10 min at 1300 rpm.

13. Supernatant was carefully discarded.

14. Again, added 1ml of 70% chilled ethanol. The samples were carefully vortexed for 3-5 min and centrifuged for 10 min at 1300 rpm.

15. The supernatant was carefully discarded.

16. 1ml of chilled absolute ethanol was added and vortexed carefully for 3-5 min, then centrifuged at 13000 rpm for 10 min.

17. The supernatant was carefully discarded.

18. Eppendorfs were kept in incubator overnight at 37° C for air drying.

19. 0.5 ml of Tris-EDTA (TE) buffer was added in each eppendorf and vortexed carefully, till DNA dissolved completely.

20. After that, DNA samples were dissolved in TE buffer and stored at -20° C till further analysis.
The detailed information regarding composition of all the chemicals and buffers used in the process are given as annexure VIII appended after the summery.

**Quantitative and Qualitative analysis of extracted DNA**

The quality and quantity of DNA can be assessed using two different methods; agrose gel electrophoresis and Spectrophotometry. While the former is a conventional method, spectrophotometry is the preferred method being quicker and precise.

**Quantity and quality check by Agarose Gel Electrophoresis**

**Principle:** Pure DNA, when analysed on agrose gel, forms intact bands whereas smeared and dull bands represent contamination of proteins. Using a 500bp DNA ladder, one can quantify the size of DNA.

**Method:**
1. 1% agarose gel was prepared by taking 1g of agarose and 100ml of Tris borate EDTA (TBE) buffer in a conical flask.
2. The gel was melted and in a separate conical flask, 15ml gel of it was poured in a separate flask and mixed with 4μl of ethidium bromide (EtBr).
3. Gel was casted after sealing the sides of cast with the help of parafilm.
4. Once the gel was solidified, the comb and parafilm were removed gently.
5. TBE buffer was poured in electrophoresis chamber, where gel was placed.
6. 5μl of DNA sample was mixed with 1μl of gel loading buffer (GLB) in agglutination plate.
7. The samples were gently mixed using micropipette and loaded in wells.
8. The gel was run at 100 Volts for 30 minutes.
9. After the run, gel was examined under UV light in the gel documentation (GEL-DOC) unit. Intact bands were visible, which showed that the quality of DNA was good.
**Quantity and quality analysis by Spectrophotometer**

**Principle:** Nucleic acids absorb ultra violet light in a specific pattern. When UV light is passed through nucleic acid samples, the photo-detectors sense the amount of light passing through the solution. The concentration of the nucleic acid present in the solution can be calculated using Beer-Lambert law. According to it, concentration is directly proportional to the amount of light absorbed. Hence, greater the intensity of the light absorbed; greater will be the concentration of nucleic acids in the sample. At wavelength of 260nm, the average extinction coefficient for double stranded DNA is 0.020 (μg/ml)⁻¹cm⁻¹. The optical density or absorbance of 1 corresponds to a concentration of 50μg/ml for double stranded DNA (dsDNA).

**Method:**
1. For 1.5ml of sample preparation, 1490μl of water was added in 10μl of genomic DNA (dissolved in TE buffer) and mixed well.
2. It was incubated for 30 minutes at 37° C.
3. Reading of dilution factor was set to 150 for double stranded DNA (dsDNA).
4. Spectrophotometer was blanked at 260nm using nuclease free water sample as blank in a cuvette.
5. In a separate cuvette, DNA sample was poured and kept in spectrophotometer and optical density (OD) was noted alongwith the quantity of DNA in μg/ml. The absorbance of DNA samples was checked at A_{260} and the reading was noted down. (Spectrophotometer itself gives final reading after multiplying it with 50).

**Calculations**

For quality of the DNA, two readings were noted down, one at A_{260} and other at A_{280} and ratio was calculated. If the ratio is more than 1.8 then it is considered that there is RNA contamination in DNA samples. If the ratio is less than 1.8 then there is protein contamination in DNA samples. The readings were noted down for all samples and ratios were calculated.
**Formula for quantization**

Concentration of the sample\(= A_{260} \times \text{Dilution Factor} \times \text{Conversion Factor}\)

Where \(A\) is the absorbance at wavelength of 260nm, dilution factor (150) and conversion factor = 50.

**Polymerase Chain reaction (PCR)**

**Principle:** PCR is a technique of amplifying targeted DNA or RNA fragments. The amplification is carried out using a thermostable DNA polymerase enzyme such as Taq polymerase which binds the nucleotides into strand through repeated heating and cooling in a process called thermal cycling. PCR involves three major steps:

1. Denaturation
2. Annealing
3. Extension

**Reaction volume**

Hot-start PCR was used to amplify the respective genes. For a total reaction volume of 25\(\mu\)l, 12.5\(\mu\)l of Taq polymerase Mastermix, 2.5\(\mu\) each (10X) of forward and reverse primers, 2\(\mu\)l of template DNA and 5.5\(\mu\)l of nuclease free water were added. For primers having higher GC content as in primers of APOE SNPs; rs429358 and rs7412, 2\(\mu\)l of dimethyl sulfoxide (DMSO) was added while reducing the water to 3.5\(\mu\)l. Detailed information about the product size of individuals SNPs within COMT and APOE genes along with their PCR conditions is given in table 1 and 2.

**Electrophoresis I (amplification checking)**

Agarose gel electrophoresis implies the separation of charged particles on the basis of their charge and molecular weight when placed in an electric field. The electric current moves via charged ions of TBE buffer from negative to positive direction, while DNA moves through the pores in agarose structure along with the direction of charge. The smaller the size of a fragment, greater will be its mobility. Detailed information regarding the gel concentration and voltage conditions has been given in table 3.
**Preparation of gel**

Gels of variable concentrations were prepared ranging from 2.75 to 3.5%, by dissolving 3g (3%) and 3.5g (3.5%) of agarose in 1X TBE buffer. 15-20ml of melted gel was mixed with 8μl of EtBr and casted and left for solidifying.

For analyzing PCR products, 6μl of amplicon was mixed with 0.5μl of GLB and loaded. For the size estimation, 3.5μl of 50 base pair agarose DNA ladder was added in one of the wells.

**Restriction enzyme digestion**

Restriction enzyme digestion is a method of cleaving DNA/RNA sequences using enzymes which recognize and cleave at specific sequences. A restriction enzyme buffer enhances the efficacy of the enzyme.

**Method**

The restriction digestion mixture of 9.25μl volume was prepared using following composition

a. Nuclease free water: 1μ
b. Amplicon: 5μl
c. Restriction buffer: 2.5μl
d. Restriction enzyme: 0.75μl

Samples were gently mixed and kept at 37°C overnight and electrophoresed the next day. Detailed information about the product size of individuals SNPs within COMT and APOE genes along with their PCR conditions and respective restriction enzyme used for each SNP is given in tables 4 and 5.

**Electrophoresis II (restriction digestion checking)**

For analyzing PCR-RFLP products, 8μl of amplicon was mixed with 0.5μl of GLB and loaded. For the size estimation, 3.5μl of 50 base pair agarose DNA ladder was added in one of the wells. The gel was inspected under UV-light and genotypes were noted. Detailed information regarding agarose gel concentration used and voltage conditions for individual SNPs are given in table 6.
The genotypes were noted down from the gels and were transferred on MS excel programme, keeping the identity of the subjects blinded in order to avoid biasness.

Statistical analysis

Baseline characteristics, variables and their differences

The data regarding general characteristics of the subjects with mild, moderate and severe pain are presented as mean ± standard deviation and numbers. The differences in proportions were analysed using chi square test after Yates’ correction, wherever necessary and continuous variables were analysed by student’s t-test. A linear regression was applied to investigate the association between risk variables and pain. Those variables which showed linear relationship with the dependent variable (P<0.05) in univariate testing were further included in multivariate logistic regression analysis (backward stepwise) to identify independent association of the significant variables.

To measure an association between an exposure (risk variable) and an outcome (risk of moderate and severe pain), odds ratios (OR) were generated by the formula given by McHugh (2009). The 95% confidence intervals were used to estimate the precision of the OR and P value <0.05 was considered significant except for multiple comparisons where Bonferroni correction was applied.

Allele frequencies and genetic model analysis

Allele frequencies were calculated by gene counting and their departure from Hardy-Weinberg Equilibrium (HWE) was analysed by $\chi^2$ test. Logistic regression was used to test the association between the variants as well as wild-type genotypes and the risk of moderate and severe pain under additive, recessive as well as dominant models of inheritance.

Haplotypes and their association with the risk of musculoskeletal pain

Expectation maximization (EM) algorithm was employed on the genotype data for deducing haplotypes by using ARLEQUIN software (ver. 3.01) (Excoffier et al. 2007). The comparison of haplotype frequencies was performed.
using equation $\chi^2 = \frac{(ad - bc)^2 \times n}{(a + b)(a + c)(b + d)(d + c)}$, where a and b represents the number of chromosomes of each respective haplotype in subjects with moderate pain and subjects with mild pain respectively or subjects with severe pain and subjects with mild pain respectively, while c and d represent the residual chromosomes. The total number of chromosomes in subjects with either moderate or severe pain and subjects with mild pain is represented by n. Linear regression (GLM procedure) was applied to investigate the association of haplotypes with moderate and severe pain. Odds ratios were generated and analysed by taking most common haplotype as a reference.

**Gene-Gene and Gene-Environmental interactions analysis**

Gene-Gene (G x G) interactions in relation to the risk of moderate and severe pain were observed from the genotype data implementing extended Kempthorne model by software epiSNP (Ma et al. 2008). Out of several hundred SNP-SNP interactions only significant interactions for the risk of moderate and severe pain are presented. Gene-Gene (SNP-SNP) analysis test corresponds to the interactive effect (I), allele x allele (A x A), allele genotype (A x D), genotype x allele (D x A) and genotype x genotype (D x D) effect.

Gene x Environmental (G x E) analysis was also done with the software epiSNP. All the significant quantitative variables (from backward stepwise regression model) were included in the model comprising every single SNP vs. variables. Out of several G x E effects only significant effects are presented. G x E tests correspond to overall marker effect (M), dominant effect (D) and additive effect (A).

**Two way epistatic effects**

The epistatic interactions between two loci were determined by epiSNPPlot programme. In the SNP-SNP epistatic effects, line colors between two respective SNPs corresponded to; Black color: I- Interactive effect between SNPs, Green color: DD-Dominance x Dominance effect, Blue color: DA-Dominance x Additive effect, Purple color: AD-Additive x Dominance and Red color: AA-Additive x Additive effect. A pair of lines between two respective SNPs represents significant (P<0.01) two way epistatic effects.
Table 1. SNPs, domains and their primer sequences of studied SNPs in COMT gene

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Alleles&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Domain</th>
<th>Primer (5’-3’)</th>
<th>Reference</th>
</tr>
</thead>
</table>
| COMT | rs165599         | G>A              | 3’-UTR | F: GACTTGCCACCAACATTC  
R: TGCTTGTCAGAAAGTTG | Ghosh <i>et al.</i> 2013 |
| COMT | rs4633           | C>T              | Exon 3 | F: TACCGGCTGGAACGAGTTCAT  
R: CTCTGCTCGAGTAGGTTC | George <i>et al.</i> 2008 |
| COMT | rs4818           | C>G              | Exon 4 | F: CAACCTGACAGGCAAGAT  
R: GCCCTTTTTCCAGGTCTGACA | George <i>et al.</i> 2008 |
| COMT | rs4680           | G>A              | Exon 4 | F: CTCATCACGATCGATCA  
R: CCAGGTCTGACAAGGGTCA | Albaugh <i>et al.</i> 2010 |

<sup>a</sup>SNP IDs according to the database dbSNP (www.ncbi.nlm.gov/SNP).  
<sup>b</sup>Bold face letters are minor alleles.  
COMT: Catechol-o-methyltransferase.
Table 2. SNPs, domains and their primer sequences of studied SNPs in APOE gene

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Alleles&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Domain</th>
<th>Primer (5’-3’)</th>
<th>Reference</th>
</tr>
</thead>
</table>
| APOE | rs440446         | G>C              | 5’-UTR | F:TCCCCAGGAGCCGGTGA  
               |                   |      |                    | R: CCCCAAGCCGACC | Long <i>et al.</i> 2003 |
| APOE | rs7412           | C>T              | Exon 4 | F:ACAGAATTCCCCCGGCCTGGTACAC  
               |                   |      |                    | R: TAAGCTTGCGACGGCTGTCCC | Singh <i>et al.</i> 2010 |
| APOE | rs429358         | T>C              | Exon 4 | F: ACAGAATTCCCCCGGCCTGGTACAC  
               |                   |      |                    | R: TAAGCTTGCGACGGCTGTCCC | Singh <i>et al.</i> 2010 |

<sup>a</sup>SNP IDs according to the database dbSNP (www.ncbi.nlm.gov/SNP). <sup>b</sup>Bold face letters are minor alleles.

APOE: Apolipoprotein E
Table 3. Agarose gel electrophoresis for the quality check of the amplified products

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP</th>
<th>Agarose Gel (percentage)</th>
<th>Voltage conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMT</td>
<td>rs165599</td>
<td>3 percent</td>
<td>70V for 1 hour 10 min</td>
</tr>
<tr>
<td>COMT</td>
<td>rs4633</td>
<td>2.75 percent</td>
<td>70V for 1 hour 15 min</td>
</tr>
<tr>
<td>COMT</td>
<td>rs4818</td>
<td>3 percent</td>
<td>100V for 1 hour</td>
</tr>
<tr>
<td>COMT</td>
<td>rs4680</td>
<td>3.5 percent</td>
<td>70V for 1 hour</td>
</tr>
<tr>
<td>APOE</td>
<td>rs440446</td>
<td>3 percent</td>
<td>70V for 1 hour 10 min</td>
</tr>
<tr>
<td>APOE</td>
<td>rs7412/rs429358</td>
<td>3 percent</td>
<td>70V for 1 hour 15 min</td>
</tr>
</tbody>
</table>

COMT: Catechol-o-methyltransferase, APOE: Apolipoprotein E, V: Volts
Table 4. PCR conditions, restriction enzyme and fragment size of SNPs within COMT gene

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP</th>
<th>PCR conditions</th>
<th>Restriction Enzyme</th>
<th>Fragment size</th>
</tr>
</thead>
</table>
| COMT | rs165599 | Initial denaturation: 95°C for 10 min | MspI | 235bp  
Denaturation: 94°C for 40 sec  
Annealing: 60°C for 40sec  
Extension: 72°C for 40sec  
Total cycles= 35  
G= 235bp  
A=181, 54bp |
| COMT | rs4633  | Initial denaturation: 95°C for 10 min | BsaAI | 341bp  
Denaturation: 94°C for 40 sec  
Annealing: 61°C for 40 sec  
Extension: 72°C for 40 sec  
Total cycles=32  
T= 341bp  
C= 82,79bp |
| COMT | rs4818  | Initial denaturation: 95°C for 10 min | BclI | 296bp  
Denaturation: 94°C for 40 sec  
Annealing: 61°C for 30 sec  
Extension: 72°C for 40 sec  
Total cycles= 35  
C= 296bp  
G= 116, 145,35bp |
| COMT | rs4680  | Initial denaturation: 95°C for 15 min | NlaIII | 109bp  
Denaturation: 95°C for 30 sec  
Annealing: 54°C for 30 sec  
Extension: 72°C for 40 sec  
Total cycles= 35  
G= 86, 23bp  
A= 68, 23,18bp |
Table 5. PCR conditions of restriction enzyme and fragment size of SNPs within APOE gene

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP</th>
<th>PCR conditions</th>
<th>Restriction Enzyme</th>
<th>Fragment size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APOE</td>
<td>rs440446</td>
<td>Initial denaturation: 95°C for 15 min</td>
<td>HhaI</td>
<td>238bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denaturation: 94°C for sec</td>
<td></td>
<td>G= 73,145bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annealing: C for sec</td>
<td></td>
<td>C= 238bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extension: 72°C for sec</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total cycles:35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOE</td>
<td>rs7412</td>
<td>Initial denaturation: 95°C for 15 min</td>
<td>HhaI</td>
<td>244bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denaturation: 94°C for 1 min</td>
<td></td>
<td>T= 91, 83, 38, 32bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annealing: 50°C for 1 min</td>
<td></td>
<td>C= 91, 48, 38, 35, 32bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extension: 72°C for 1 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total cycle: 35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOE</td>
<td>rs429358</td>
<td>Initial denaturation: 95°C for 15 min</td>
<td>HhaI</td>
<td>244bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denaturation: 94°C for 1 min</td>
<td></td>
<td>C=72, 48, 38, 35, 32,19bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annealing: 50°C for 1 min</td>
<td></td>
<td>T=91, 48, 38, 35, 32bp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extension: 72°C for 1 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total cycle: 35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Agarose gel electrophoresis and voltage conditions for genotype counting

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP</th>
<th>Agarose Gel (percentage)</th>
<th>Voltage conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMT</td>
<td>rs165599</td>
<td>3 percent</td>
<td>70V for 1 hour 20 min</td>
</tr>
<tr>
<td>COMT</td>
<td>rs4633</td>
<td>2.75 percent</td>
<td>70V for 1 hour 30 min</td>
</tr>
<tr>
<td>COMT</td>
<td>rs4818</td>
<td>3 percent</td>
<td>100V for 1 hour 15 min</td>
</tr>
<tr>
<td>COMT</td>
<td>rs4680</td>
<td>3.5 percent</td>
<td>70V for 1 hour 15 min</td>
</tr>
<tr>
<td>APOE</td>
<td>rs440446</td>
<td>3 percent</td>
<td>70V for 1 hour 15 min</td>
</tr>
<tr>
<td>APOE</td>
<td>rs7412/rs429358</td>
<td>3 percent</td>
<td>70V for 1 hour 30 min</td>
</tr>
</tbody>
</table>

COMT: Catechol-o-methyltransferase, APOE: Apolipoprotein E, V: Volts